UNDERSTANDING THE VALUE OF VISUAL DATA DISCOVERY — A GUIDE TO VISUALIZATIONS
TABLE OF CONTENTS

EXECUTIVE SUMMARY ................................................................. 3
CHAPTER 1 - DATAWATCH VISUALIZATIONS ................................. 4
CHAPTER 2 – SNAPSHOT VISUALIZATIONS ................................. 5
Bar Graph ........................................................................................ 5
Bullet Graph ..................................................................................... 6
Cross Tab Pivot Table ................................................................. 6
Dot Plot ............................................................................................ 6
Heat Map ......................................................................................... 7
Heat Matrix ..................................................................................... 8
Map Plot ........................................................................................... 8
Numeric Line Graph ........................................................................ 9
Numeric Needle Graph .................................................................. 9
Numeric Stacked Needle Graph .................................................. 10
Pie Chart ......................................................................................... 10
Scatter Plot ..................................................................................... 11
Geographic Scatter Plot ............................................................ 11
Shapes/Choropleth ........................................................................ 12
Surface Plot .................................................................................... 12
3D Surface Plot ............................................................................... 13
Treemap ........................................................................................ 13
CHAPTER 3 – TIME SERIES VISUALIZATIONS .............................. 14
Candlestick Graph .......................................................................... 14
Horizon Graph ............................................................................... 14
Line Graph ....................................................................................... 15
Needle Graph ................................................................................ 15
OHLC Graph .................................................................................. 16
Percentage Area Graph ............................................................... 16
Spread Graph ................................................................................ 16
Stack Graph ................................................................................... 17
Stacked and Grouped Needle Graph ........................................ 17
Time Series Scatter Plot .............................................................. 18
Time Series Combination Graph ............................................... 18
CHAPTER 4 – MIXED MODE VISUALIZATIONS ............................ 19
Table ............................................................................................... 19
EXECUTIVE SUMMARY

Visually Design, Discover and Explore New Insights from Any Data in Real Time

VISUAL DATA DISCOVERY

In today’s market, businesses must leverage every shred of data to stay competitive. And with the amount of data continually growing, it’s critical for organizations to analyze, understand, and interact with this data regardless of its type (variety), its size (volume) or the speed in which it is delivered (velocity).

Visual Data Discovery is critical if you are truly looking to get more from your data and not just graph something where you undoubtedly already know the answer. The Datawatch Desktop solution lets you quickly start asking questions to see hidden patterns, spot problems and identify missed opportunities without programming or scripting.

ANALYZE DATA IN MOTION AND AT REST

The idea of “real time” can be very confusing, since virtually all software companies that do visualizations say they can handle “real-time” requirements. However, what most of them mean is that their software goes out and requests an update from an external data source every time fresh data is needed. The data is accurate as of that exact moment, but it becomes out of date immediately since no further updates are available until a complete refresh is done. The information that is being queried comes to rest in a “real-time” data warehouse or database.

Datawatch is unique in its ability to visualize data in motion – by consuming data streams from sources like CEP engines and message brokers which constantly push information into the system instantly, as it happens, on a tick-by-tick basis. So you can see exactly what is happening, as it happens and know precisely how your business is performing.

EXTRACT AND TRANSFORM DATA FROM EXISTING REPORTS

Data is diverse and rarely presents itself in a form perfectly structured and ready for analysis. Some of the most valuable information inside your organization is locked in static operational reports that provide a necessary and trusted set of information but is inflexible. You also have critical data that come from outside the four walls of your company (like invoices, statements, forms, market data) where you don’t have access to the underlying systems.

Datawatch allows users to access, extract and transform any static data into live data for visualization, analysis and sharing with other users and systems. Without programming, a business user opens the report or file in Datawatch and can point and click on the data to be extracted.

Now, let’s get started. Datawatch supports a wide range of visualizations to give you an easy and fast understanding of all of your data, however, there is no one visualization that is ideal for every purpose. This paper outlines and provides guidance for choosing the appropriate visualization for the right analytical task at hand.
CHAPTER 1
DATAWATCH VISUALIZATIONS

Datawatch Desktop software supports a wide range of information visualizations, including our well-known Treemaps, Heat Maps, Scatter Plots, Horizon Graphs, and a wide range of other great visualizations designed for fast comprehension and easy interpretation of static, time series, real-time streaming, and historic data sets.

As no one visualization is ideal for every purpose, the appropriate visualization for the analytical task at hand must be used. Here are some general recommendations:

<table>
<thead>
<tr>
<th>Analytical Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read numeric values quickly</td>
</tr>
<tr>
<td>Performance against a KPI</td>
</tr>
<tr>
<td>Performance across a single variable for a small number of data elements, with different magnitudes</td>
</tr>
<tr>
<td>Performance across a single variable for a small number of data elements, each with similar magnitudes</td>
</tr>
<tr>
<td>Performance across a single variable for a large number of data items</td>
</tr>
<tr>
<td>Performance across a single variable for a large number of data items, which have different importance values</td>
</tr>
<tr>
<td>Performance across a hierarchical or grouped dataset</td>
</tr>
<tr>
<td>Correlation between two categories of data</td>
</tr>
<tr>
<td>Correlation between two or more numeric data columns</td>
</tr>
<tr>
<td>Geographic correlations of data</td>
</tr>
<tr>
<td>Correlation over both a single numeric data column and various categories of data</td>
</tr>
<tr>
<td>Trending performance across ordered categories</td>
</tr>
<tr>
<td>Trending performance between two numeric variables</td>
</tr>
<tr>
<td>Trending performance between three numeric variables</td>
</tr>
<tr>
<td>Trending performance across time</td>
</tr>
<tr>
<td>Time based Ranking</td>
</tr>
<tr>
<td>Time Based Contributions</td>
</tr>
<tr>
<td>Time Based Correlations between time series</td>
</tr>
<tr>
<td>Time Based Transactions</td>
</tr>
<tr>
<td>Financial Time Series Distributions</td>
</tr>
<tr>
<td>Auction Price &amp; Interest/Volume Distribution</td>
</tr>
<tr>
<td>Geospatial Area Densities</td>
</tr>
<tr>
<td>Spread between two time series</td>
</tr>
</tbody>
</table>
CHAPTER 2
SNAPSHOT VISUALIZATIONS

Some of the most common use cases for data visualization software require the system to display information about a data set as it exists at a particular point in time. These “snapshot” visualizations are extremely useful for understanding relative quantitative and qualitative measures and enable users to gain a comprehensive understanding of very complex data sets very quickly.

BAR GRAPH

Bar Graphs are probably the best known visualization for quantitative data.

You can display Desktop Designer Bar Graphs either horizontally or vertically. These graphs are available in three variants:

- Standard
- Grouped
- Stacked

In each case, you can sort the layout of the bar graph according to your requirements, and, with hierarchical data, the graph represents the netted position at each aggregated depth level.

You can also use the Bar Graph visualization to display demographic data in so-called Tornado Charts or Population Pyramids.
BULLET GRAPH

Bullet Graphs were designed by Stephen Few to remove unnecessary clutter and instead focus on visualizing metrics like Key Performance Indicators (KPI).

Research has shown that Bullet Graphs are easier to interpret in less time than the radial gauges or speedometers often seen in BI dashboards.

FIGURE 3-6. A HORIZONTAL BULLET GRAPH.

FIGURE 3-7. A VERTICAL BULLET GRAPH.

CROSS TAB PIVOT TABLE

Although selected visualizations can be cross*tabbed into small multiples, each showing subsets of the original data set, the Cross Tab can itself be used to display a Pivot table.

Pivot Tables support a single numeric value being represented at the cross point of hierarchical rows and columns.

Each intersection cell can display the aggregated numeric value and its associated color range, which can be subdued or intense. Numeric labels can also be removed to produce a Heat Matrix.

FIGURE 3-8. A PIVOT TABLE WITH SUBDUEd COLORS.

FIGURE 3-9. A PIVOT TABLE WITH INTENSE COLORS.

DOT PLOT

Dot Plots have two primary use cases:

- A more effective alternative to a Bar Graph
- A distribution display similar to a Scatter Plot

Dot Plots are an effective alternative to Bar Graphs, particularly in cases where the data being analyzed contains many similar numeric values.

In comparison with the Bar Graph, Dot Plots do not use a zero baseline and are less cluttered. This makes it easier to add additional data variables to the visualization.

FIGURE 3-10. A SAMPLE HORIZONTAL BAR GRAPH SHOWING REVENUES VERSUS FORECASTS.

Dot Plots can also be used to represent data distributions in which one axis is numeric while the other axis is categorical.
Scatter Plots using such data sets can be misinterpreted; Dot Plots of the same data are unambiguous and easy to understand.

**FIGURE 3-11. A DOT PLOT OF REVENUES VERSUS FORECASTS**

**FIGURE 3-12. A DISTRIBUTION DOT PLOT.**

**FIGURE 3-13. A BUSINESS PERIOD CATEGORICAL LINE GRAPH**

**HEAT MAP**

A Heat Map is a special type of color-based data visualization that is well suited for analyzing large flat data volumes using an intuitive graphical display. Heat maps are good at representing large numbers of data points in ways that would be unwieldy and hard to interpret using traditional tables or charts.

A Heat Map represents each item in the data set as an equally-sized cell, unlike a Treemap that uses the size of the box to represent a qualitative value and location to represent hierarchical relationships. In a Heat map, the color of the square represents a quantitative value relative to the other boxes in the Heat map, while the location can represent the sorting of another quantitative or categorical value. This allows the analyst to see all of the data items simultaneously. The user can also hover over any item to bring up more detailed information on demand.

**FIGURE 3-14. A HEAT MAP WITHOUT SORTING.**

**FIGURE 3-15. A HEAT MAP WITH SORTING.**
HEAT MATRIX

A Heat Matrix is similar to both the Heat Map and Treemap in that it displays many different data items and represents the value for each item using colors. However, unlike its cousins, the Heat Matrix has a defined structure where two data attributes define each axis, thus producing a correlation matrix. Within the Heat Matrix, each column and row represents a unique attribute, and the point where two items intersect represents a unique combination of the two attributes.

The matrix can display labels within each intersecting tile or simply display color.

Our Heat Matrix data visualization helps our clients identify correlations within their data sets using an intuitive graphical display.

MAP PLOT

Use Map Plots to display geographic data, where you have longitudes and latitudes associated with individual data points. These plots clearly show data correlations and clustering that is geographic in nature.

In a Map Plot, the visualization expects Latitude and Longitude measures to be associated. It will then retrieve from the selected map tile provider the appropriate background map to display under the data points. This background map is constructed by retrieving individual map tiles at set zoom levels.

As the background map is provided automatically, it relies on:

a) A range of supplied longitudes & latitudes to provide a bounding area.

b) An active Internet connection to retrieve the map tile images.

Datawatch ships with a number of cross reference datasets to determine the appropriate latitude/longitude for datasets. These have been provided through subsets of the data available at GeoNames.org (http://www.geonames.org).

More detailed geo-coding data is available from this website, and many others.

Zooming into the map will cause, new map tiles to be retrieved, and a new background map image behind the data points to be displayed.

For example increasingly zooming into Northern Europe would produce:
NUMERIC LINE GRAPH

Numeric Line Graphs differ from the standard line graph in that they have a numeric X axis, rather than one based upon time.

They are commonly used in both scientific and financial scenarios to show trends in functions that are based on two numeric inputs (X and Y).

Common uses include the display of Yield Curves.

Numeric Line Graphs can also be used to display selected cuts through a Surface Plot.

NUMERIC NEEDLE GRAPH

Numeric Needle Graphs display price distributions.

Unlike a traditional Bar Graph, the X Axis is numeric rather than categorical. Bars are positioned along the X axis according to their X value, and their height is determined by their Y values.

This allows gaps, and clustering in price to be more accurately identified.
NUMERIC STACKED NEEDLE GRAPH

Numeric Stacked Needles again display price distributions.

Unlike the standard Numeric Needle Graph, multiple items can be identified at a single price.

A common usage is displaying client activity within an order book.

FIGURE 3-25. A SEPARATED NUMERIC STACKED NEEDLE GRAPH.

FIGURE 3-26. A NUMERIC STACKED NEEDLE GRAPH.

PIE CHART

Pie Charts are one of the oldest and best known visualizations for displaying contributions to a total.

Desktop Designer can produce standard Pie Charts in which the pie slice represents a numeric variable that is proportional to the total size of the pie. The color variable can represent either a category or another numeric variable.

Pie Charts can be flat, showing a single set of slices. They can also be hierarchical and display multiple levels of data in a variant called a Multilevel Pie Chart. This is also known as a Sun Burst or a Radial Treemap.

The user can modify the visible depth level and drill into particular slices to investigate further detail.

The center of a multilevel Pie Chart can be cut to form a Donut Chart. However, rather than simply leaving the area blank, Datawatch Designer Pie Charts show the aggregate color for the complete data set.

FIGURE 3-27. A TYPICAL PIE CHART.

FIGURE 3-28. A MULTILEVEL PIE CHART (SUN BURST).

FIGURE 3-29. A MULTILEVEL PIE CHART WITH DEEPER HIERARCHY.

FIGURE 3-30. A MULTILEVEL PIE CHART WITH THE CENTER SHOWING AGGREGATE COLORS.
SCATTER PLOT

Scatter Plots are used to identify trends, clustering and outliers across a number of numeric variables, especially when investigating large data volumes.

Each scatter point is represented by:

- X Position
- Y Position
- Size
- Color (numeric or categorical)

A line of best fit can also be added to highlight outliers.

Desktop Designer’s Scatter Plot data visualizations are easy to set up and highly customizable. You can configure your display in ways that will make the most sense to you and your users, and users have all the tools they need to filter and manipulate the Scatter Plot to concentrate on the most relevant subsets in the data.

GEOGRAPHIC SCATTER PLOT

Use Geographic Scatter Plots to display data where physical location is important, and the background map image can be manually provided. These plots clearly show data correlations and clustering that is geographic in nature, and typically used for non-standard mapping.

If a standard map is required then it is likely that the Map Plot should be used instead.

In Geographic Scatter Plots, the X and Y coordinates can correspond to longitude and latitude. The color and size of each scatter point represent other data variables.

As with standard statistical Scatter Plots, you can zoom and pan within the visualization to focus on specific areas of interest, but the underlying map image will not change.

As the background map image is manually provided, the visualization can be used for non-traditional maps, such as internal floor plans.

FIGURE 3-31. A SCATTER PLOT WITH LINE OF BEST FIT.

FIGURE 3-32. A SCATTER PLOT WITH SQUARE SCATTER POINTS.

FIGURE 3-33. A GEOGRAPHIC SCATTER PLOT.
SHAPES/CHOROPLETH

The Shapes visualization allows the display of Choropleth Graphs and other displays built from SVG Paths.

The Shapes visualization can be used to display data where both physical location and size are important.

They clearly show data correlations and clustering that is geospatial in nature.

Unlike the Geographic Scatter Plot, the size of each shape is fixed, imparting the importance of the item. As a consequence, data should be relative to each shape size, such as area densities.

FIGURE 3-34. A SHAPE VISUALIZATION.

SURFACE PLOT

Surface Plots are used to identify trends and outliers within numeric surfaces.

The Surface is made up of a series of points where each point has:

- X Position
- Y Position
- Color (which represents the Z axis).

The Surface Plot can support data sets where the X and Y positions can both be regular and irregular in their distribution.

Additionally, the color scale can be continuous or stepped to show a surface gradient.

FIGURE 3-35. A SURFACE PLOT WITH STEPPED COLORS.

FIGURE 3-36. A SURFACE PLOT WITH CONTINUOUS COLORS.

3D SURFACE PLOT

3D Surface Plots are a 3D perspective version of the 2D Surface Plot.

They provide a clearer understanding of the overall “shape” of the surface but they also introduce occlusion problems; not all data points can be seen due to the display perspective.

The Surface Plot 3D is made up of a series of points where each point has:

- X Position
- Y Position
- Z Position (encoded by color)

The Surface Plot 3D can support data sets where the X and Y positions can both be regular and irregular in their distribution.

The color scale can be continuous or stepped to show a surface gradient.
Grid lines, a ground plane, and markers for data points can be shown if required.

FIGURE 3-37. A 3D SURFACE PLOT WITH STEPPED COLORS.

FIGURE 3-38. A 3D SURFACE PLOT WITH CONTINUOUS COLORS.

**TREEMAP**

Treemaps represent hierarchical data sets, showing both each level in the hierarchy and how they interact with each other.

They are represented by a colorful mosaic of rectangular cells based on your data. The size of a cell reflects its importance. The color conveys urgency or variance:

- White – Target/Benchmark Performance
- Red – Under Performance
- Blue – Over Performance

The intensity of the red or blue shades indicates the level of under– or over–performance.

Most people can learn to understand the information presented in a Treemap in under a minute — even if that Treemap is showing data representing an underlying data set of thousands of records.

Our Treemaps are not static pictures. The real value of the visualization is quickly apparent when you interact with the data. Users can zoom, filter, and view details on demand, as well as link to and highlight other sources of information. For example, fund managers can link to a trading system directly from within the Treemap.

EX supports three different styles of Treemaps:

- Classic Treemaps
- Windows Treemaps
- Cluster Treemaps

FIGURE 3-39. CLASSIC STYLE TREEMAP

FIGURE 3-40. WINDOWS STYLE TREEMAP

FIGURE 3-41. CLUSTER STYLE TREEMAP
CHAPTER 3
TIME SERIES VISUALIZATIONS

The ability to handle very large quantities of multivariate time series data is an essential element in a complete visual analysis system. Desktop Designer offers a range of specialized data visualizations, including Horizon Graphs, Stack Graphs, and Line Graphs, designed specifically to make analyzing historical data easier and more efficient. The software’s ability to connect to traditional row-oriented relational databases or column-oriented databases is key to supporting fast, responsive multi-dimensional analysis of large data sets. Our time series capabilities are especially important for users in global investment banks, hedge funds, proprietary trading firms, and exchanges.

CANDLESTICK GRAPH

Candlestick graphs are a traditional financial visualization for display of time-based price distributions. Specifically, for each time slice, they display:

- Opening Price
- Highest Price
- Lowest Price
- Closing Price

The Candle is filled if the closing price is lower than the open and empty if the closing price is higher than the open.

The vertical line (or candle wick) displays the range of traded prices across the period.

HORIZON GRAPH

Horizon Graphs are a fantastic way to overview a large number of time series in a limited rectangular space. Since this visualization packs the information in a line graph in 1/6th the space through smart pre-attentive color encoding, it allows for an overview of a large number of time series. Users can scan huge amounts of data points across all relevant time series and immediately identify areas of concern that require closer scrutiny.

Our Horizon Graph visualization is particularly useful when you need to see a large number of time series on a single screen. This makes it easy to compare trends and spot patterns that would be very difficult or impossible to see in a standard report.
**NEEDLE GRAPH**

Needle Graphs display time-based transactions or occurrence frequencies, rather than time-based trends. They are simply time-based Bar Graphs where each bar is located at a particular time point on the axis.

They work especially well when combined with a Line Graph.

The most common use of a Needle Graph is when showing the trading volume for a stock, typically underneath the price performance.

**LINE GRAPH**

Line Graphs are easy to understand and are a great way to communicate important time-based trends, clustering, and outliers.

They work especially well when comparing ten or fewer data sets (our Horizon Graph is a good solution for displaying time series data with ten or more data sets).
OHLC GRAPH

OHLC Graphs also display time based distributions of price data. For each time slice, they display:

- Opening Price
- Highest Price
- Lowest Price
- Closing Price

Similar to the Candlestick Graph, a vertical line defines the range of traded prices across the period. However, in this case, the left notch determines the opening price and the right notch determines the closing price.

PERCENTAGE AREA GRAPH

A Percentage Area Graph is like a Treemap spread out over time; you can see how each constituent part contributes to the total at any point in the time series. It is an excellent choice for visualizing time series data when you are interested in seeing the relative contributions for each data set in the series, regardless of the absolute total.

SPREAD GRAPH

The Spread Graph displays the variance or spread between two time-based data series.

Typical use cases include comparing a stock’s price performance to an Index or a bond’s yield to a benchmark rate.
STACK GRAPH

Stack Graphs let you visualize quantitative changes to several data sets over time, and you can see how each data point contributes to the total. As with the Treemap, the height of the stack relates importance, while the color relates urgency or variance.

Stack Graphs are a great way to look at revenue or gross profit figures over time across several product lines. Stack Graphs are also good to use when you have up to ten or eleven time series data sets to look at, especially for data sets that have a large number of data points.

STACKED AND GROUPED NEEDLE GRAPH

Stacked and Grouped Needle Graphs display time-based transactions or occurrence frequencies, similar to the standard Needle Graph.

It allows each transaction to be split into its components, allowing contributions to the total to be viewed across time.

Common uses include splitting of transaction volumes by venue or by direction (Buy/Sell).
TIME SERIES SCATTER PLOT

Time Series Scatter Plots display time-based transactions, similar to the Needle graphs. Like the scatter plot, it displays individual data points (or transactions), with a given numeric Y value and a given timestamp X value.

Common uses include displaying transaction volume across time relative to the price at which the volume was executed.

Typically, the graph is combined with line graphs to show the scatter points relative to defined boundaries.

Figure 2-25 shows trade volumes and prices relative to the best bid and offer across time.

TIME SERIES COMBINATION GRAPH

The Time Series Combination Graph, combines a series of time series visualizations as individual layers of the total display. As a consequence more complex time series visualizations can be built from the “base” visuals.

Each visual can be assigned to either the left or right Y axes, allowing multiple scales to be represented.

For example the following visualization includes:

- Candle Stick Graph – Showing the distribution of prices (OHLC)
- Line Graphs – Showing moving averages of the closing price
- Needle Graph – Showing traded volume across the period
- Spread Graph – Showing a price band across the period

Each of the visuals has a defined “Z” order, which in this case places from back to front:

Spread, Needle, Candle Stick, Line
CHAPTER 4
MIXED MODE VISUALIZATIONS

Mixed Mode Visualizations are capable of displaying time series or snapshot data. In some cases, these types of visualizations can display both time series and snapshot data simultaneously.

TABLE

A table can be used to display a small dataset where all the values are visible or the aggregate values of a larger data set.

The table can be configured to show hierarchies, allowing sub totals and grand totals to be displayed. Additionally, branches of the hierarchy can be expanded and collapsed.

The table can be sorted by clicking on a column heading, and sorting is applied across the defined hierarchy.

Columns cells can be represented in their value form or, alternatively, graphically as a series of micro-charts including:

- Bullet Graph
- Bar Graph
- Dot Plot
- Line Graph

FIGURE 4-1. A SIMPLE TABLE.

FIGURE 4-2. A TABLE WITH HIERARCHY, TOTALS, AND MICRO-CHARTS.

FIGURE 4-3. A TABLE SHOWING SNAPSHOT AND TIME SERIES TRENDS

About Datawatch Corporation

Datawatch Corporation (NASDAQ-CM: DWCH) provides visual data discovery software that optimizes any data — regardless of its variety, volume, or velocity — delivering next generation analytics to reveal valuable insights for improving business. Its unique ability to integrate structured, unstructured, and semi-structured sources like reports, PDF files and EDI streams with real-time streaming data into visually rich analytic applications allows users to dynamically discover key factors that impact any operational aspect of their business. This ability to perform visual discovery against any data sets Datawatch apart in the big data and visualization markets. Organizations of every size, worldwide use Datawatch products, including 99 of the Fortune 100. Datawatch is headquartered in Chelmsford, Massachusetts with offices in New York, London, Munich, Stockholm, Singapore, Sydney and Manila, and with partners and customers in more than 100 countries worldwide. See the Whole Story for yourself by downloading the free trial at www.datawatch.com/trial.

Contact Datawatch

For more information about Datawatch, contact us directly, www.datawatch.com, +1.978.275.8222, or sales@datawatch.com.

Datawatch Corporation, 271 Mill Road, Quorum Office Park, Chelmsford, MA 01824, USA
Toll free US (800) 445.3311 or +1.978.441.2200