tableau® SOFTWARE

visual analysis for everyone

AUTHORS:

Pat Hanrahan
Chris Stolte
Jock Mackinlay

DATE:

January, 2007

UNDERSTANDING DATA EXPLORATION AND VISUALIZATION

Although visual interfaces and databases are two of the success stories of the computer revolution, their synergy to date has been modest, probably because visual interfaces have focused on human capabilities while databases have focused on efficient query processing. Unlocking this synergy will shift us from data workers to data thinkers, people who can use data at the speed-of-thought. Three problems must be addressed:

1. ineffective information presentations,
2. poor exploratory capabilities, and
3. difficult user interfaces.

This whitepaper describes how Tableau Software solves these problems with a declarative visual query language (VizQL) that simultaneously describes how to query data and how to present it visually. The result is a fundamentally new way of interacting with databases and spreadsheets. Software products based on VizQL can be used for tasks that range from everyday question-answering that require small data collections and simple visual analysis to mission-critical tasks that require very large data collections and deep multi-dimensional visual analysis.

ABOUT THE AUTHORS

Pat Hanrahan, CTO
Pat is Tableau Software's CTO and also the CANON USA Professor of Computer Science and Electrical Engineering at Stanford University. Pat's research has included visualization, image synthesis, and graphics systems and architectures. He was a founding employee of Pixar (NASDAQ: PIXR), where he was the chief architect of the RenderMan™ Interface - a protocol that has revolutionized the modern graphics and entertainment industries. Pat is the winner of two Academy Awards, the Spirit of America Creativity Award, the SIGGRAPH Computer Graphics Lifetime Achievement Award and an IEEE Lifetime Achievement Award.

Chris Stolte, VP, Engineering
At Stanford University, Chris researched the analysis and exploration of large, complex databases. This research resulted in a dozen landmark research publications and two large-scale visualization systems. Chris was also CTO and co-founder of BeeLine Systems, a visualization software company that developed a revolutionary map rendering system (now owned by Microsoft). Chris is a co-inventor on five software patents related to information visualization. He holds a Ph.D. in Computer Science from Stanford University, and a B.S. in Computer Science from Simon Fraser University.

Jock Mackinlay, Director of Visual Analysis
Jock is Tableau's Director of Visual Analysis. Jock has a Ph.D. in computer science from Stanford University, where he pioneered the automatic design of graphical presentations of relational information. He joined Xerox PARC in 1986, where he collaborated with the User Interface Research Group to develop many novel applications of computer graphics for information access, coining the term "Information Visualization." Much of the fruits of this research can be seen in his book, "Readings in Information Visualization: Using Vision to Think." Jock is a co-inventor of numerous software patents related to information visualization.

INTRODUCTION

The need for visual analysis

Although visual interfaces and databases are two success stories of the computer revolution, their synergy to date has been modest. Popular data analysis and reporting applications, for instance, suffer from three problems:

Ineffective Information Presentations. Trying to answer a question by working with a grid of numbers is difficult. The charting wizards contained in most applications, including spreadsheets, do not provide a satisfactory alternative. The visual presentations in such interfaces are not good at revealing answers to deep questions, are restricted to small data, and often distort the truth. What is needed is a high performance software interface for effectively visualizing meaningful multi-dimensional questions.

Lack of Exploratory Capabilities. Most applications claim to contain ad hoc analysis capabilities. Yet they are typically based on analysis models employing multi-step wizards and templates that get in the way of the thought process. Some of them require IT departments to pre-define the ad hoc queries people are likely to ask, a self-contradictory proposition. What is needed is a software application for free-form exploratory analysis.

Difficult User Interfaces. The most common customer complaint about analysis applications today is that they are too hard to use. Introductory training courses for products from SAP, SAS, Business Objects, Crystal and Cognos require the greater part of a week. Spreadsheets become unwieldy for analysis as tables get large. What is needed is an easy analysis interface for knowledge workers.

This whitepaper describes how Tableau Software solves these problems with a declarative visual query language (VizQL) that ushers in a new analysis model called visual analysis. The result is a fundamentally new way of interacting with databases and spreadsheets.

VISUAL ANALYSIS

A description of visual analysis

Tableau Software was founded on the idea that analysis and visualization should not be isolated activities but must be synergistically integrated into a visual analysis process. Visual analysis means specifically:

Data Exploration. Visual analysis is designed to support analytical reasoning. The goal of visual analysis is to answer important questions using data and facts. In order to support analysis, it is not enough to only access and report on the data. Analysis requires computational support throughout the process. Typical steps in analysis include such operations as (1) filtering to focus on items of interest, (2) sorting to rank and prioritize, (3) grouping and aggregating to summarize, and (4) creating on-the-fly calculations to express numbers in useful ways. A visual analysis application exposes these exploratory operations to ordinary people through easy-to-use interfaces.

Data Visualization. Visual analysis means presenting information in ways that support visual thinking. Data is displayed using the best practices of information visualization. The right presentation makes it easy to organize and understand the information. For example, critical information may be quickly found, and features, trends, and outliers may be easily recognized. One powerful way to evaluate any analysis tool is to test its effectiveness in answering specific questions. At the most fundamental level, does the tool have the analytical power needed to answer the question? At another level, how long does it take to answer the question?

A successful visual analysis application unites data exploration and data visualization in an easy-to-use application that anyone can use.

WHAT IS VISUAL ANALYSIS?

Visual analysis means exploring data visually. A story unfolds as you navigate from one visual summary into another.

A visual analysis experience has certain characteristics. It allows you to do two things at any moment:

- Instantly change what data you are looking at (1-click). This is important because different questions require different data.
- Instantly change the way you are looking at it (1-click). This is important because different views of data answer different questions.

The combination creates the exploratory experience required for ordinary people to answer questions quickly.

The technical architecture of a visual analysis system flows directly from the requirements.

- The visualization engine needs to be fused to a query engine.
- The visualization engine must be capable of generating an infinite class of picture expressions.

Question	Action	Discovery
What's a fair price for a condo in Florida?	Begin by summarizing all sales (1-click).	A view that organizes the data. Commonly called a Pivot Table or cross-tab.
But what are the general trends?	With 1 click, convert the summary to a visual analysis.	The average price of a condo has increased significantly this year (2005), even though the total number of transactions fell.
Does condo complex affect this analysis?	Drag the Condo Complex field onto the worksheet. Turn on reference lines.	Harbor Landing prices have been lower than those of other condo complexes. In fact, a $300K price appears expensive for Harbor Landing.
What factors determine the value of specific condos?	Drill into the 2005 records (double-click), and pivot the view to show Square Feet vs. Price (drag and drop).	Square footage is clearly a strong determinant of value. But the jagged line suggests other factors are also at work.
What factors besides Square Footage matter?	Break the data into individual records (1 click). Then drop the Bedrooms field onto the worksheet.	For the same square footage, buyers tend to pay more for 3 bedrooms.
What other factors determine value?	Add the Baths field for even more detail. Then drag Condo Complex back onto the worksheet.	Condo prices fall into relatively tight groupings. For example, look at Harbor Landing. A $300K condo in Harbor Landing may be reasonably priced. In fact, based on recent sales, if it has 3 bedrooms and 3 baths and is close to 2000 sq ft, it may be a good deal.

VISUAL ANALYSIS IN PRACTICE: EXAMPLE #1

Let's consider some examples of how visual analysis is used to answer questions.

Most analysis tools allow people to interact with data using tables of numbers. The top summary in Figure 1 is familiar and typical. This data visualization is typically called a cross-tab or Pivot Table. The table shows the breakdown of Sales and Profit by Product Type, Product and Region. The underlying relational database consists of all sales transactions in the organization. The cross-tab summarizes the data by computing the total sales for all rows in the database, broken down by the dimensions shown.

Try to answer some meaningful questions using the cross-tab view at the top of the Figure. Which product-region pair is the best seller? The question can be answered with the cross-tab, but it takes time. In fact, you must sequentially scan all the numbers in the table looking for the maximum. And that scan is error prone because you must remember the bestseller as you scan the table. If you lose track of the current maximum value, you need to restart the scan. Now look at the bar graph at the bottom of the Figure. The best selling product immediately pops out: Columbian Coffee in the East, the tallest bar. This visualization makes it easier to answer the question because it presents the information in a better way.

Now answer the question: Which product-region pairs have high sales but lose money? Again, the task can be accomplished using the cross-tab but it is frustratingly slow. In contrast, the bar chart shows profit using color, while the height of each bar represents the sales total. The money losers (red) are easy to pick-out. Amongst those, it is easy to see which ones have the highest sales. For example, Decaf Irish Cream in the West sells quite well, but loses money.

Why is the second visualization better for answering these particular questions? Because it uses the power of our perceptual system. Psychology experiments have shown that people can immediately find the tallest bar amongst a set of bars of varying height. The length of a bar is a pre-attentive visual cue. There is no equivalent visual cue in the table. Our visual system performs rapid search by processing the image in parallel. The parallelism means that the time it takes to find the largest bar is independent of the number of bars in the set. In computer science terms, the search takes constant

time. In contrast, the time needed to search the table depends on the number of cells. It takes linear time. As the number of items increases, the efficiency of the bar chart vs. the table increases arbitrarily. This simple example demonstrates the power of visualization to save time and reduce errors in analytical reasoning.

For the first question we posed ("What is the best selling product in a region?"), the bar chart is better than the cross-tab. However, there are analytical tasks where cross-tabs are better than charts. For example, cross-tabs provide a powerful indexing mechanism that can be used to speed up search. Suppose we change the question to "What is the precise Sales Total for Columbian Coffee in the East?" This question is hard to answer precisely with a bar chart, but it is easy with a cross-tab. Why is that? First, the row and column headers in a table efficiently organize the information so that it is easy to look up values. Second, it is hard to estimate the length of a bar accurately. Our visual acuity is limited and visual estimates of length are only accurate to a couple of decimal places. In contrast, the number in the table is shown at a higher precision.
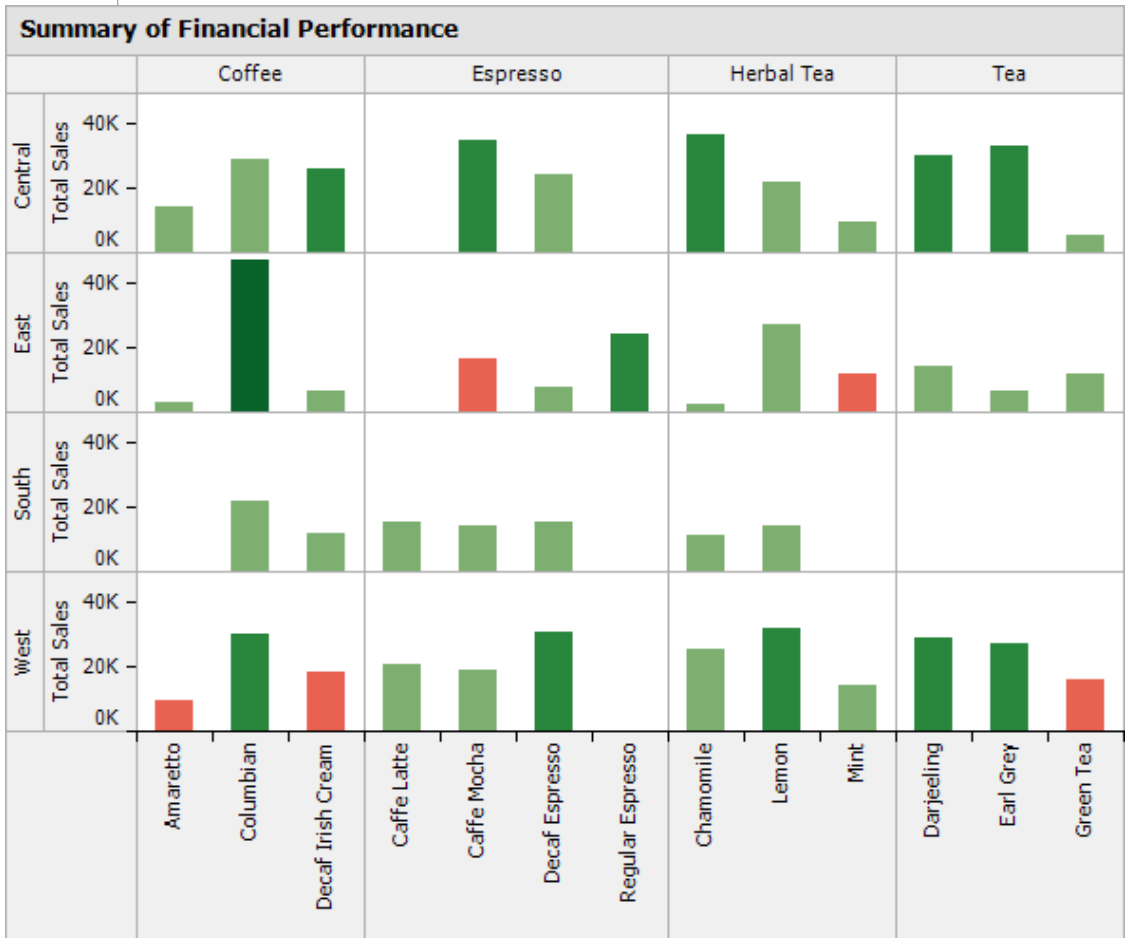
This example is meant to show the power of visual thinking, and its most important principle: For a given question or analytical task, some visual representations work better than others. Visual thinking involves choosing the right visual representation for your task.

FIGURE I: Visualizing Information

Two visualizations of financial data, both created with one click using Tableau. Each is useful for different tasks. The top view is useful for looking up the precise performance of a specific product. The bottom view is useful for examining relative magnitudes, relationships and anomalies. For instance, the bottom view clearly shows the areas of the business resulting in negative profit (red bars).

Summary of Financial Performance
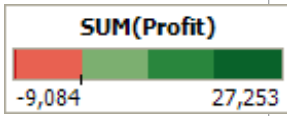
		Central		East		South		West	
		Total Sales	Total Profit	Total Sales	Total Profit	Total Sales	Total Profit	Total Sales	Total Profit
Coffee	Amaretto	$14,011	5,105	$2,993	1,009			$9,265	-1,225
	Columbian	$28,913	8,528	$47,386	27,253	$21,664	8,767	$30,357	11,253
	Decaf Irish Cream	$26,155	9,632	$6,261	2,727	$11,592	2,933	$18,235	-1,305
Espresso	Caffe Latte					$15,442	3,872	$20,458	7,502
	Caffe Mocha	$35,218	14,640	$16,646	-6,230	$14,163	5,201	$18,876	4,064
	Decaf Espresso	$24,485	8,860	$7,722	2,410	$15,384	5,930	$30,578	12,302
	Regular Espresso			$24,036	10,062				
Herbal Tea	Chamomile	$36,570	14,434	$2,194	765	$11,186	3,180	$25,632	8,852
	Lemon	$21,978	6,251	$27,176	7,901	$14,497	2,593	$32,274	13,120
	Mint	$9,337	4,069	$11,992	-2,242			$14,380	4,330
Tea	Darjeeling	$30,289	10,772	$14,096	6,497			$28,769	11,780
	Earl Grey	$32,881	10,331	$6,505	3,405			$27,387	10,425
	Green Tea	$5,211	1,227	$11,571	5,654			$16,063	-7,109

Summary of Financial Performance

VISUAL ANALYSIS IN PRACTICE: EXAMPLE #2

Figure 2 shows another example involving two different visualizations of the same data.

Each row of both visualizations corresponds to a different market segment. Each line represents sales for a different product on different weekdays in different years. The top view shows weekdays nested within years. The bottom visualization flips the order of these two variables. Suppose we want to compare sales performance of products over time. Which visualization is better at helping us do this?

The answer depends on the precise question. If we need to determine the weekly selling pattern of products, the top view is very useful. One immediate discovery, for instance, is that sales of Technology products to the Corporate segment spike on Tuesdays, and then fall through the rest of the week, before spiking again on Saturdays. But what if you needed to know if Saturday sales in that segment were improving from year to year? For instance, there might be a marketing campaign in place to drive Saturday sales higher. Is it working? This question is extremely difficult to answer in the first visualization. The second visualization, however, makes the trend abundantly clear: Saturday sales in this segment have been falling every year.

Comparisons are easier when the items to be compared are adjacent. It is more difficult to compare items at a distance. Our visual system is foveated, which means that our eye's processing power is concentrated at the center of the visual field where the resolution is highest. Even though we monitor our entire field of view, we focus our attention in the direction we are looking. For this reason, nearby items are seen all at once. In order to compare items at a distance, we need to move our eyes back and forth between them. Thus, when answering a question involving comparison, it is important to order the values next to each other to facilitate comparison. Changing the visualization to show the data in the right order makes the answer to certain questions obvious. A different question might require a different order. Hence it is imperative that a visual analysis system allows people to easily change the visualization of a dataset at any moment.

FIGURE II: Alternate Visualizations of the Same Database Query
Alternative visualizations of the same data. A visual analysis system must allow people to quickly switch between visualizations, because different views answer different questions. For instance, the upper view clearly answers the question, "What is the peak sales weekday for each product in each segment in each year?" Only the bottom view, however, clearly answers the question, "Year to year, have Saturday sales been improving or worsening for each product in each segment?" In Tableau, users can toggle between these visualizations with one click.
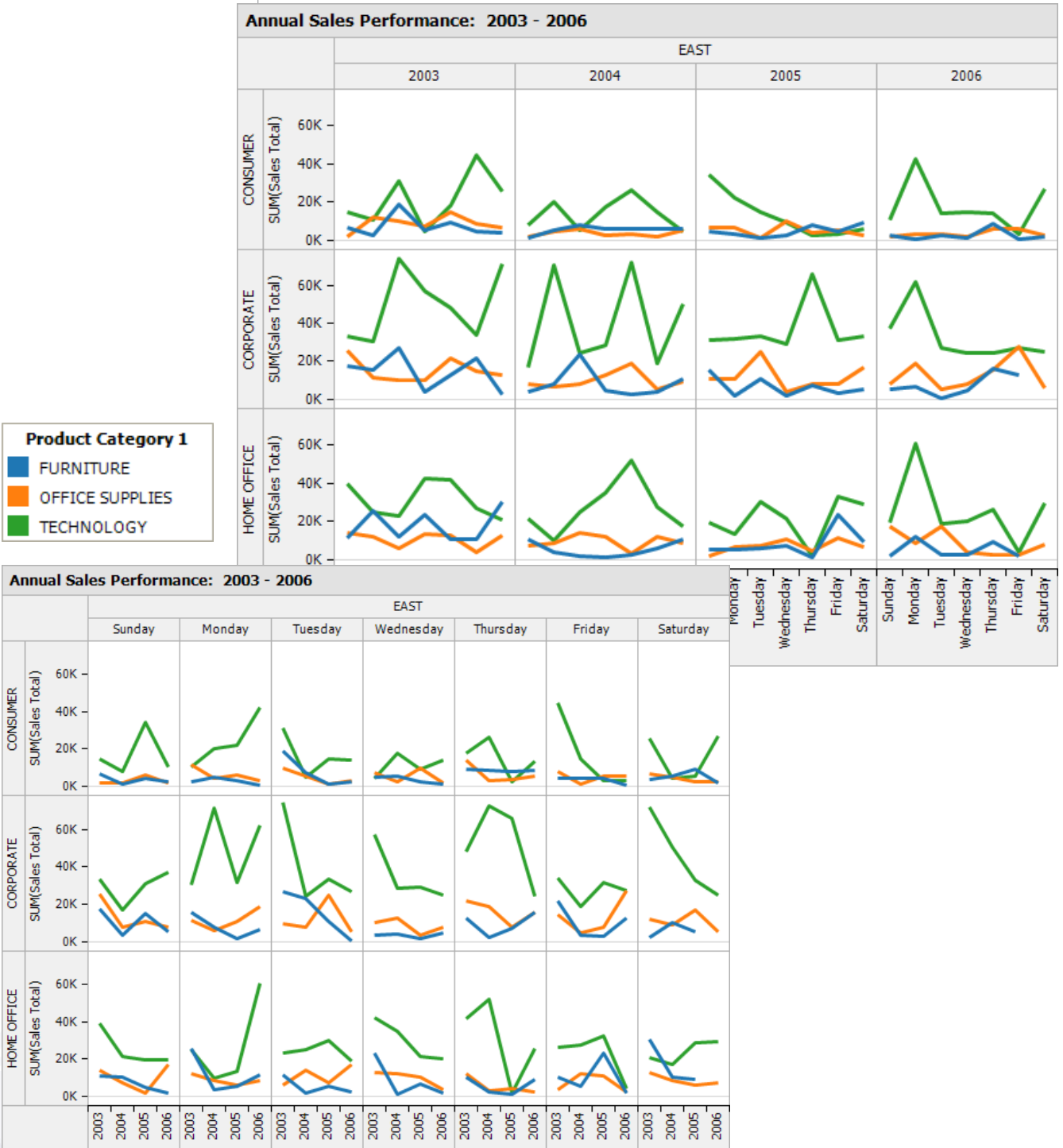
VISUAL ANALYSIS IN PRACTICE: EXAMPLE #3

Since using the right representation is so important to visual analysis, we will show one more example. Shown in Figure 3 is a study of the relationship between Florida condo prices and their attributes.

Each symbol corresponds to a sale of a condo unit. The symbols are colored by the number of bedrooms, with shape depending on the number of bathrooms. This example shows our perceptual power to form groups and organize information. Note that it is easy to form groups like the two three-bathroom units in Harbor Lakes (squares), which command much higher prices than the two-bathroom unit with the same number of bedrooms and square footage. Our visual system readily groups items using color and shape. Mentally we draw a lasso around items with similar visual attributes, making the analysis of physical condo attributes easier. We are thinking visually.

Once data is organized into groups, new findings are evident and new questions jump out. For instance, all condo complexes in the view experience sales across a similar price range ($200-350K), but each has a slightly different median. We can also immediately cluster items into subsets with similar trends. For instance, we can see that three-bathroom units almost always sell at or above the median for the condo complex. With this visualization, we can see and begin to understand the complex dynamics underlying our business.

FIGURE III: Thinking with your Visual System
A visualization of Florida real estate transactions. Mentally our perceptual system draws a lasso around items with similar visual attributes (in this case making analysis of physical condo attributes easier). We can immediately cluster items into subsets with similar trends. For instance, we can see that three-bathroom units almost always sell at or above the median for the condo complex. An effective visualization helps us to think visually.
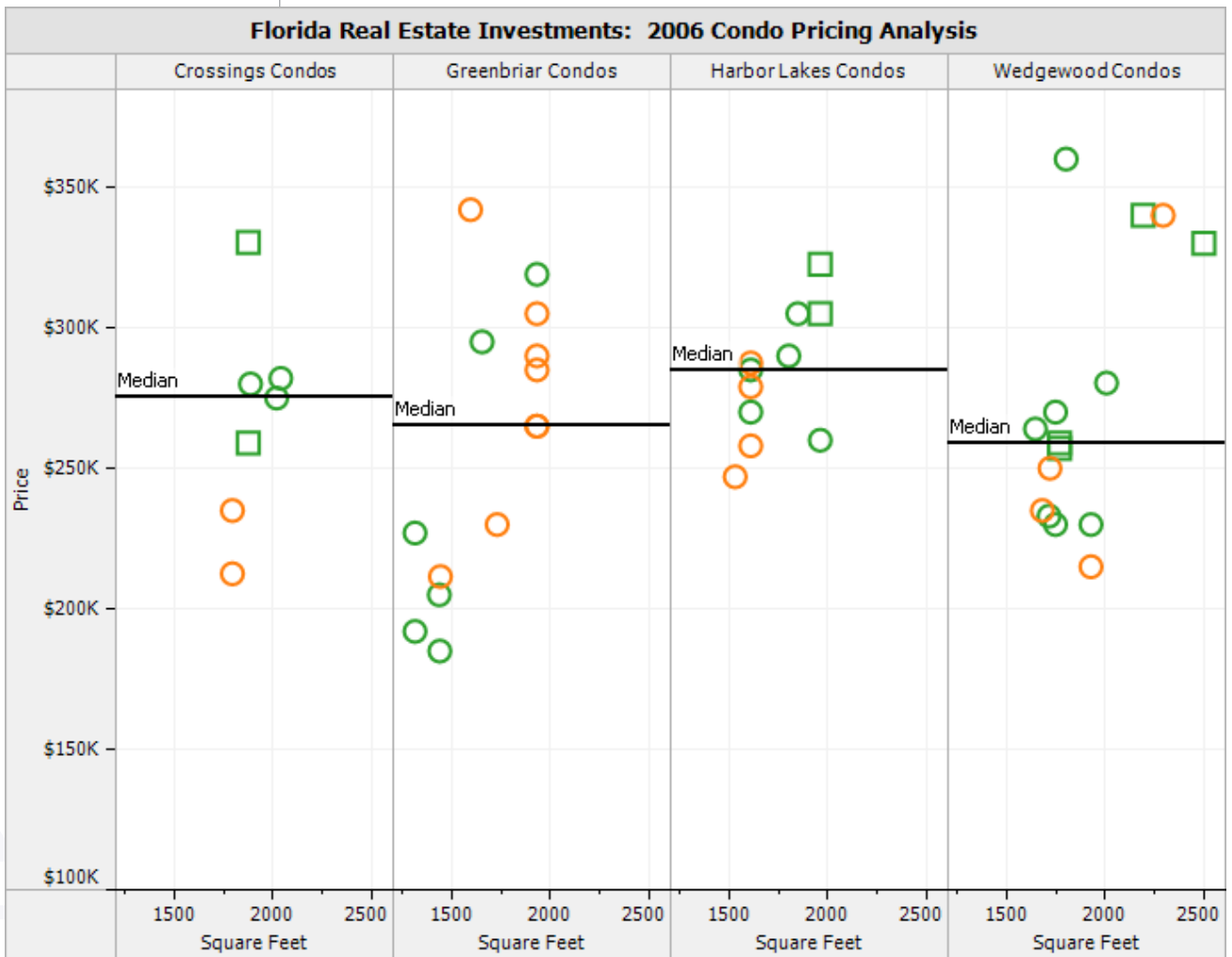
INTRODUCING TABLEAU

A visual analysis product suite

Tableau builds visual analysis products that help people ask and answer analytical questions involving data stored in databases and spreadsheets. Tableau enables people to better analyze data using their natural ability to think visually. How specifically does Tableau accomplish this?

The answer involves advancements in technology and careful product design. Tableau's development effort began when the US Defense Department challenged computer graphics inventors to collaborate with leading database researchers at Stanford University. Tableau's product line is devoted to combining graphics, databases and analysis into a unified visual analysis framework. This framework is based on five principles.

Principle #1: Easy Interface

Analyzing structured data requires subtle distinctions but queries are hard to author. They are cumbersome to compose using raw query syntax. Menu choices, multi-step query generation wizards and report design frameworks don't make the analysis experience much easier. Typing formulas into a spreadsheet offers little relief, particularly when data changes frequently.

A visual analysis product must have a carefully designed user interface that makes query generation easy. What if you could analyze information by simply creating the visualization that presents the answer to your question? If such an interface could be invented, it might bring exploratory analysis to a new class of people.

Figure 4 shows Tableau's user interface. A Tableau worksheet starts as a blank canvas. By dragging fields onto a worksheet, users visually analyze data sources and produce instant visual results. The act of dragging a field onto a worksheet both (1) creates a visualization of the data, and (2) generates queries to the data source to fetch the appropriate records.

Let's take a closer look at the interface. On the left is the data viewer, which shows the fields in the database organized into categorical dimensions and numerical measures. Users can opt to interact with all of the fields in a data source, or just a subset of them. In the main work area there are several shelves. There is one for Rows, another

for Columns, etc. Dragging a field (e.g. year) onto the row shelf causes a row to be created for each value in the set of values comprising that field. For example, if you drag Months onto the row shelf, twelve rows are created, one for each month of the year. The contents of these shelves determine the structure of the visualization.

Inspecting the Tableau interface you see a blend of analysis and visualization. Shelves exist for analytical tasks such as filtering as well as visual tasks such as the laying out rows and columns, and mapping data values to color and size. For example, if you add a category to color, it both changes the level of detail of the summary and the mapping of the category to color. Analysis and visualization are intertwined, not separated.

FIGURE IV: Tableau's User Interface

Tableau's user interface is based on drag and drop analysis. Here, a user examines a database of Seattle 911 calls. She is able to generate an informative visualization of the data in seconds.
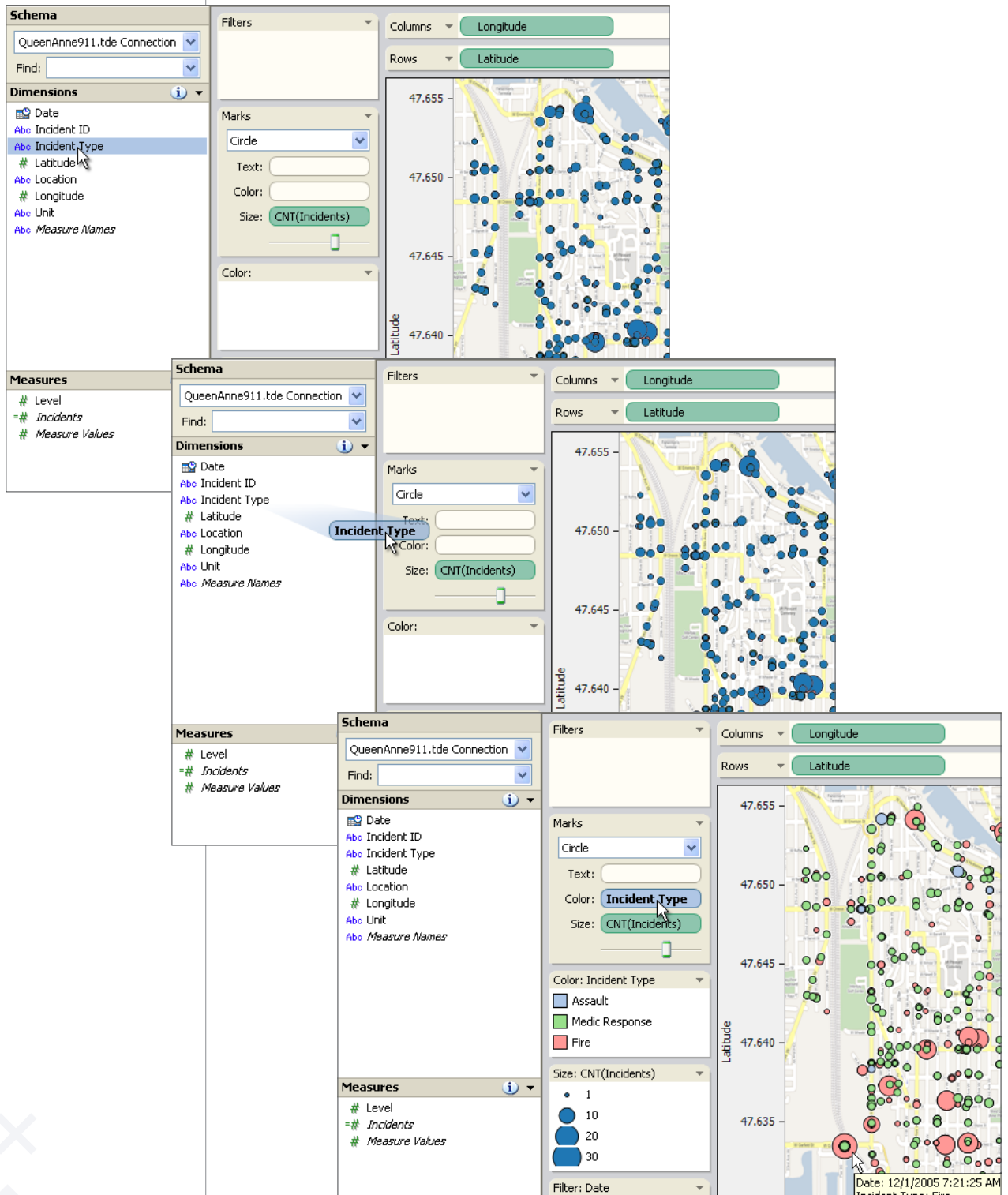
Principle #2: Data Exploration

Tableau helps you think visually. It's not just a system for creating high quality visualizations. It is an interactive visual analysis system for answering questions. Every time you compose a picture, Tableau automatically composes the queries and analytical computations needed to create the picture.

Figure 5 summarizes the analysis process. This cycle of forage, experiment, discover and act is typical of problem solving. We call this data exploration. It is essentially a visual cycle of analysis that results in an answer. Tableau is designed to support analysis of this variety. It does this in several specific ways.

First, posing a question and getting the answer involves only a few actions. Visual thinkers quickly learn to compose pictures that answer their specific questions. Cumbersome user interfaces derail your train of thought by asking you to interact with templates, wizards, forms and modules. The naturalness of the Tableau interface keeps you focused on the problem at hand at all times. For the first time, an analysis session flows from a stream of consciousness. Figure 6, for instance, shows some of the computational functions (e.g., create set) that Tableau users can perform by interacting directly with a visualization.

Second, Tableau visualizations often show the unexpected. A surprise finding stimulates the thought process, encouraging deeper analysis or a different path of exploration. Tableau takes great care to present users with the most effective visualization of every query.
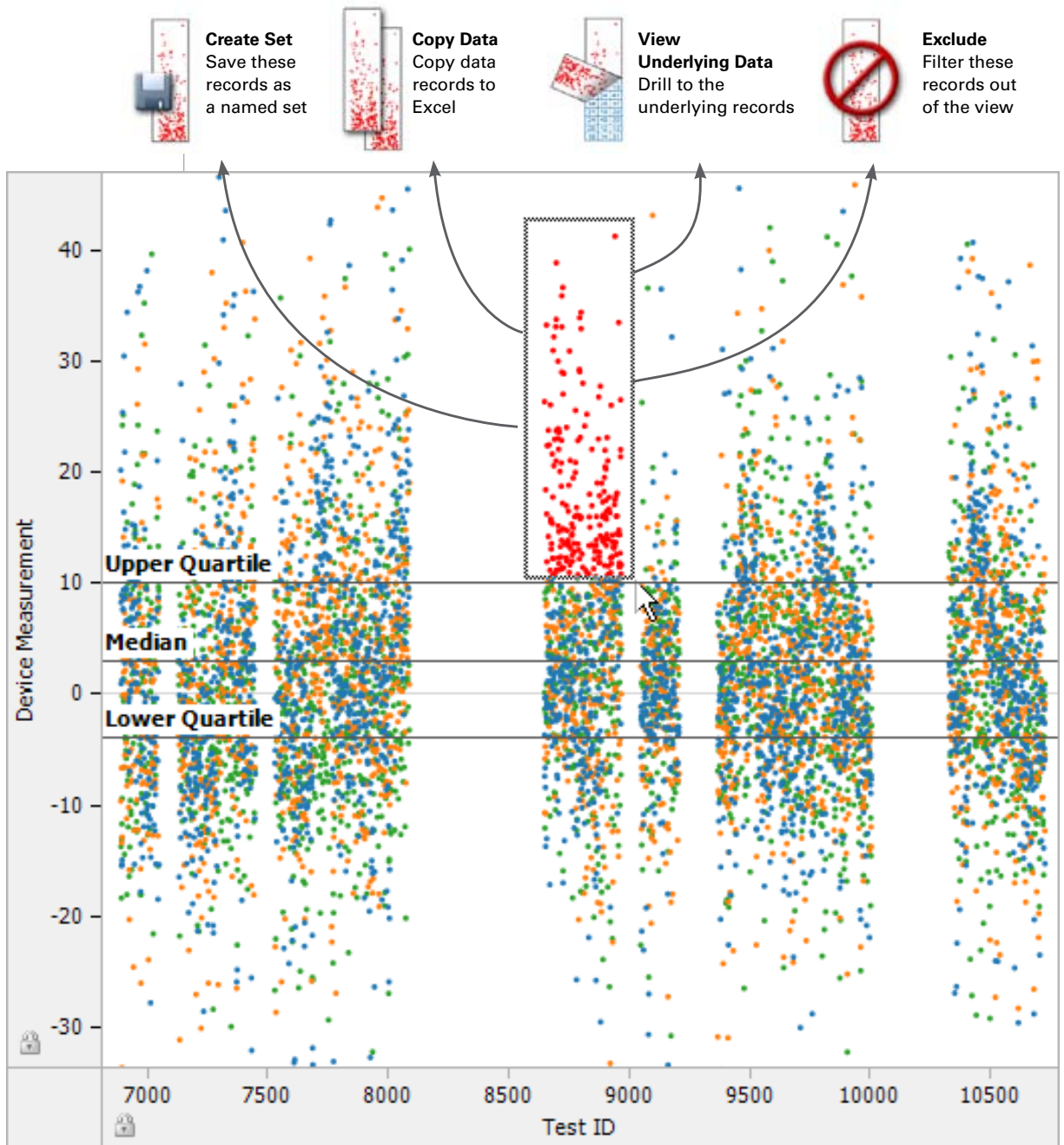
FIGURE V: Data Exploration
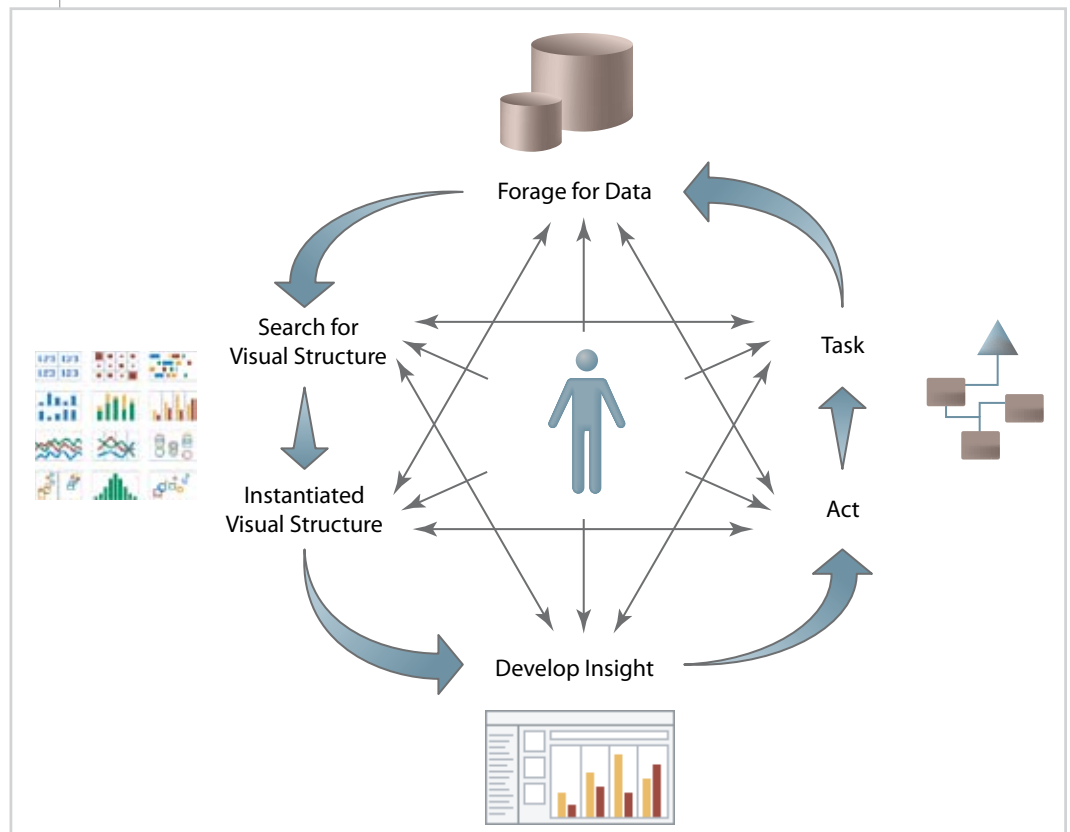
Third, Tableau's interaction model is based on the concept of incremental change: Whenever you perform an action (e.g., filter), Tableau instantly shows you the new result. Why is incremental change important? It lets us intuitively explore a broad space of possible visualizations to find the right one. It allows us to focus on the exploration task, where questions lead not just to answers but also to more questions. It also lets us learn visual literacy at our own pace. We can build sophisticated pictures of data slowly and incrementally as we learn how to look at information. In essence, Tableau's interface is based on the process of incrementally refining a question into an answer. It's a satisfying system to use.

Finally, Tableau records the analysis process as it proceeds. It is always possible to undo, redo or return to a previous state. Indeed, Tableau is based on a global undo architecture that makes every action safe, because it can be instantly undone. It is also possible to create bookmarks signifying important findings or branch points in the analysis. The key point is that exploration is easy and hence encouraged. Typically an analysis does not involve a single finding, but a collection of findings. In decision making scenarios, the analysis consists of an argument for a certain course of action. This argument is based on evidence and findings that are assembled into a coherent structure. Tableau makes it easy to generate such lines of thought.

FIGURE VI: Exploring Engineering Data With Tableau

A visual summary of a million test results. The horizontal axis shows consecutive
test numbers while the vertical axis plots a device measurement. Consecutive
test runs form visual clusters. Reference lines are added to designate important
statistical thresholds. In Tableau, users can lasso the data points of interest (this
selects the underlying records) and take action.

Create Set
Save these
records as
a named set

Copy Data
Copy data
records to
Excel

**View
Underlying Data**
Drill to the
underlying records

Exclude
Filter these
records out
of the view

Principle #3: Expressiveness

An open set of questions cannot be addressed by a fixed set of visualization types. Every answer leads to new questions. What is needed is an analysis interface that supports an infinite class of visualizations.

Most analysis applications rely on templates, wizards, widgets and their associated chart types to provide visualizations of data. This is true of Excel, but it's also true of enterprise business intelligence applications. Tableau is based on a fundamentally different approach: Tableau's is based on a declarative visual query language (VizQL™) that is infinitely expressible and composable. This is the technology at the heart of Tableau's product suite.

Each visual analysis in Figure 7, as an example, can be described by a single VizQL™ expression.

These examples represent a tiny subset of VizQL's expressive capabilities. None of the views above is based on a template populated with data, rather each is described by a simple VizQL statement that Tableau's interpreter translates into an interactive information visualization.

What is VizQL? It's like SQL. The development of SQL led to a revolution in database technology. Suddenly, complex queries were easily described as readable text. Programmers were freed from writing custom procedural programs in order to query databases and instead could compose simple statements. VizQL marks a similar step forward. What SQL did for database interactions in text form, VizQL does for database interactions in visual form. Sophisticated visualizations are described by declarative VizQL statements. The standard approach when developing analysis systems is to provide a set of canned objects — bar chart objects, scatter-plot objects, etc. This legacy approach is similar to selling a database with six pre-approved queries rather than with a complete query language like SQL. With VizQL, people have a single visual analysis interface to generate a broad range of visualizations. The language based approach supports the expressive flexibility required by different questions.

FIGURE VII: Visual Language ≠ Visual Widgets

What SQL did for data interactions in text form, VizQL does for data interactions in visual form. VizQL statements describe an infinite class of sophisticated multidimensional visualizations. Users can transform any database or spreadsheet into information visualizations tailored to the data.
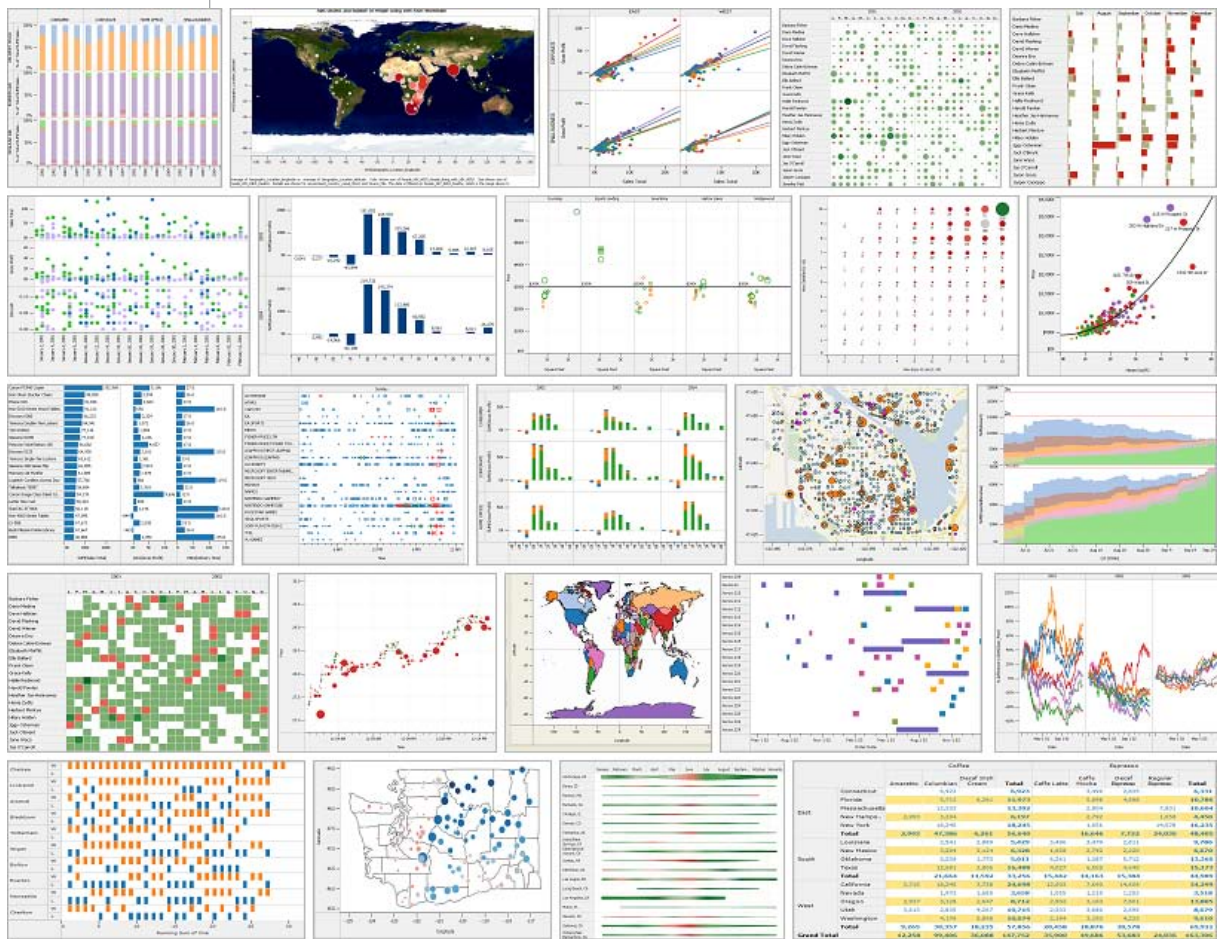
Principle #4: Visualization Best Practices

When information is presented effectively, new findings become evident and new questions jump out. An effective visual analysis system needs to encourage the creation of effective visualizations.

Many software packages are based on exotic visual metaphors of questionable utility (e.g., tree maps, bubbles and 3-D graphics). They are based on the belief that in order to advance decision

making, it is necessary to invent elaborate and complex graphical summaries. Tableau believes that instead people need a rich set of proven visualizations that grows as their needs and visual literacy changes. Conventional types of business graphics such as tables, bars, maps and time series are known to be effective. These visual representations have been optimized for over a hundred years, so Tableau emphasizes them. Moreover, they are understood by everyone. Although exotic visual metaphors may be useful in some situations, they are seldom needed. Tableau uses the best practices in information visualization to display data effectively.

Tableau's approach is to encode the best practices into graphic design principles. Design rules are used to choose defaults intelligently. Any visualization may be created, but the ones conforming to best practices are the ones that emerge naturally from the product.

Figure 8 shows an example of Tableau's approach to the hard problem of understanding multidimensional data. Here, an operations analyst needs to understand the relationships between five independent business dimensions: Year, Market Segment, Ship Method, Ship Container and Sales Total.

In this case, Tableau creates a table of common size bars. The example shows several examples of best practices:

- Tables of graphs are used to compare across more dimensions. The visualization expert Edward Tufte calls this technique small multiples. Small multiples make comparison easy and can be used with many different types of visualizations. Tableau makes it possible to replicate any visualization for comparison. You can make a table of line graphs, a table scatter plots, a table of maps, etc.

- Color and shape are used effectively. Distinctive colors are chosen so that different classes are easy to distinguish. Data is shown using vivid, but not garish, colors so that information is highlighted and easy to understand.

- The clean visual design of the Tableau interface is designed to emphasize your data, not the interface. The interface is in the background and the data is in the foreground. The display is also uncluttered, so that it is easy to focus on the information presentation.

FIGURE 8: Using Best Practices in Information Visualization
Analysis of an operations problem involving five dimensions: Year, Market Segment, Ship Method, Ship Container and Sales Total. Tableau supports best practices in information visualization that make the process of producing useful visualizations easy. In this case, the table of small multiples facilitates comparisons. Colors, though customizable, are automatically chosen to be clear and distinct. The interface is clean and uncluttered. The result is that insights emerge. For instance, every year the small business segment has become increasingly reliant on Jumbo Drums to fulfill their orders.
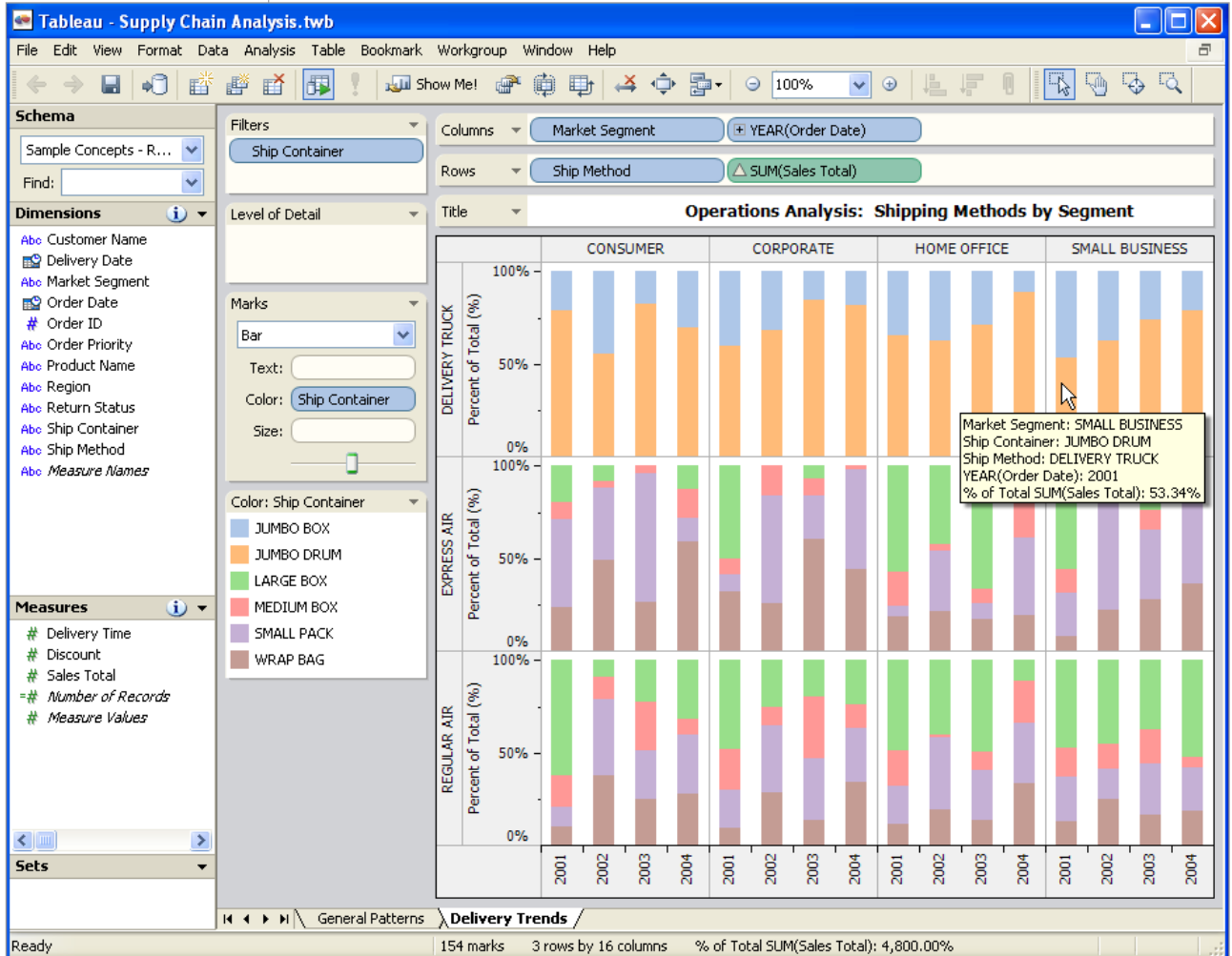
Principle #5: Database Independence
Database independence is an important attribute of Tableau's visual analysis products. Even though different types of databases have different capabilities, Tableau's interface is the same regardless of how the data is stored; there is no need to learn a different system if you change the database technology. Different workgroups using different databases may use the same front-end, saving on training costs and encouraging people to learn from their colleagues.

As your needs grow, you can upgrade to more powerful database engines and Tableau will still work as before. In fact it is possible to retarget existing Tableau workbooks to different database technologies. Figure 9 show Tableau connected to Oracle, SQL Server Analysis Services and Excel simultaneously.

Out of the box, Tableau supports relational databases, OLAP cubes, spreadsheets and files. No special configuration is required. Tableau also supports ad hoc relational database joins, from simple star-schemas to complex referential structures. These joins can be defined with a few clicks and in a manner that is easily understood by knowledge workers without technical skills. There is no limit to the number of databases and spreadsheets that can be analyzed simultaneously.

Tableau can be used to perform real-time analysis against large quantities of data and has no preset data scalability limits. Customers have used Tableau as a visual analysis interface to databases containing 100s of millions of records.

FIGURE IX: Operational Dashboard Using Multiple Databases
A visual summary of category management data based on multiple source systems. This example is based on data in Oracle tables, SQL Server Analysis Services cubes and an Excel spreadsheet. Tableau supports the ability to add different analyses to a dashboard arrangement.
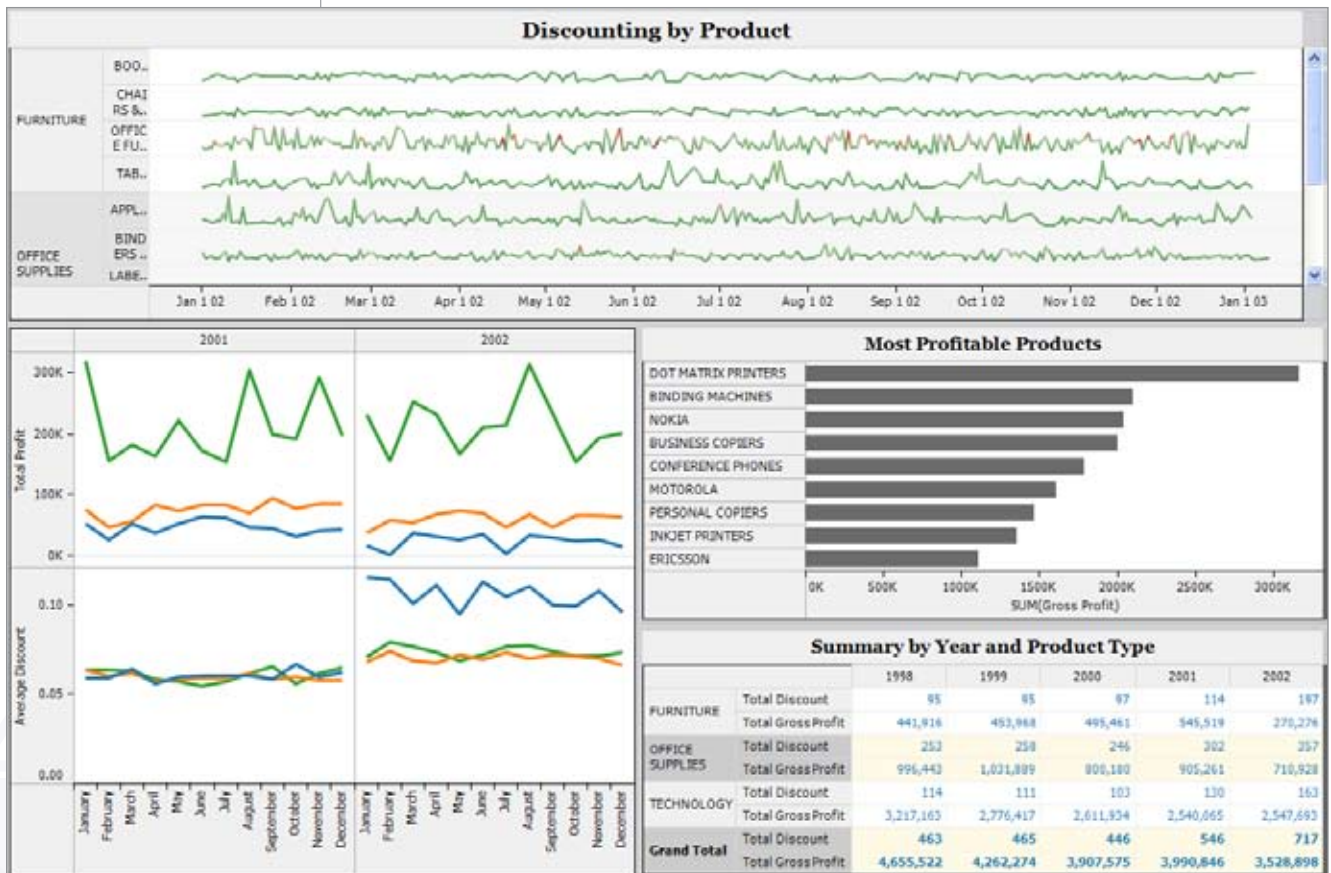
INTRODUCING VIZQL™

A visual query language for data

The five core design principles of Tableau's visual analysis products are: Easy User Interfaces, Easy Data Exploration, Expressiveness, Visualization Best Practices and Database Independence. All of these principles are embodied in or enabled by Tableau's VizQL™. VizQL™ is a visual query language for databases. What SQL did for database interactions in text form, VizQL™ does for database interactions in visual form. Sophisticated interactive visualizations can be described by declarative VizQL™ statements.

The language-based approach of VizQL™ facilitates three important goals that underlie Tableau's product design principles:

- **Expressiveness.** People have a single visual analysis interface to generate an infinite class of picture expressions.
- **Incremental Construction.** People can incrementally construct views that answer questions. The language-based approach enables people to start with a simple visualization and then incrementally add data or computations until the view is refined into an answer.
- **Alternatives.** People can easily switch between different visualizations. A declarative language removes the need for new templates and widgets to be populated every time a user needs to change a visualization. Displaying data at just the right angle is crucial to making discoveries.

Inside VizQL™

VizQL™ goes a step beyond the two industry standard database query languages, SQL and MDX. VizQL has the query and analytic capabilities of these query languages. However, it extends them by adding support for creating visualizations of the results returned from the query. If you have used SQL or MDX, you know that they have primitive formatting capabilities. Any result set returned from a query may be printed on a terminal. Each column in the tuple is a column of the table, and the rows are the individual records. Fields may also be formatted using a variety of formatting functions. In MDX, special print versions of the fields may be present (these are sometimes called aliases), and color may even be specified. Although these formatting capabilities are sometimes useful, they cannot generate the visual representations needed to support the analysis process. VizQL™ fills this void.

Like SQL and MDX, VizQL™ contains a WHERE clause for filters, an ORDER BY clause for sorting, a GROUP BY clause for controlling level

of detail and aggregation, etc. VizQL™ is unique in that it contains a rich syntax for creating the row and column structure of a table. VizQL™ also lets you select different types of marks (text, symbols, lines, etc.) and control their visual attributes (color, size, etc.). VizQL has the capabilities needed to create all the common types of visualizations necessary for analysis and reporting applications. Tableau's VizQL™ technology is unique within the industry.

VizQL™ has advantages compared to existing methods for building analysis and reporting tools:

- VizQL™ is a computer language for describing tables, graphs, charts, maps, time series and tables of visualizations. VizQL™ unifies all these different visual representations into a single framework. Unlike conventional component architectures that underlie charting wizards, a visual language permits an unlimited number of picture expressions; every expression is a unique, different and customizable visualization.

- VizQL™ statements define the mapping from records returned from the database to marks (drawing primitives). Some fields in the record control the geometric properties of the mark, including position, size and orientation; other fields control visual attributes like color and pattern.

- VizQL™ is a declarative language. The advantage of a declarative language is that the user describes "what" picture should be created, not "how" to make it. The user need not be aware of the machinery under the hood, the query, analysis and rendering operations that run behind the scenes. Database languages such as SQL (or the relational algebra) are also declarative. The result is a more portable and scalable system.

- The VizQL™ query optimizer compiles into SQL and MDX. SQL provides access to relational databases; MDX to OLAP data cubes. VizQL™ therefore provides a uniform interface for interacting with databases of any size, format or subject domain. By exposing the full power of SQL and MDX, VizQL™ incorporates all the important operations needed for analysis, including selection, filtering, sorting and calculations.

- The VizQL™ interpreter is highly optimized for interactive use, querying, processing, and drawing over 100,000 records per second. VizQL™ is designed to take advantage of modern computer graphics hardware (fast rendering chips developed for gaming that are standard on every personal computer).

FAQ

What type of user is Tableau designed for?
Excel users. Anyone comfortable with Excel can use Tableau.
Tableau is known for bringing ad hoc analysis to a new class of user
in an organization. Our award-winning UI team hails from Stanford
University's HCI department, Xerox PARC and Visio. People become
comfortable with Tableau's user interface in the first use. Tableau
also provides online training.

Is Tableau a desktop application or a web-based application?
Both. Tableau can be used in both desktop and web-based configurations.
Both products can be downloaded and installed in a few minutes.

Can people publish Tableau's interactive visual analyses to others?
Yes. Tableau's web product supports publication and collaboration.
Among other features, it allows people to publish interactive visual
analysis that can be accessed through a web browser.

What are Tableau's performance characteristics?
Tableau has no preset data scalability limits. Out of the box, Tableau
can be used with enormous data warehouses, small spreadsheets,
and everything in between. Large customers use Tableau as a visual
analysis interface to 100s of millions of records. In these configurations,
Tableau pushes calculations to the source database, leveraging the
processing power of database servers without in-memory limitations.
With one click Tableau users can even toggle between in-memory
analysis and live database server analysis.

**What value does Tableau add to environments that employ
dashboards, reports or scorecards?**
Corporate environments populated with business intelligence
packages add Tableau due to its visual analysis capabilities. Visual
analysis is needed due to the shortcoming of business intelligence
tools. Specifically, Tableau provides data exploration, data
visualization and easier ad hoc analysis.

How large is Tableau's customer base?
Tens of thousands of users. Leading customers include Microsoft,
Google, eBay, Wal-Mart, Merck, Hewlett-Packard, Countrywide, Fidelity,
GE, Charles Schwab, Harvard University, and the US Government.
Tableau's products can be used with data of all languages.