INTERNATIONAL INSTITUTE FOR ANALYTICS

Enterprise Research Service

Research Brief

Data Visualization:
7 Considerations for Visualization Deployment

Authored by:

Analise Polsky

Big Ideas

- Data visualization tools are powerful means to democratizing data within an organization and enabling non-statistically trained users to rapidly gain insights about critical business decisions.
- Before implementing a visualization tool, it is critical to understand what decisions it will support within your organization and establish a plan for rollout and training of users.
- As with all analytics applications, data quality, definitions, governance and management must exist.

Introduction

It's 8 a.m. After a long, slow sip of coffee, your hand falls to your mouse. The Web browser pops open to Bloomberg News. Your eyes scan the morning's story highlights as stock prices rapidly stream by on your ticker toolbar. A cluster of familiar letters whiz past. You have your eye on a small tech company that went public a few months ago and wonder how it's doing. A quick search reveals a line graph showing the percentage changes in price over the last six months. In just the last few days, the tech company's stock price is up 5 percent and looks to continue this trend. You contentedly lean back in your chair, happy you bought early.

A glance at the stock ticker may not suffice. Sometimes numbers alone cannot tell us everything we need or want to know. We rely on visual cues to absorb and process information. Visuals are the single best way our brain processes information. It's not surprising then that data visualization – the pictorial or graphical representation of data – plays a pivotal role in how we explain and interpret the data we have come to depend on. Data visualization is steadily growing in popularity and spreading to the mass market. But why now?

Rising data volumes, increasing data complexity, and the availability of new data sources can make it seem hard to find and disseminate important nuggets of information. Visualization tools have risen to the challenge, making it possible to bring together dimensions of data from disparate data sources, drill down, dynamically interact with the visuals, and even produce forecasts in a single user interface. Visualization software harnesses the power of analytics and adds a visual presentation layer. From animations to geospatial analysis, there are many ways to work with the data. And the best part is, the software is easy to use. You don't have to be a technical user to make data visualization software work for you. Problem solved. Sort of.

Visualization tools do not function in isolation. They are part of a larger operational picture. Getting the most out of your tools means understanding how people will use it and which users it affects. You may need to create new policies or redefine your internal data management processes. Whether you are on the cusp of implementing a visualization solution or already have one, the considerations below will give you a way to think beyond features and functions, so you can derive the maximum business value from it.

1. Deploy with purpose

Having a clear idea of how users will apply the technology to solve specific business issue(s) prior to purchasing or implementing it will maximize value from the investment.

A purpose-driven approach to deploying data visualization will highlight key use cases and reveal the optimal features of the visualization software. It will also help you gain the support of leadership, which is essential for any project to gain momentum. A strong purpose also makes it easier to identify whom the software affects and how. Who are your advocates and stakeholders? Are there easily identified naysayers or saboteurs? Advocates and detractors are valuable sources of information. Based on their insight, you can plan a formal communication strategy and avoid potential user roadblocks.

Evolving regulatory compliance mandates in financial services illustrate a compelling opportunity for data visualization. In the European Union, the Solvency II Directive brings together insurers under one regulatory framework, protecting consumers by making sure insurers hold enough capital to offset their risk of insolvency. Insurers are required to provide detailed risk reports. They must also trace and monitor data quality on the criteria of accuracy, appropriateness and completeness.

Solvency II compliance monitoring can be facilitated through a dashboard generated in a data visualization tool. Working from the requirements established in the directive, color-coded gauges are a means of monitoring and tracking key performance indicators for data quality. A dashboard also facilitates tracking of executed data quality rules and the frequency of rule violations. Seeing violations and alerts in conjunction with the data is extremely valuable, enabling you to rapidly identify problems and their potential sources allowing action to be taken quickly.

Public health monitoring of rates of disease across different patient populations and understanding regional variations or patterns in the concentration of chronic diseases, such as diabetes, are another opportunity to take advantage of data visualization tools. The ideal way

for the agency to view the information is through geospatial analysis or mapping patient information to geographic location. Traditionally, spreadsheets would capture this information with line item subtotals for regions and the country as a whole. Using geospatial analysis makes the hot spots and trends far more obvious with very little effort on the part of the analysts. This information helps organizations make decisions about staffing, facility management and community health programs. Visualization tools can also provide Web or mobile access to key reports and data sets for public health entities to distribute to decision-making bodies, such as policymakers and other health agencies.

2. Understand what data supports your decisions

Dedicate time to understanding the decisions you need to make and the data that will support those decisions.

Data is the core of every operational and systematic business process. The more you know about the data, the more empowered you will be when it comes time to interpret analytic results.

Data comes from an ever-expanding list of sources including data warehouses, CRM systems, spreadsheets, Web applications, flat files, XML documents and cloud providers. Once you have identified your business issue, you may discover that the data you need to resolve it will come from a combination of sources. Often, feeding more data into the analysis is better.

For instance, a large shoe retailer is facing a downward turn in sales. Their marketing campaigns are slow to develop, and they are not sure they have a single view of the customer across their brands. The data they will need to analyze and subsequently visualize will come in part from CRM and transactional data. The retailer may also want to look at data coming from online reviews or social data.

The requirements for visualizing and analyzing structured transactional data differ from those of social and other forms of unstructured data. Text analysis can extract, analyze, categorize, and apply metadata to unstructured sources like warranties, reports, Twitter feeds and online comments. In general, the text must go through a preparation process, followed by linguistic processing, and then loading into a data table so it will have the necessary structure for visualization and further analytics. The sub-steps will depend on the data you are working with and will shift based on the problem you are trying to solve. For example, you may need content-specific taxonomies or hierarchies to organize the data. Text analysis is a technical

process that requires data mining, which is often conducted by a data scientist. Just be aware of the additional steps up front, so you can plan for and execute your analyses efficiently.

The types of data and the decision you are making also determine which type of visualization will best represent the data. For text analysis, you may use traditional bar charts to show sentiment, employ network diagrams to show connections between consumers, or use concept linkage to show associations between words. Knowledge about the structure and content of the data will inform which approach to visualization you need to use, ultimately improving your understanding of the results.

3. Define your data

Analytics cannot be performed if the data is not consistently understood and of a reasonable level of quality.

A shared understanding between individuals and groups as to what data means is critical. Variation in definitions does nothing but lead to confusion and reduce the validity of analysis. The sales department may view the region "South" as 10 states in the southeast corridor of the US. Product marketing, on the other hand, may consider "South" to be 12 states. IT's definition could contrast with business' entirely. Analytics software cannot account for these variations.

Standard definitions promote consistency in data comparisons and facilitate communication about the results because everyone is working from the same set of basic assumptions. It also eliminates disagreements about how to interpret the results of analysis. Remember, you do not have to define every single data element up front. Use the business issue to identify a small control project so there is a clear starting point and scope for creating the definitions. Then expand to other critical data elements.

In addition, you need to determine the acceptable quality of the data. In other words, establish what amount or degree of bad data is acceptable for the issue you are working on. For example, a 3 percent error in mailing addresses may be acceptable to a manufacturing firm, but greater than a 2 percent error in defect description codes may not.

Visualization tools can act as collaboration platforms for the conversations around all of these issues.

4. Formalize accountability and oversight

Establish data stewards, data governance and data management.

You may be wondering if the responsibility for overseeing the data shifts or changes as you introduce visualization software; in particular, software that enables more types of users to work with the data. No, it doesn't. In fact, it becomes even more important that you have data stewards who are accountable for your data and robust data governance and management.

Data stewards are the "go to" people when there are questions, concerns or doubts about the quality or overall management of the data. They speak IT and business, often mediating the relationship between the two groups. In other words, they know what the business drivers are and how the data supports them. As mentioned in consideration 2, which discusses the importance of standardizing our data definitions, creating the initial definitions will involve individuals from different business units and departments. The upkeep, recording and maintenance of the definitions should fall on someone accountable for the data's meaning and usefulness. The data steward is a formal role, and depending on the size, data volume and demands of your industry, you may need more than one in your organization.

Data stewards play a significant role in implementing data governance. Governance explains where the data is generated, its architecture, and how it is administered, as well as security and access rights. Governance ties business strategy to the policies and processes that surround data. It is an organizing framework, not just policy. Data management is the execution component of data governance.

An investment bank that has multiple lines of business including private equity funds, risk management, mutual funds, and foreign exchange is highly siloed. The units do not share data, proliferating it through spreadsheets. If you were to ask, "Who in this type of environment is responsible for the data?" you would see multiple individuals raising their hands on behalf of their data. Most users would not even know where the data came from originally. If this sounds familiar, you probably need to launch data stewardship.

Democratized data does not mean free rein to change the data or use the data as we see fit. Ensure that only the people who should have access in fact do. Tracking usage and monitoring access are important when so much data is sensitive in nature. Even if the data is not sensitive, you want to ensure that rules are in place that enable people to explore without making changes to the source data (unless they should be). In the end, establishing clear boundaries over data use and the dissemination of data can break down silos. (Data stewards can help with this too.)

5. Consider the user

Each function within a company will benefit from data visualization tools differently.

Data visualization is a powerful means of finding trends, patterns, gaps and inconsistencies in the data. “Consider the user” means recognizing that, aside from the overall benefit of data visualization, each function in your company will likely benefit in different ways.

IT can use charts and graphs to monitor and track usage rates, processing speeds, hardware resources and user session information. Are there spikes in CPU and memory usage? How heavily are the operating systems (i.e., Linux, UNIX or Windows) using our servers?

What about a business user? A cost accountant for a manufacturing firm, for example, will review the data on cost variances and determine the profitability of projects. The visualization software should help him figure out which assets are the most costly or how much maintenance costs by asset groups. He will also need the flexibility to perform calculations against the data and provide regular reports to management. The tools can easily

accommodate different categories of business users, who may not all need to see the data the same way.

Know your audience. Think about who consumes the data and how they consume it. This will help you as you try to figure out the most effective way to get people to adopt a visualization tool and promote enterprise-wide use of analytics.

6. Develop skills

Users of visualization tools must have training to create and correctly interpret analytically derived visualizations.

In most cases investing in a new data visualization toolset is not a rip and replace scenario, wherein the tool is replacing existing analytical toolsets. More often than not, a data visualization tool fills a discrete need to deliver analytic results in a manner that makes it usable

by non-statisticians – a new user community inexperienced with fact-based decision making or complex data that needs to be distilled into graphs to make it more comprehensive.

Visualization tools alone can't produce better decision making or insight, nor will they create a self-sufficient user community. They are a key component of self-service environments as the intuitiveness of today's tools and the automation of many analytical calculations facilitates the user's ability to develop a deeper understanding of analytics. Training is still required, regardless of the user's prior skillsets. Visualization tools offload some of the development tasks from BI/analytics teams; however, some of the gains are offset by ongoing enablement and training.

Consideration 2 outlines the basic ideas behind understanding the data and highlights the importance of establishing standard definitions for the data. Users will need to be educated on those definitions, as well as a basic knowledge of data and analytic principles.

Users should also have access to information on effective graphing and charting techniques. Why and when would you use a bar chart or a pie chart? For users building dashboards and reports, a basic sense of design is also essential. Do they know to limit the number of colors and visuals per screen? Or which colors are most effective? You do not have to be a visualization expert to learn basic design principles.

Visualization software can also liberate IT from some of the reporting demands of business users. A business user who does not need to write code to create a data visualization relieves IT of performing that task. The same is true of creating basic reports. It is easy to overlook the workload associated with distributing and providing data that rests on IT.

7. Manage expectations

Functionality shouldn't be dropped on the user community wholesale. The rollout process should be incremental and include training, education and a communication strategy. Small controlled projects are a way of providing proof-of-concept and allow for changes during the rollout process.

What Jill Dyché and Evan Levy call "neat to know" analysis in their book, *Customer Data Integration* (Wiley, 2006), is still all too frequent in today's companies and something to beware of when implementing a new tool. Exploring data in a "neat to know" fashion is analysis without a clear link back to a business goal or initiative. It's data analysis without teeth. Here are three key questions you can ask: What is the problem? Can you act on it? Can your data

help? Simply reviewing reports or exploring data, however intuitive, is no substitute for making practical business decisions. Organizations should strive to focus analytics on improving key business decisions.

Conclusion

Data visualization tools are only now coming of age. We will see tremendous growth in this area over the coming years, whether it be through increasing functionality of mobile applications or dynamic visual interfaces. How data visualizations are used and exploited will continue to depend on the community consuming the resulting information. Data scientists may have the first crack at sifting through the data, interpreting patterns and trends, and using data visualizations to represent them in a meaningful way. Likewise, business analysts will work with the data scientist to interpret the results and decide if he needs to know more. Eventually, the results may work their way to senior management. And at this point, they might be further distilled into a report or dashboard. The audience consuming the resulting information will be the final arbiters of the value of data visualization.

Managing expectations before, during and after the adoption of visualization software is crucial. Users should know what the rollout process will look like and how it will take place, and have clear goals for using the tool. Make sure that the desired outcome isn't just look-and-feel. Creating beautiful charts and graphs is not a substitute for practical business decisions.

Deploying a data visualization tool is one part technology, one part analytics, and two parts change management. Data visualizations themselves are the driver in what makes adoption possible. They will continue to play a central role in how you communicate information internally and to the world. The more attention you give to the enabling pieces, the better the visualization tools will support and benefit the growth and evolution of your business.

About the Author:

Analise Polsky is a thought leader on the SAS Best Practices team. The focus of her work is developing and delivering content on the subjects of data visualization, quality and stewardship, as well as culture and change management. She has also created courseware for SAS® DataFlux® products. Prior to joining SAS, she provided training on database and application products to private and public sector cliental in English, Spanish and Portuguese. She holds a master's degree in public health and has conducted extensive academic research in the fields of economics and anthropology. She has contributed to research projects in Peru, Brazil and Mexico, and has been published in academic journals.