What Do Data Analysts Most Need from Their Tools?

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter
April/May/June 2015

A decade or so ago, analytics became recognized as a critical activity and largely untapped opportunity of organizations—especially businesses. This led to the current popularity of *big data*, *data science*, and *data visualization*, as well as a renewed appreciation for *statistics*. As a result, existing software vendors in the business intelligence (BI) space and enterprising entrepreneurs jumped onto the analytics bandwagon in droves. Unfortunately, few of the tools that have emerged during the last decade to support the needs of data analysts have provided useful solutions. None of the tools are anywhere near what they should and could be. Skilled data analysts—not the masses who have added "analyst" to their job titles—are frustrated by the paucity of viable tools and the clumsiness of the few that exist. What do data analysts actually need from their tools? It isn't what most of the vendors, including the few that have produced decent tools, are focusing on.

Two recent encounters prompted me to write this article: 1) a product briefing by a well known company that is automating the process of anticipatory chart generation (i.e., generating charts before they're requested by the analyst), and 2) a request for consulting assistance from a software company that says it is incorporating *artificial intelligence (AI)* and *natural language processing (NLP)* into its data analysis product. Both companies are attempting to build intelligence into their products to reduce some of the work that data analysts routinely do. There is a problem with both of these attempts, however. Their attempts to assist data analysts are actually getting in the way and slowing them down. In both cases the vendors don't understand where to draw the line between the work that a computer can do and the work that requires the human brain.

In an email from the second of these companies, the co-founder explained:

> We have added AI and NLP to our product in order to help the user choose from multitudes of paths that he can go down while performing analytics. Our interest is building a product that truly becomes an extension of the data analyst thought process.

> Our hypothesis is…analysis by rejection rather than selection. We believe it's easier for users to reject vizualisations rather than imagine what a vizualisation should be. So we offer every user a buffet of limited relevant charts to look at and select one of them.

You have perhaps noticed a recent rise in vendors and thought leaders touting the potential benefits of AI and NLP for BI. These are just the latest in a perpetual stream of short-lived and ill-conceived fantasies that pique the interest of CIOs and other corporate decision makers when shopping for products. I find it hard to imagine how either AI or NLP can be meaningfully used to support the work of a skilled data analyst who neither needs nor wants the computer to think for her and certainly doesn't want to direct the computer by writing sentences. Perhaps my imagination is stilted. Then again, perhaps I understand the needs of data analysts better than most.

When I explore a data set or pursue a path of analysis, I usually don't want my tool to guess what I need to see other than a simple transition in graph type based on the nature of the data that I've selected. For example, if I suddenly add a temporal variable (i.e., date or time) to the data that I'm viewing, it is a fair guess that I'd like to switch to a line graph to easily see and compare patterns of change. Automating this switch would save me a click of the mouse (or more with some tools), which would rarely result in an undesirable change. If my tool tries to anticipate my needs beyond this, however, it will almost always guess wrong, forcing me to undo what it's done, and thus wasting my time and breaking the flow of analysis. Based on the data that I've selected, I certainly don't want my tool to automatically generate an entire collection of graphs that I must review to determine my next view, which is what these two vendors believe that we need. Chances are, none of the

graphs that were automatically generated would match what I want as my next view, and even if one did, I would have to review several graphs to get to it.

When I look at data, the next view that I need comes to mind as soon as I've absorbed what I'm seeing. I don't want to review my tool's naive guesses about my needs. What I really want is a simple and efficient path to the next view that I've already envisioned. In other words, I want a really good *human-computer interface*. Perhaps every once in a while I'll get stuck and find it hard to imagine my next useful view, in which case it might make sense to ask the tool to generate a few possibilities based on a good set of heuristics, but only then.

Unlike a skilled data analyst, a novice might benefit from an automatically generated set of potentially useful graphs. Even for the novice, however, this intervention would only be useful if it is designed to teach him appropriate choices, but this will only work if the tool explains in the moment why particular choices are useful. A novice who lets the computer make these judgments for him without understanding the reasoning behind these judgments will forever remain a novice.

Data analysis software should never attempt any of the following:

1. Do what a data analyst could do better
2. Take the data analyst out of the loop of awareness and thinking
3. Pretend that it can think

Most vendors, rather than focusing their efforts on the functionality that data analysts actually need, are looking for new and sexy features that make the people with purchasing authority—not the actual users—flush with excitement (e.g., spinning pie charts with variable speeds, both in forward and in reverse).

What data analysts most need from their tools today is not innovative functionality, but the following:

1. *Effective graphs.* Analysts need a limited set of graphs that are useful and well designed. They don't need a huge gallery of graphs, few of which are actually useful. These graphs should be properly formatted to feature the data by default and should possess functionality and formatting flexibility to adapt them as needed.

2. *Effective data interactions.* Analysts need a limited set of interaction methods (filtering, sorting, annotating, etc.) that are useful, well designed, and efficient to use. There are a few ways that analysts frequently interact with data to change what they're viewing or how they're viewing it. These methods should be readily available and easy to initiate without distracting from the flow of analysis.

3. *Useful statistics.* Proven statistical methods should be readily available and well designed to produce reliable results. The statistics that are most frequently needed should be close at hand and those that are less frequently needed should be easy to find. What's provided should be extremely well documented with clear examples and instruction for proper use.

4. *Effective human-computer interfaces.* Although last in the list, this is of equal importance to the other requirements. The interface should make it easy for analysts to efficiently interact with data without distraction. The interface should seamlessly distribute the work, sometimes to the analyst and sometimes to the computer, in a manner that leverages the strengths of both.

Instead of being constantly enamored by the latest claim to technological nirvana (e.g., AI), we should demand tools that finally get these basic requirements right. Effective graphs, interactions, and interfaces can only be provided by vendors that understand the human brain, and, to state the obvious, useful statistics can only be provided by vendors that understand statistics. It is the rare vendor that possesses adequate expertise in both of these areas.

I didn't include the need for *data extraction*, *cleansing*, *transformation*, and *loading (ECTL)* functionality because I'm strictly limiting the scope of data analysis to direct data sensemaking tasks, excluding the work that still too often must be done by analysts to prepare data for sensemaking. If you work in an IT department, you might wonder why I left *fast performance* off the list. Actually, I didn't miss this, for it is part of what makes an effective human-computer interface. As an analyst, if you must wait too long for the next view of data that you need, the interface isn't working. Anything that technologies must do to support the four basic needs of

data analysts listed above are not actually requirements; they are merely the means to meet a requirement. It is always a mistake to focus on specific technological means to achieve our objectives as objectives in and of themselves. To do so disconnects us from our purpose: effective data sensemaking to support better decision-making.

Contrary to what you might believe, so-called Big Data has not changed what data analysts need. Neither the volume, velocity, nor variety of data affect the fundamental nature and process of data analysis. What data analysts most need from their tools today does not differ substantially from what they needed when I first became an information technology (IT) professional over 30 years ago. Until we shift our focus to the real needs of data analysts, we'll continue to wander in the wilderness, forever dreaming of the promised land of enlightened analytics that lies just over the hill.

Despite the fact that data analysis technologies have made some progress during the last 30 years, only a handful of tools exist today that provide viable support for *exploratory data analysis (EDA)*. Programming tools such as R are powerful analytical tools, but they are not designed for EDA. Writing code is an appropriate interface for constructing specific analyses or for building applications that support specific analytical tasks, but it is not appropriate for data exploration, which must enable the analyst to get to that next view that's needed without becoming distracted by the mechanics of interacting with the software.

Based on the data analysis products that I've seen, only the following tools, in no particular order, are viable for EDA today:

- TIBCO Spotfire
- Tableau
- Advizor Analyst
- SAS JMP

A few more are potential candidates for this list if and when they mature:

- QlikSense
- SAS Visual Analytics
- DataWatch Designer (formerly Panopticon)

If others exist, I haven't seen them. It is certainly possible that other viable EDA tools exist for specific data domains (e.g., for oil exploration or for genetic research), but I'm only considering generic tools that can be used across many domains.

Even the products in my list of four viable tools are all far from what they should and could be.

- TIBCO Spotfire has a venerable pedigree, emerging from doctoral work that was done by Christopher Ahlberg at the University of Maryland under the direction of Ben Shneiderman, but its interface could be improved, some of its functionality (e.g., parallel coordinates as well as brushing and linking) remains hobbled, and since its acquisition by TIBCO, its focus has become less pure and more susceptible to influence from the marketing department.

- Tableau also has rich roots, emerging from doctoral work that was done by Chris Stolte at Stanford University under the direction of Pat Hanrahan and highly influenced by earlier work by Jock Mackinlay, but the integrity of its original vision was compromised as the company grew much too rapidly and its attention shifted to pleasing Wall Street. Some of its functionality has been implemented incompletely (e.g., brushing/linking and tree maps) or sloppily (e.g., the formatting interface), and some of it should not have been implemented at all (e.g., packed bubbles and word clouds). As with all of the products, its interface has grown less friendly per release as the product has become more complex.

- Advizor Analyst from Advizor Solutions similarly began as the work of an academic researcher, but it became disconnected from its roots long ago. Despite good basic functionality, its interface, graphs, and functionality exhibit a path that veered from its expert beginnings.

- SAS JMP has been around longer than the others and has always been a favorite among statisticians. As you would expect from SAS, it is rich in statistical functionality, but its primary shortcoming is a clumsy interface. Sufficient focus on good interface design has always been lacking, resulting in a fragmented experience of disconnected views and a system of navigation that only a statistician could endure. A part of the product, named Graph Builder, has improved the interface in recent years, bringing multiple graphs and interactions together into an integrated experience, but this improvement remains segregated from the product as a whole.

I would love to see these vendors commit to the improvements that are needed, but their current paths make this unlikely, at least in the near term. I would also love to see any of the products among the three potential candidates that I listed above exhibit a commitment to excellence as they mature, but hoping for this is probably naive. I would also welcome a newcomer with passion and clarity of vision, but I've heard no trustworthy whispers of this possibility.

Software vendors don't seem to be talking to data analysts when determining how to improve their products. By "data analysts," I referring to people who actually spend most of their time making sense of data and have acquired the skills over time that are necessary to do so productively. Most of the people who use data analysis tools are not skilled data analysts. They haven't been trained. They have little or no knowledge of statistics. They know little about the data. They may know how to use a tool, but they don't understand the principles and practices that are necessary to use it effectively. Most of them spend their days producing reports in response to requests with little knowledge of how those reports will be used.

The four viable EDA tools that are available today fall far short of what's needed for some or all of the following reasons:

1. They are out of touch with the needs of data analysts. Typically, they are focused more on the customers who yell the loudest or have the biggest wallets.
2. They are hard-driven by the demands of a product release schedule rather than a commitment to quality.
3. They are herding cats (large, disorganized development teams) rather than tracking to a clear and coherent vision.
4. They implement functionality in the ways that best fit their existing architecture rather than in ways that work best.
5. They follow the whims of managers (including those in sales and marketing) and software developers rather than the advice of seasoned experts.
6. They have conflicting interests (e.g., how well their stock is doing rather than how well their product is performing).

Here are a few guidelines that providers of EDA tools could follow to provide useful and effective tools:

1. Commit to making products that will provide maximum benefit for data analysts above all else, rather than focusing on revenues, profits, the value of their stock, and the size of their bonuses.
2. Select and design needed functionality based on interaction with skilled data analysts.
3. Benefit from the advice of experts when selecting which functionality to provide and when designing that functionality.
4. Say "No" to silly feature requests, even when they come from big customers.
5. Strive for *effective* functionality above *innovative* functionality. New features or new ways of designing existing features are rarely better than simple, time-proven methods. Innovation is certainly needed to improve data sensemaking, but not for its own sake.
6. Never release new functionality until it actually does what it's intended to do without compromise. Never let the limitations of the existing architecture lead them to make such a compromise.
7. Staff their design and development teams, in part, with people who understand data analysis and

also understand the humans who work as data analysts. This includes an understanding of human perception and cognition, which is required to determine what tools should do (and not do) and how to do it in a way that works for humans.

Is this asking too much? I've never encountered a software vendor that follows most of these guidelines; not for long, anyway. A few have started with these intentions but all have to some degree set them aside as they faced the demands and temptations of business. It would certainly take courage and skilled management to stay committed to these guidelines, but I'm convinced that the vendor that manages to do this will eventually rule the roost. This vendor would have my gratitude and that of all data analysts. It would be well deserved.

Discuss this Article

Share your thoughts about this article by visiting the <u>What Do Data Analysts Most Need from Their Tools?</u> thread in our discussion forum.

About the Author

Stephen Few has worked for nearly 30 years as an IT innovator, consultant, and teacher. Today, as Principal of the consultancy Perceptual Edge, Stephen focuses on data visualization for analyzing and communicating quantitative business information. He provides training and consulting services, writes the quarterly *Visual Business Intelligence Newsletter*, and speaks frequently at conferences. He is the author of three books: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Second Edition, *Information Dashboard Design: Displaying Data for at-a-Glance Monitoring*, Second Edition, and *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. You can learn more about Stephen's work and access an entire <u>library</u> of articles at <u>www.perceptualedge.com</u>. Between articles, you can read Stephen's thoughts on the industry in his <u>blog</u>.