★★★★★★★★

Practical Data Visualization – Businesses Competing on Analytics with Data

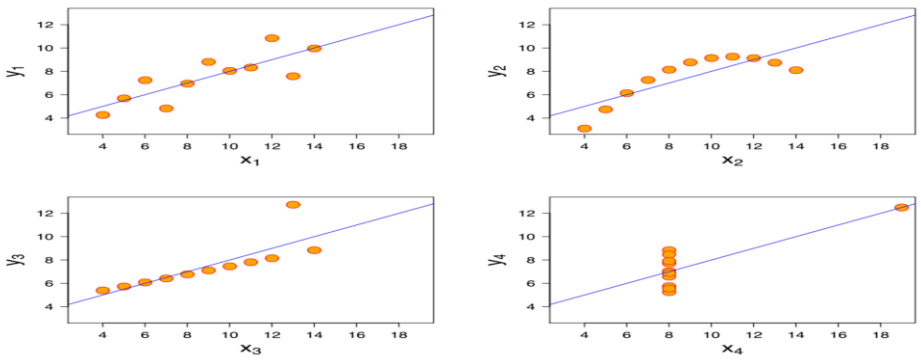
2012

PCA Business Notes Series

Data Visualization for Business Data Analytics

William Playfair said 200+ years ago: "As the eye is the best judge of proportion, being able to estimate it with more quickness and accuracy than any other of our organs, it follows, that wherever relative quantities are in question … [the Line Chart] … is peculiarly applicable; it gives a simple, accurate, and permanent idea, by giving form and shape to a number of separate ideas, which are otherwise abstract and unconnected." William Playfair invented four types of Data Visualizations: Line and Bar Charts (1786), Pie and Circle Charts (1801). The **Origin** of Data Visualization is the **eye's superiority**!

Pure Analytics – often **misses patterns** like in the famous Anscombe Quartet (each of four datasets has just 11 data points!):



For these four completely different datasets we have **identical statistical properties**: Mean and Variance for both X and Y, Correlation and Regression!

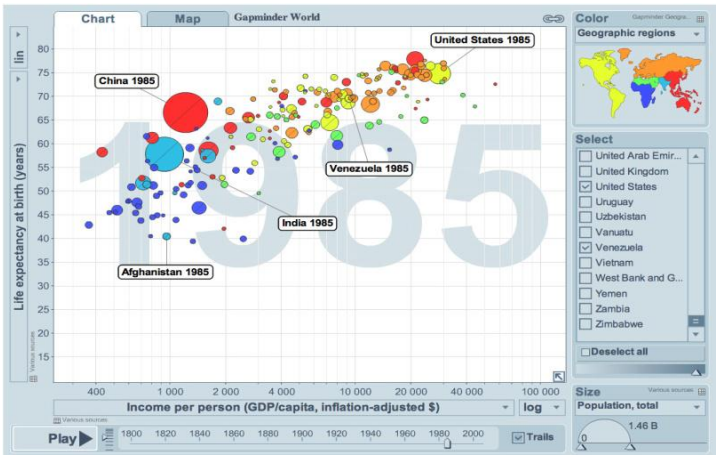
The main reason users are migrating to Data Visualization from Static Reports and Business Intelligence Paradigm is simple: it easy to use and **explore data** for average users and allows management to quickly **see trends, detect outliers** and interactively **drilldown** to data exceptions (outliers) and actionable information. Users instantly see information without special training using the **self-evident user interfaces** of modern Data Visualization tools.

175 Highland Avenue
Needham, MA 02494
617.527.4722

Data Visualization: how much Data can a Human Eye see?

The human eye has its own Curse of Dimensionality. In most cases the data (before it's visualized) need to be organized in Data Marts and multidimensional Cubes and then needs to be projected into synchronized and less-dimensional datasets (e.g. 3-Dimensional Cubes) before it can be exposed through an interactive, synchronized set of charts on the 2-dimensional surface of computer monitor in a Data Dashboard form.

During the last 200+ years people designed charts to be printed on paper or shown on screen. Most charts show 2- or 3-dimensional datasets. Dynamic charts can show more dimensions – as many as six dimensions can be shown by a Motion Chart:



Each Bubble has 6 dimensions represented on the x, y, Size, Color, Name, Animation and Shape. In the example the dimensions shown are: Income, Life expectancy, Country Population, Continent, Country, and Year (animated 6th Dimension).

Data Collection

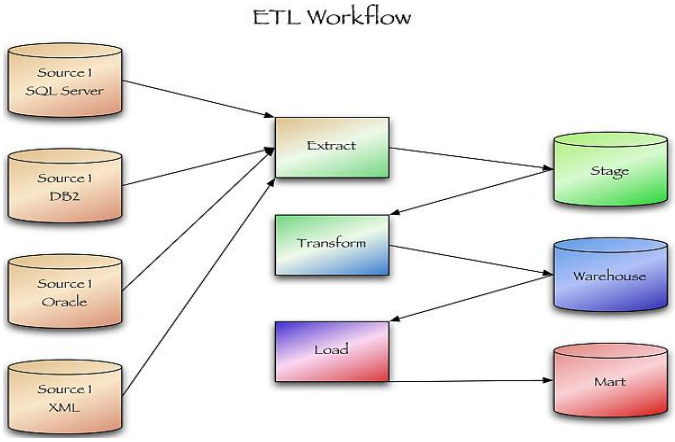
Before data can be visible and interactive, it needs to be collected into Data Collecting relational databases, Email Systems or Document Repositories. A significant amount of data is collected automatically, such as metered electricity usage. Other data is collected manually by computer users via surveys, such as the US Census or by scanning objects (with scanners or just … eyes) and then entering the data through a Data Collector (application or form) into data stores. Still other data is derived (calculated) from primary data sources.

Extract, Transform and Load (ETL)

Extraction takes data from data sources with different structure and format, pre-validating extracted data and parsing valid data to a Staging Database.

Transformation applies predefined rules and functions to it, including selection, data cleansing, encoding, deriving, calculating, sorting, joining data from multiple sources (e.g. merging), aggregation (e.g. summary for each month), transposing (columns to rows or vice versa), splitting, disaggregation, lookups (e.g. validation through dictionaries), predefined validation etc. which may lead to rejection of some data. Transformed data are stored in a Data Warehouse (DW).

Load takes transformed data and places it into the end target, in most cases called Data Mart. Load appends, refreshes or/and overwrites preexisting data, applies constraints and enforces data integrity, provides data backup and replication into Data Marts and Columnar Databases:



Big Data Analytics and Structures for Big Data

We are already overloaded with data: "Big Data Analytics" (BDA) is coming your way (like it or not) regardless of the nature of your work, profession and business. Big Data Analytics is the newest and most compelling reason for using Data Visualization. We are witnessing a rapidly growing size of data, which cannot be explored, analyzed and utilized without Data Visualization Tools.

According to IBM, in 2010 we created 2.5 Exabytes of data per day and "90% of the data in the world today has been created in the last two years alone". In 2011 the world will produce almost 2000 Exabytes and in 2020 it may reach 35,000 Exabytes! Big Data streams everywhere from climate sensors to posts on social media sites, digital pictures and videos posted online, EBay and Amazon transaction records of online purchases, Google searches and from cell phone GPS signals.

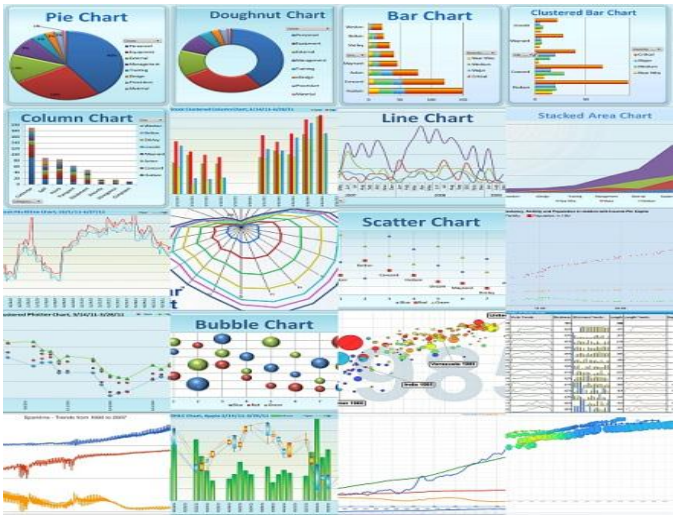
Business users interact with huge datasets through front-end Data Visualization tools and are concerned about trends, outliers, clusters, patterns, drill-downs and other visually intensive data phenomena.

Visual Data User Interfaces

When we are ready to visually explore our data, drilldowns, slice and dice, and analyze data patterns, such as data clusters, data trends and outliers**,** we need simple ways to view and interact with our data. We need to present data as interactive,

Practical Computer Applications, Inc.
175 Highland Avenue | Needham, MA
(617) 527-4722 |www.pcapps.com

searchable and synchronized Charts. Groups of related and synchronized charts when combined in one window or webpage are usually called Data Dashboards.



Some of the more popular chart designs consist of: Pie, Bar, Line, Area, Plot, Scatter, Bubble, Radar, Time Series, Sparklines, Stock and Motion Charts etc., all used to express the different combination of Dimensions, Attributes and Measures for datasets with hundreds or thousands of "currently" visible datapoints (and millions and billions of temporary hidden datapoints).

Data Exploration and Data Patterns

Data Visualization works well for good reasons:

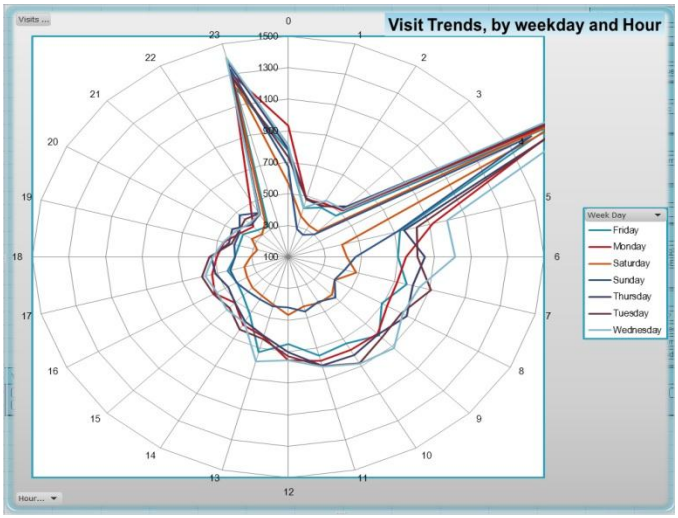
1. Data Visualization fits many more (tens of thousands) data points into one screen or page as compared to numerical information and data grids;
2. It enables users to visually drilldown and zoom through interactive and synchronized charts;
3. It conveys a visual story behind the data;
4. Analysts and decision makers can see patterns (**clusters, trends, outliers** etc.) in data with Data Visualization. An example is the 37+ year old example above, known as Anscombe's quartet, which comprises four datasets that have identical simple statistical properties, yet appear very different when visualized.

See it 1st: Trends, Outliers, Clusters

Good Data Visualization allows users to "see this" before "analyze this" and to take advantage of human eye and brain's ability to recognize trends, outliers and data clusters more quickly. Visual Trend analysis helps to "spot" a pattern, or trend, in data or predict future events.

The following chart shows that the number of visitors to a website depends on the time of the day and on the day of the week. Hourly trends are visualized as well as day of week picture!

(Note: Hourly traffic is on the circumference, day of week denoted by color.)



It is easy to see that around 11pm and 4am of each day the number of visits far exceeded any other time of the day (perhaps due to spiders). The user may need to analyze those visits in more detail.

Visual Drilldown: Business Intelligence (BI) Problem: The Needle in a Haystack

The eternal Ad-Hoc problem between BI Users and BI Developers: Users want to have flexible reports, but developers in most cases can deliver only pre-defined ("rigid") or ad-hoc reports. Visual Drill-down capabilities allow users to have their cake and eat it too. Drill-down capabilities enable users to narrow the data selection by setting data filters to specific values or/and by visually choosing regions (e.g. rectangles) on interactive charts, so that only those values associated with regions will stay in the data selection. Chart sections behave as filters eliminating the need for filters since many users prefer to use the mouse to select significant data.

Drilldown results in the repainting of all synchronized charts with only the selected data. After drilling-down to the appropriate data, the user can switch to Visual Data Exploration to examine trends, outliers, clusters, similarities and other data patterns.

Practical Computer Applications, Inc.
175 Highland Avenue | Needham, MA
(617) 527-4722 |www.pcapps.com

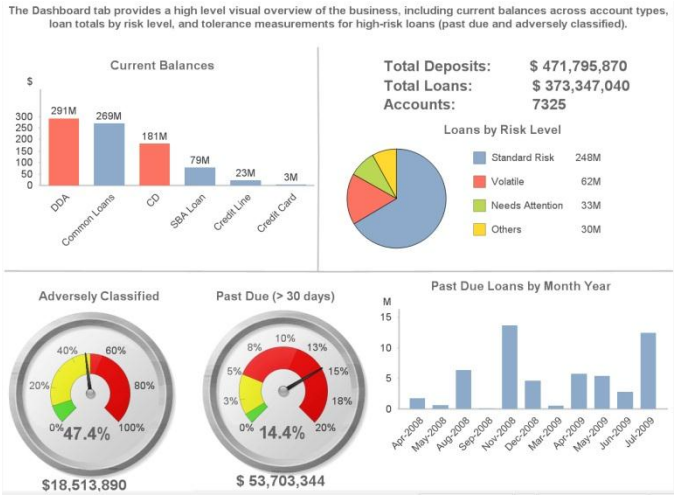
Analytical Business Dashboards

Dashboards can be used for strategic, analytical, operational, or informational purposes. For example, an automotive dashboard allows the driver to monitor the state of a car (speed, fuel, heat level etc.) and acts on it (e.g. push a pedal or stop for a refuel).

Business dashboards do the same for business user. Dashboards are typically limited to summaries, key trends, comparisons, and exceptions. Well-designed business dashboards provide powerful means to monitor and act on business information.

A business dashboard integrates multiple Data Charts, Key Performance Indicators (KPIs), Balanced Scorecards, Tabular Snapshots of Data, and/or Pivot Tables. These may be synchronized with each other or be synchronized with real-time data feeds, transactional databases, multidimensional data cubes, in-memory columnar databases and other of data repositories. These United Synchronic Data Views show relationships between different business measurements enabling users to act on changes and exceptions in the data.

Analytical Dashboards typically support interactions with the data, such as drilling down into the underlying details.



Operational Data Dashboards

Operational Dashboard enable users (managers or operators) to monitor operations, activities and events in real-time that might require attention and response at a moment's notice. These dashboards summarize large volume of information and highlight changes and exceptions in the data.

Data Visualization Market and Vendors

The Data Visualization Market recently reached 1 billion dollars in size and is growing faster than any other software market, including traditional Business Intelligence (BI) tools. Among Data Visualization Leaders we mention important players:

Tableau Software – the fastest growing Data Visualization Vendor (113.5% Year-over-Year growth in 2010), easiest for end users with minimal need for IT support, has best access to OLAP cubes and best Pivot Control among DV vendors. The recent Tableau release has a capable and fast in-memory (64-bit) data engine and ability to perform incremental data updates.

Spotfire – lost (in 2007-11) its Data Visualization focus with TIBCO's acquisition of Spotfire, but still has the best web client, best Visual Analytics (S-Plus, IronPython) and API; still the best option for large enterprises. Spotfire has an excellent in-memory (64-bit) data engine and unlimited ability to scale-up to disk.

Microsoft PowerPivot, Power View and Excel – Microsoft is clearly the elephant in the room. Microsoft tools are catching up quickly, and may overtake their competitors. Microsoft's biggest advantage is the integration of Excel into Microsoft's BI suite, and the ability to effectively program any type of visualization using those tools. Microsoft is clearly the most "Open" approach. With some energy and effort, excellent results can be obtained very flexibly.

QlikTech – fast growing, with good Data Visualization functionality, especially visual drilldown but has limited flexibility for large enterprises. Qlikview has a good in-memory columnar database and so called AQL ("Associative Query Language"). QlikView is less flexible than the other products and the company is more rigid.**See a comparison of Tableau vs. QlikView and Tableau vs. Spotfire**

PCA Recommendations for Data Visualization

Data Visualization is only as good as the data and analytics behind it. Good Data Visualization will enable timely monitoring of operations, trends and outliers; alert decision makers ahead of deadlines, and enable management to quickly drilldown to actionable data items and simplify enterprise reporting and analytical processes.

The PCA process, business discovery, and expertise in construction of effective Data Warehouses and easy-to-use **Data Visualization** applications for interactive reports, analytical and business dashboards can greatly improve the Return on Investment (ROI) of fast growing corporate databases, help our clients choose and embed appropriate Data Visualization technology into business processes, or effectively apply such technology if it is already in use.

Contact us at (617) 527-4722, (877) 843-3405, or pdbinfo@pcapps.com

Practical Computer Applications, Inc.
175 Highland Avenue | Needham, MA
(617) 527-4722 |www.pcapps.com