An Oracle White Paper
April 2010

Value of Spatial Analytics in Business Intelligence

Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Introduction 2
Information Visualization 2
Geospatial Analytics 8
Spatial Business Intelligence 8
Location and Context 12
Enhancing Location Based Analytics 15
Challenges 16
Conclusion 16

Introduction

The ability to display data using an appropriate visualization is essential to providing insights to business intelligence users. For data with a geographical dimension, geo-spatial views can often be most appropriate. Wider adoption of geo-spatial analytic visualizations in companies has hampered because of the challenge of acquiring mapping and geospatial data, of integrating mapping applications into existing BI products, and of managing these deployments. In this paper, we show how different types of data visualizations can be useful for analyzing data, and how data with a spatial dimension can be visualized and analyzed more effectively. Furthermore, we will also provide examples how spatial visualizations themselves can be enhanced to provide further insights.

Information Visualization

There is an explosion of information that is captured by enterprise systems. This information makes its way into data marts, data warehouses, though often also staying as-is in transactional systems. From these data marts and data warehouses information is then analyzed. The purpose of this analysis is to provide the user(s) with insights in to the data, and therefore, into the business itself. By understanding trends and patterns in the data, business analysts can make more effective decisions about planning and operations.

Apart from the myriad challenges that are faced by business analysts when analyzing data, one of the biggest is deciding how best to visualize the data. Too high level a view and it becomes more or less meaningless; too detailed a view and any insights are buried under massive amounts of data. Visualizations, in order to provide insights, need to be effective, interactive, and bite-sized. Examples of commonly used visualizations used are:

- Tables
- Pivots or crosstabs
- Charts (pie, bar, line, …)
- Maps
- Trellis views
- Text based views
- Scorecards
- Gauges
- Funnel charts
- Hierarchical views
- Tag Clouds
- Treemaps
- Dashboards
- Sparklines
- Heatmaps

Visualizations need not be esoteric to be effective. The simple table and chart views can be just as effective in visualizing data. Let us start with an example. The table below displays one dimension – "District" – and one measure – "Dollars". This can be visualized by means of a table. The benefits of such visualization are obvious – all the data is laid out, precisely, for the user to consume. However, it is not possible to do a comparison of the different districts without scanning all the values individually. This is sub-optimal. An alternative is to sort the values so that the district with the most revenue (Dollars) is displayed at the top.

ORACLE Answers

District	Dollars
BOSTON DISTRICT	$2,802,567
DENVER DISTRICT	$538,484
LOS ANGELES DISTRICT	$1,212,312
NEW YORK DISTRICT	$480,788
PHILADELPHIA DISTRICT	$407,341
PORTLAND DISTRICT	$181,777
SAN FRANCISCO DISTRICT	$105,209
YOUNGSTOWN DISTRICT	$952,287

ORACLE Answers

District	Dollars
BOSTON DISTRICT	$2,802,567
LOS ANGELES DISTRICT	$1,212,312
YOUNGSTOWN DISTRICT	$952,287
DENVER DISTRICT	$538,484
NEW YORK DISTRICT	$480,788
PHILADELPHIA DISTRICT	$407,341
PORTLAND DISTRICT	$181,777
SAN FRANCISCO DISTRICT	$105,209

Figure 1– (a) Table, and (b) Table with data sorted

However, if the purpose is to simply compare "Dollars" across the different districts, and not so much focus on the precise values, then a table is providing more information than is needed at this stage. A chart would convey the information more meaningfully, and facilitate faster visual search of the data. However, for visual search to be effective, it is important to make use of the correct chart.

One of the most common types of charts is the pie-chart. While visually appealing to most, it also suffers from a host of problems. With a large number of slices it becomes difficult to decode the relative values of the slices. It is also difficult to perform a quick visual comparison of the data when represented in a pie chart. In the example below, it is difficult to look at the pie chart and figure out what the values are for the different districts, and also to estimate whether the value for 'Philadelphia District' is more or less than 'New York District'.

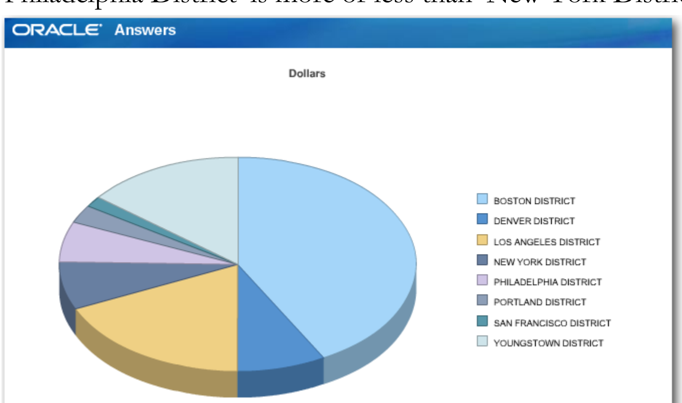


Figure 2 – Pie Chart

Many of these problems can be avoided by using a bar chart. However, even with a bar chart, it is important to avoid adding artifacts of presentation that detract from the task of visual decoding of the data. Such artifacts, as the one shown below, where the bar chart has been formatted with a '*triangle*' effect, do not add anything to the chart, and in fact, lessen the efficacy of the task at hand, viz. to visually inspect the data in the shortest amount of time.

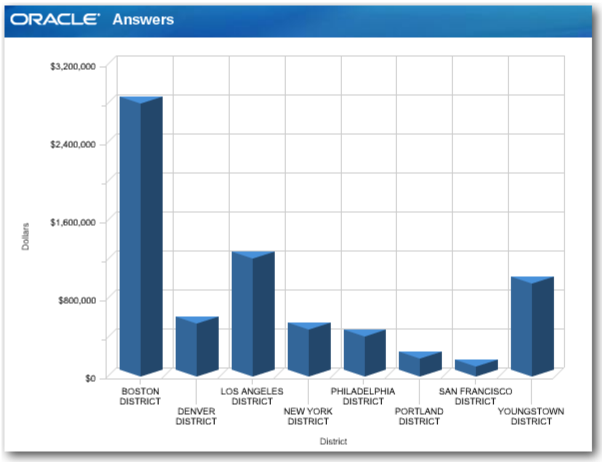


Figure 3 – Bar chart

If we were to draw a simple bar chart, without any needless formatting artifacts, we would end up with the bar chart shown below. Sort the values in a descending order, as shown in the figure below, and it now becomes simple enough for the analyst to perform a quick visual inspection of the chart and see that '*BOSTON DISTRICT*' has the most '*Dollars*', followed by '*LOS ANGELES DISTRICT*', and '*SAN FRANCISCO DISTRICT*' has the least '*Dollars*'. For more detailed analysis, interactivity in the bar chart, for example – tooltips, or zooming capabilities, can be employed.

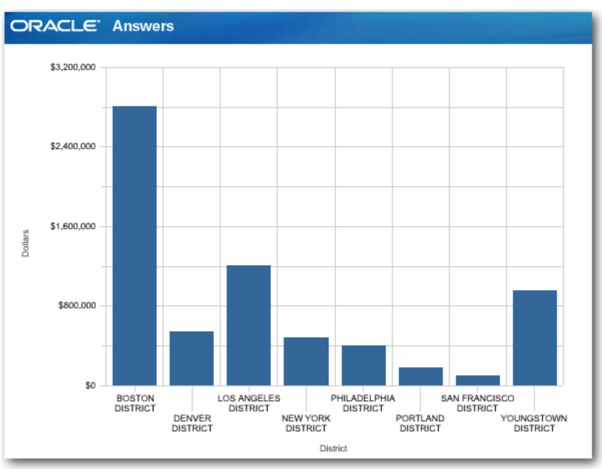


Figure 4 – Simple bar chart

By sorting the bar chart, data can be analyzed at a glance: the relative rankings of the districts are visible, with pre-attentive processing coming into play. For more detailed analysis, interactivity in the bar chart, for example – tooltips, or zooming capabilities, can be employed.

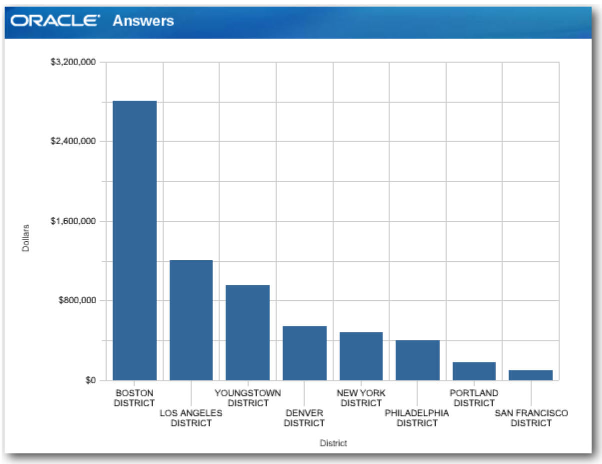


Figure 5– Bar chart with data sorted

Let us look at one more example. The table here presents monthly 'Dollars' for five quarters, beginning with Q1 of 2001. If were to try and identify all months where 'Dollars' were greater than $500,000, the table in the first figure makes the task difficult, since all individual values have to be scanned visually, sequentially. The second table makes this task easier, by displaying the data, sorted in a descending order. The user has to quickly scan the *most significant digit* in the measures column, till you see that the digit has changed from '5' to '4', at which point you know that all rows above match the search criteria. A further improvement can be made to this tabular display of data by color-coding the data, so that all values exceeding the threshold are formatted with a different color, in this case red. This type of a presentation makes for the fastest visual processing of the data. A fourth type of display is a combination of the second and third figures below, sorting the data while also applying a color-coded format.

ORACLE Answers

Year	Month	Dollars
	1/1/2001	$548,504
	2/1/2001	$425,983
	3/1/2001	$600,889
	4/1/2001	$562,739
	5/1/2001	$547,457
	6/1/2001	$688,249
2001	7/1/2001	$401,049
	8/1/2001	$453,513
	9/1/2001	$513,129
	10/1/2001	$515,511
	11/1/2001	$445,619
	12/1/2001	$541,274
	1/1/2002	$383,429
2002	2/1/2002	$354,942
	3/1/2002	$416,159

ORACLE Answers

Year	Month	Dollars
2001	6/1/2001	$688,249
2001	3/1/2001	$600,889
2001	4/1/2001	$562,739
2001	1/1/2001	$548,504
2001	5/1/2001	$547,457
2001	12/1/2001	$541,274
2001	10/1/2001	$515,511
2001	9/1/2001	$513,129
2001	8/1/2001	$453,513
2001	11/1/2001	$445,619
2001	2/1/2001	$425,983
2002	3/1/2002	$416,159
2001	7/1/2001	$401,049
2002	1/1/2002	$383,429
2002	2/1/2002	$354,942

Figure 6 – Unsorted and sorted table

ORACLE Answers

Year	Month	Dollars
	1/1/2001	$548,504
	2/1/2001	$425,983
	3/1/2001	$600,889
	4/1/2001	$562,739
	5/1/2001	$547,457
	6/1/2001	$688,249
2001	7/1/2001	$401,049
	8/1/2001	$453,513
	9/1/2001	$513,129
	10/1/2001	$515,511
	11/1/2001	$445,619
	12/1/2001	$541,274
	1/1/2002	$383,429
2002	2/1/2002	$354,942
	3/1/2002	$416,159

ORACLE Answers

Year	Month	Dollars
2001	6/1/2001	$688,249
2001	3/1/2001	$600,889
2001	4/1/2001	$562,739
2001	1/1/2001	$548,504
2001	5/1/2001	$547,457
2001	12/1/2001	$541,274
2001	10/1/2001	$515,511
2001	9/1/2001	$513,129
2001	8/1/2001	$453,513
2001	11/1/2001	$445,619
2001	2/1/2001	$425,983
2002	3/1/2002	$416,159
2001	7/1/2001	$401,049
2002	1/1/2002	$383,429
2002	2/1/2002	$354,942

Figure 7– Unsorted and sorted tables with color-coded values

While this may appear an ideal presentation of the data, there are two problems here. The first problem is that the tabular display is providing more information than the user is interested in at this point. If the object of the user's search is to simply find all months where '*Dollars*' exceeded *$500,000*, then it is immaterial whether the number is *$513,129* or *$513,000* or even simply '*>$500,000*'. The second issue is that the chronological linearity of the data has been lost in the second and fourth tables above. The data, time-series in nature, is now no longer displayed as such.

Both problems can be alleviated by using a chart. However, a line chart is the best chart type in this case. The same data as in Figure 6 above is displayed in this line graph below:

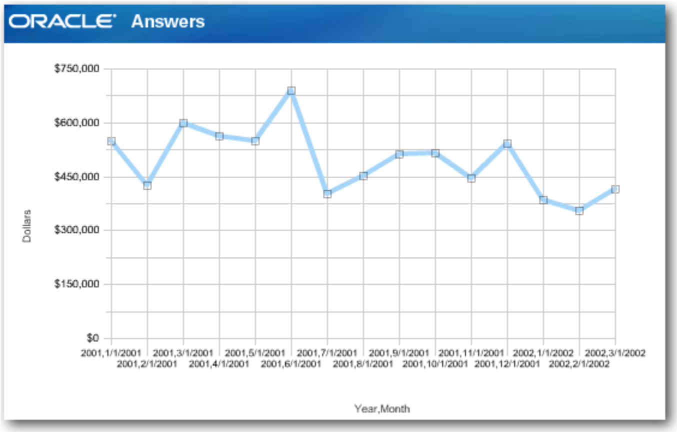


Figure 8– Line chart

Following the same principles as followed in the table, where we color-coded the values of interest, we can color-code the points on the line chart that match our requirements. That is, mark all months where '*Dollars*' were greater than or equal to *$500,000* with a red colored marker.

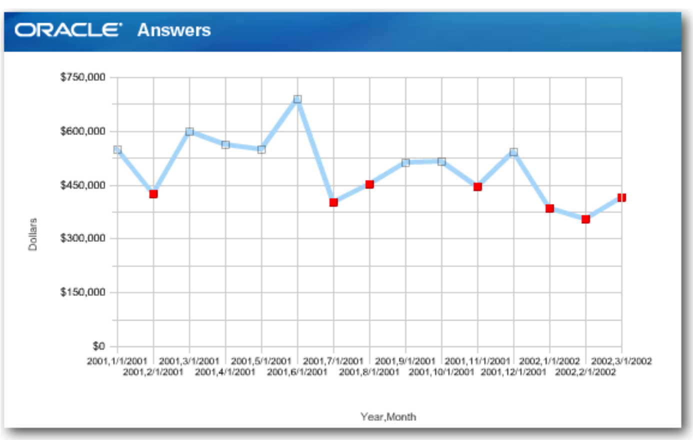


Figure 9 – Line Chart with matching values color-coded

Finally, instead of sorting the values in the line chart, what we can do is draw a threshold line can be drawn for greater emphasis:

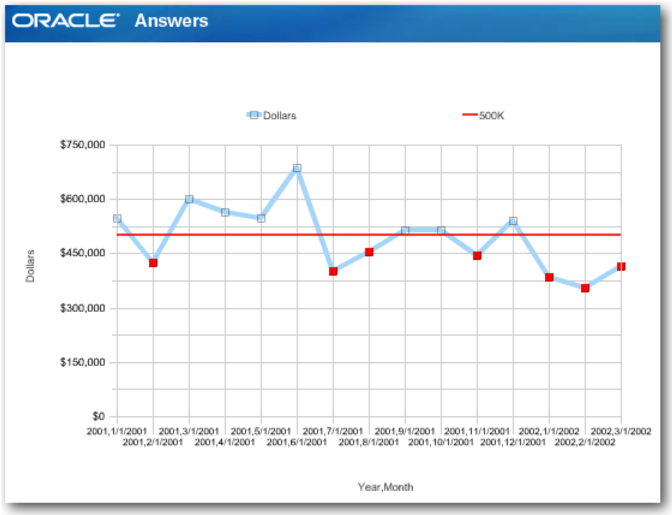


Figure 10

Geospatial Analytics

Thus far we have seen that tabular data is useful when the amount of data to be analyzed is small, and its readability can be improved through the use of judicious formatting. Where the amount of data starts to increase, charts offer a convenient way of viewing trends and also obtaining a high level overview of the data. Line charts are ideal for visualizing time-series data, while bar charts should be used for the vast majority of other data.

However, data with a spatial dimension is, in most cases, best visualized via the use of maps. Maps are used to represent spatial phenomena or relationships such as flow or proximity. They are often used for

1. Planning – determining where to locate a store or service facility;
2. Understanding spatial relationships – patterns and locality in crime;
3. Display jurisdiction, ownership, or assessment – school districts, property and tax or zoning maps;
4. Navigation and route planning – route guidance or planning and scheduling school bus, or postal delivery, routes; and
5. Forecasting or warning – weather maps or assessing the likely risk to insurance portfolios from a predicted event as a hurricane or wildfire.

Spatial Business Intelligence

Maps also allow for the display of statistical measures for an area or region. For example, average household income by census block or population density by county. In addition maps allow for multi-measure displays. For example, one could display variable-sized circles, whose size indicated a range for average family housing prices by census block, over a color-shaded map of average household income by census block. Similarly, one might display a pie-chart indicating percent of population in various ranges (< 15, 15 -- 25, 25 -- 65, and > 65) on a population density map by county.

Interactive digital mapping technology empowers end-users. One can further enhance multi-measure displays described above, for example, by providing adjustable sliders (so users can modify the ranges or colors), or a layer selection tool so displayed measures can be switched on or off. For example, a population density map could show racial mix, gender profile, educational level, or other measures as layer choices on the same map and the user selects the one(s) to display at any time. Though for most cases no more than two layers (or formats) should be used at any given point in time on a map. More formats can be created, but the end-user should be given the option of choosing to hide them as needed.

The key benefit of interactive mapping and geospatial technology, however, is that some types of analysis possible only with spatial analytics and best visualized on maps. For example, analysis such as finding the top 10 performing stores, by sales volume, within a user-drawn region of interest. Or finding the number, and types, of public safety incidents within a certain distance (e.g. 100 meters) from a specified location (e.g. public parking lot or park).

Consider the following example which draws upon a crime data warehouse maintained by a city. Data is captured and available based on a number of dimensions such as the service area, the region, time, type of offence, etc… and measures. The requirement then is to be able to report on this data, and also to plot this data on a map to allow for hitherto un-discovered insights to be gleaned.

In the example below, a simple thematic map is displayed, with the number of offences per service area for a given region displayed as a color overlay. The end-user can adjust the color thresholds by adjusting the color slider available at the top of the map).

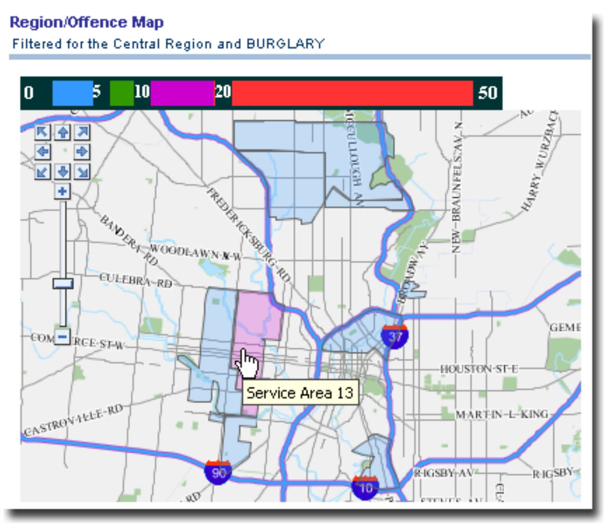


Figure 11– Map with number of offences overlain on region map

This simple map can be enhanced by providing information rich popups that serve to display further details about the layer over which the user's mouse is hovering, and also by providing a link for the user to click. This link can provide more details as needed – for example, in the form of a tabular report listing all the offences for the selected service area.

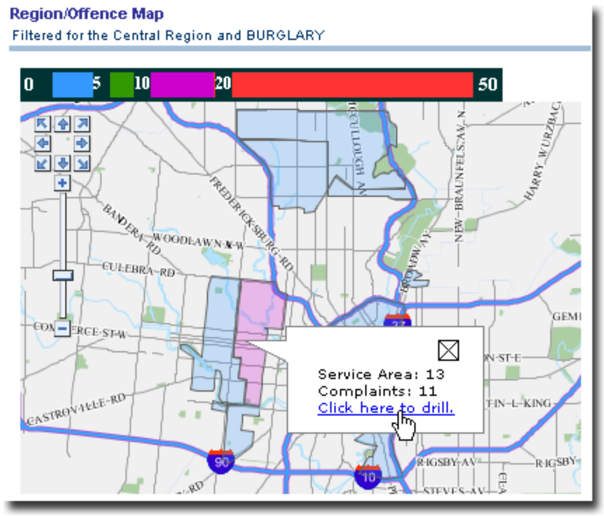


Figure 12– Thematic map with a drillable information popup.

The following (Figure 13) is an example of a compound visualization that combines the best uses of a tabular report and of a map into one, linked visualization. The dropdown at the top serves to filter the data based on a geographical dimension – *Service Area* – and on a second attribute – *Offence Type*. Each row in the resulting table is displayed on the map at the right. Furthermore, in addition to the information popup available to the end-users, the ability to 'mark' the spot by pinning it is also now available, thus serving to immediately draw the user's attention to this point.

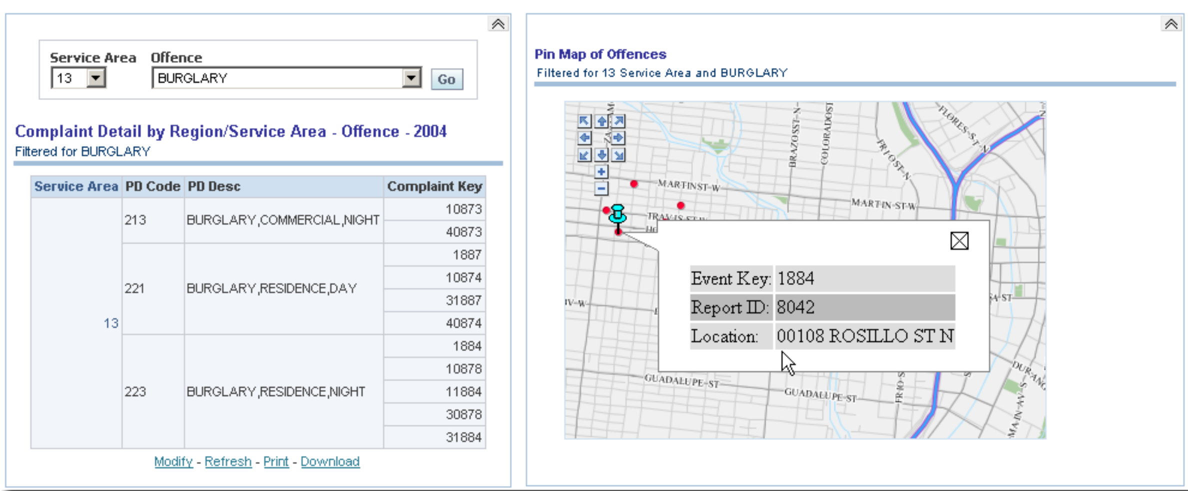


Figure 13– Compound visualization, with the ability to mark a spot on a map via a pushpin

The figure below is a slight modification of the above example. Instead of first rendering the table and the map and then placing a pushpin marker on a map, in the example below, what we are doing is to use the pushpin marker as a filter to the table that now displays data only for the selected point. The table can be used to display several more attributes and details about the selected point on the map. Therefore, using a combination of two views, the map is serving as a high-level overview view, while the table serves as a detail view, and the user can select slices of data to analyze in more detail. That is, interactive maps facilitate spatial filtering.

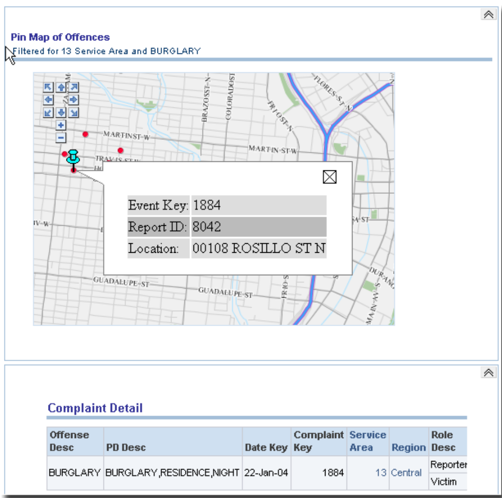
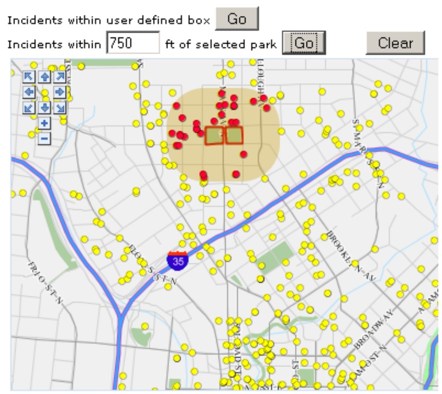


Figure 14– Pushpin on map is used to filter data in the table

Spatial bounding boxes can be used to empower end-users in a very unique manner. Whereas filters and parameters have been traditionally used to filter data based on a textual view of the values available to be filtered on, spatial bounding boxed provide a very visual means and metaphor to filter data, directly on a map. For example, users can select a point of interest such as a part, in the example below, and then apply a geo-buffer to select all incidents that fall within a user-defined distance. Such type of analyses is very difficult without geo-spatial visualizations and spatial analytics.



Complaint Detail

Offense Desc	PD Desc	Date Key	Complaint Key	Service Area	Region
CRIMINAL MISCHIEF & RELATED OF	MISCHIEF, CRIMINAL 4, OF MOTOR	18-Feb-03	1026	28	Central
DANGEROUS DRUGS	CONTROLLED SUBSTANCE, POSSESSI	10-Nov-02	30099	28	Central
		10-Mar-03	40099	28	Central
HARRASSMENT 2	HARASSMENT,SUBD 1,CIVILIAN	02-Aug-03	1064	32	Central
	HARASSMENT,SUBD 3,4,5	04-Mar-03	1027	28	Central
		04-May-03	31027	28	Central
		04-Sep-03	41027	28	Central
		19-Sep-03	41028	28	Central
ROBBERY	ROBBERY,UNCLASSIFIED,OPEN AREA	09-Jan-04	41032	28	Central

Figure 15– Geo-spatial buffer used to filter data

Lastly, not all spatial analysis or filtering results require a map. Consider, for example, the following query to find all competitors within 2 miles of Northport Branch:

```
SELECT c.holding_company, c.address
  FROM competitor c,
       bank b
 WHERE b.site_id = 1604
   AND SDO_WITHIN_DISTANCE(c.location, b.location, 'distance=2
unit=mile') = 'TRUE'
```

A tabular result is sufficient if all we're interested in whom the competition is rather than how their proximity may influence the performance of our facility.

Location and Context

While location provides context, a map often removes content to highlight and de-clutter the information conveyed on it. Thus traditional maps go only so far in providing users with the information and the context they need in their line of work. Sometimes, however, location information too needs additional context. In some industries, the ability to obtain a real-world perspective, an eye-on-the-ground view so to say, is vital. Examples are: Retail Planning, Insurance Analytics, Telecom, Utilities. Maps with satellite imagery provide that additional context.

Consider the series of maps we saw in the examples above. These are street maps with analytics information overlain on top. The same information can also be displayed on top of a hybrid map that combines street views and satellite imagery. Without having to extract more information from the analytics data warehouse, or to feed the required information into the data warehouse in the first place, the end-user can visually inspect the incident location for further attributes of interest.

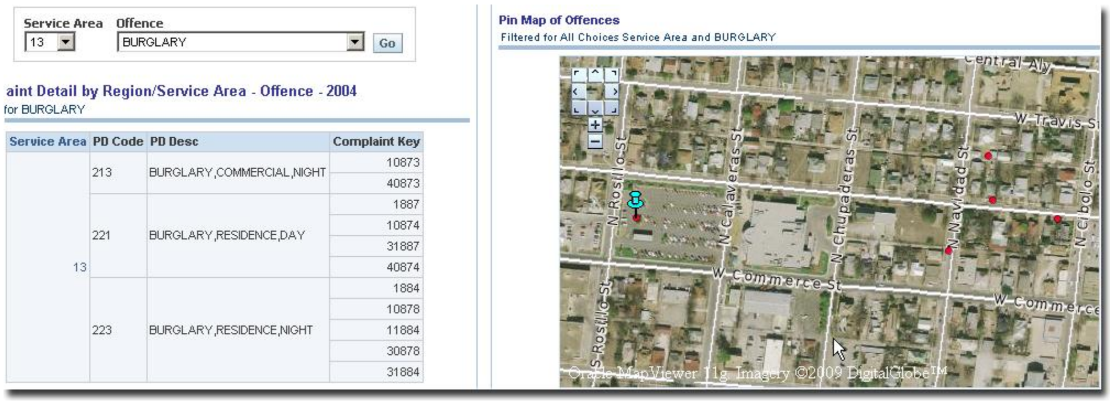


Figure 16– Hybrid map showing street and satellite views

Another example should illustrate the value of adding this additional element of locational context to map-based visualizations. In the example below, water quality for a number of regions across the United States is monitored by the Environmental Protection Agency (EPA). It then reports on this data, and visualizes it as tables, charts, and maps. As you can see below, the *Change from Prior Year* table values are supplemented by a traffic-light (or stoplight) icon that indicates whether the water quality has deteriorated or improved over the prior year. This same color coding scheme is also applied to the map-based visualization.

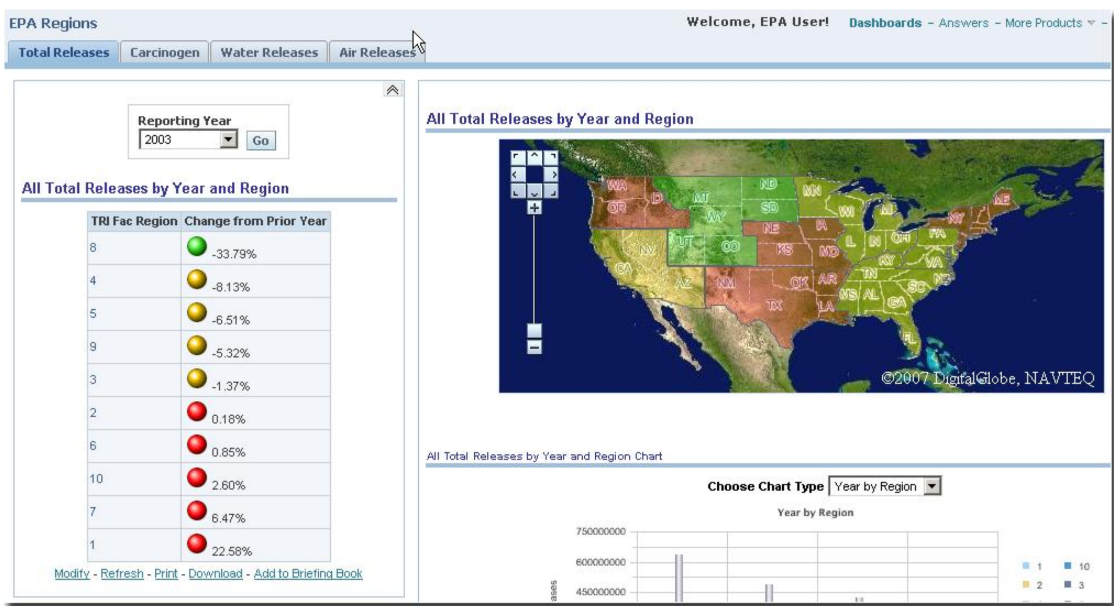


Figure 17

Users can then drill down, or zoom into the map. As they zoom in, different locations where water samples are collected are indicated by appropriate colored markers. Therefore, **red**, **yellow**, and **green** markers on the map continue to provide a visual indication of the water quality change from the prior year. As we saw before, an information tooltip over the marker displays the facility name and the precise value of the change.

As we drill down from a country level into region, state, and county it's the thematic map that conveys the relevant context. Once we drill to an individual facility level (Fig 21), however, the high resolution satellite imagery has the potential to convey a lot more relevant context in this use case.

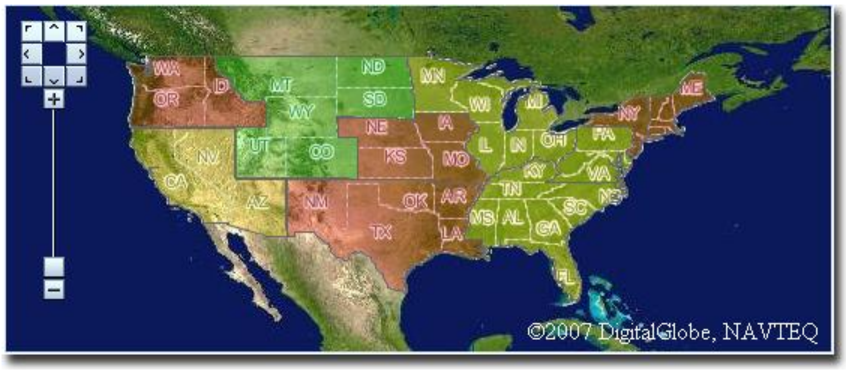


Figure 18– Color-coded satellite-map visualization

Figure 19– Water Quality map at the County level

Figure 20– Water quality map at the facility level

Figure 21– Water quality map displaying zoomed-in location

Enhancing Location Based Analytics

Finally, location-based analytics and map-based visualizations can be enhanced even further by integrating them with a time dimension. Data analytics is multi-dimensional. Time and Geography are the two most important dimensions in most analytics. Consider these examples:

- Sales by country
- Sales by Year
- Sales by Year by Country
- Y-o-Y Growth in Sales by Country / State / City / Region

Just as data changes with time, so does geography. Performing data analytics over time and geography on maps is very useful, but can be enhanced if users get the ability to look at time slices and map overlays for the same slice of time.

In the example below, sales for a region have been displayed on a map as a color overlay, color-coded. Furthermore, the user gets the option of selecting a slice of time for which to view the satellite overlay for.

Figure 22- Color-coded maps with time-slice selection capability

Challenges

Convincing businesses and analysts of the value of business intelligence is today no longer an issue. Businesses are convinced more than ever of the need to put in place an integrated, complete, open analytics solution for their users. Business intelligence, analytics, performance management, and related technologies and solutions are a means for companies to gain a competitive advantage in their respective industries. Similarly, geospatial visualizations are gaining acceptance within enterprise applications and tools. For long the preserve of select industries, geo-spatial visualizations and analytics are starting to make inroads into a wider swath of industries.

The challenges however are still considerable, and formidable. To name a few:

- The acquisition of **spatial data** continues to be a formidable one. The emergence of companies that provide this data over the net helps alleviate this problem. However, for certain use-cases, where companies do not want their data being transferred to a third-party's servers, an in-house hosted model may be the only one.
- Customers do not want to spend time and money on a custom-crafted solution that fulfills a subset of requirements at a given instant in time. They want a self-service model where their users can create map-based visualizations using a simple, point-and-click interface, without any intervention from IT, or without having to spend hours on creating these visualizations.
- Performance, Scalability, Reliability, Manageability are all very important issues that the IT department of a company grapples with in its quest to ensure that the solution available to the company's employees actually remains available on a reliable basis.
- Most out-of-the-box solutions typically fulfill 80% of the requirements of customers. The remaining 20% of the requirements are either left unmet, or are provided over a period of time in subsequent releases of a product, or are attempted to be met via custom extensions to the product. These custom extensions bring with them their own issues of manageability, cost-of-development, and upgrade issues. The answer lies in providing a pluggable and extensible framework that allows a company to extend the 80% OOTB solution as needed.

Conclusion

There is an explosion of information that is captured by enterprise systems. This information makes its way into data marts, data warehouses, though often also staying as-is in transactional systems. The ability to display data using an appropriate visualization is essential to providing insights to business intelligence users. For data with a geographical dimension, geo-spatial views can often be most appropriate. However, wider adoption of geo-spatial analytic visualizations in companies has hampered because of the challenge of acquiring mapping and geospatial data, of integrating mapping applications into existing BI products, and of managing these deployments.

Data with a spatial dimension is, in most cases, best visualized via the use of maps. Maps are used to represent spatial phenomena or relationships such as flow or proximity. They also allow for the display of statistical measures for an area or region. Their expressive power lies in the fact that they provide context, allow for multi-dimensional displays and filtering of information, and are interactive. Interactive digital mapping technology empowers end-users.

Spatial filtering, such as regions of interest or bounding boxes, empower end-users in a very unique manner. Whereas prompts and parameters have been traditionally used to filter data based on a textual view of the values available to be filtered on, spatial bounding boxed provide a very visual means and metaphor to filter data, directly on a map. Finally, location-based analytics and map-based visualizations can be enhanced even further by integrating them with a time dimension.

Despite the formidable challenges of (i) acquiring, and using, spatial data; and (ii) custom crafting a solution since most existing platforms only meet, or attempt to address, at best 80% of the typical requirements; and (iii) doing so while satisfying the scalability, reliability, security, and manageability requirements of an enterprise-wide deployment; the need and value of spatial analytics and visualization are apparent to enterprise users. Given this need there will be vendors ready to meet it and do so with an integrated, complete, open analytics platform.

Value of Spatial Analytics in Business Intelligence
April 2010
Author: Abhinav Agarwal, Jayant Sharma
Contributing Authors:

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com