Data Warehousing for Business Intelligence
Course 4: Business Intelligence Concepts, Tools, and Applications

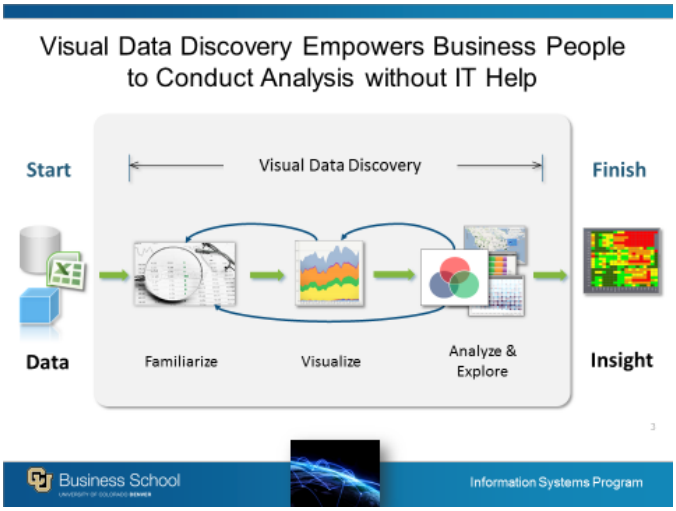
Module 3 Bonus Materials
Lesson 1: Data Visualization

We've arranged for students in this MOOC to purchase at a very low cost digital versions of chapters 1, 2, and 4 of the authoritative textbook *Business Intelligence and Analytics: Systems for Decision Support*, 10th edition, 2015 by Sharda, R., Delen, D., and Turban, E. See the optional text book link under course overview to purchase (US$4 for one chapter, US$10 for all three; the regular price for students is $15 per digital chapter).

- ***Data visualization*** *is closely related to the fields of information graphics, information visualization, scientific visualization, and statistical graphics. Until recently, the major forms of data visualization available in both business intelligence applications have included charts and graphs, as well as the other types of visual elements used to create scorecards and dashboards.*
- ***Scientific Visualization*** *is Visualization of scientific data.*

In this course we are more focus the use of data visualization for business intelligence purposes. Data discovery is an emerging style of BI that empowers Business people to conduct their own analysis without help from IT. For business people it is a brand new BI experience from going from data to Business insight.

There are three layers in data visualizations in a data Visualization Framework that helps us understand how data is processed for visualization.



Data Layer
- *Locating and obtaining data*
- *Importing data in proper format*
- *Relating data for proper correspondence*
- *Data analysis and aggregation*

Mapping Layer
- *Associating appropriate geometry with corresponding data channels*
- *Data analysis and algorithms (e.g. contouring)*

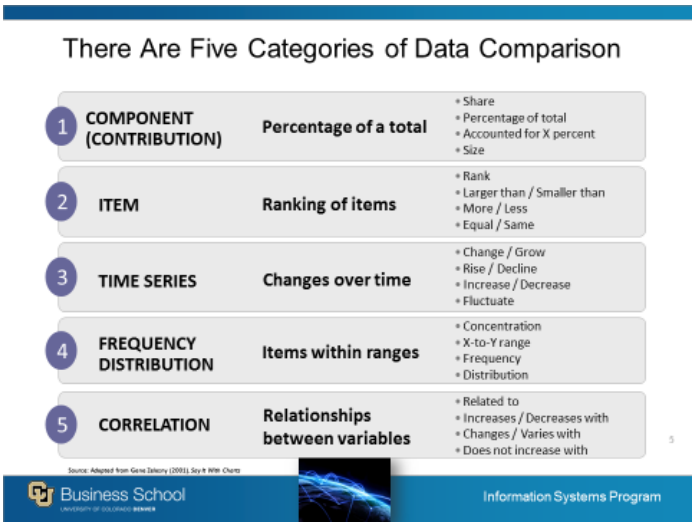
Graphics Layer
- *Conversion of geometry into displayable image*

•Decorations
• Managing interaction

You start with data. That data can be managed data from database or it can be your own data in spreadsheets. It's up to you. You start by familiarizing yourself with the data. You can look into the data and begin to understand any relationships. Next, you can apply visualizations to that data to help you discern patterns, problems, and trends. You can apply your own analysis to this by filtering the data, creating new calculations, and adding views of the data. But truthfully, the data discovery process is not so linear. At any time you can add new visualizations, add new combinations of data, and refine your data analysis.

> Until you arrive at the business insight. **Data Discovery** is an exploratory and iterative process. That exploratory process is highly visual. A good way to describe it is Visual OLAP. It's a better way for many people to understand data than the classic OLAP tabular format. And business people can do all this on their own without help from IT. At the same time, IT can make sure this can securely deploy at scale and with high performance.

See Best Practices in Data Visualization, by Vihao Pham 2014 for the types of data used for visualizations.



Excerpts from **Zelazny, G. Saying It With Charts: The executive's Guide to Visual Communication McGraw Hill Professional, March 15, 2001.**

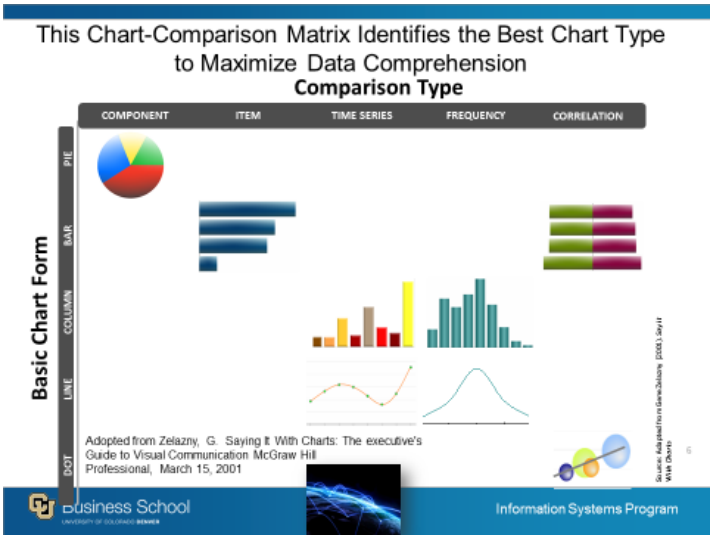
"The first step in identifying the type of visualization or chart type you want to add is identifying the message you're trying to convey. Zelazny identified 5 categories of data comparison. You want to look for these key words on the right to indicate the type of data comparison you're trying to make. …

- **The first type** of data comparison is **Component Contribution**, or percentage of a total. You would typically show this type of comparison in a 100 percent column chart or a pie chart in order to illustrate how one component compares to a whole. The idea is to easily identify one portion or slice of data. For example, you would want to use this type of comparison if you are trying to analyze how your sales compares to sales reported by your competitors in the industry, or maybe you want to analyze how much a certain activity contributes to the total cost of a product or service.
- **The second type** of data comparison is **Item Comparison**, or ranking of items. The goal is show a ranking or benchmark analysis. For example, you would use Item Comparison if you are trying to show where your company profit ranks as compared to your competitors. Maybe you

would use a deviation bar chart to illustrate favorable versus unfavorable conditions. For example, you might want to illustrate if departments reacted favorably or unfavorable to an increase in the company stock price. Did the company lose employees or gain some? This is just one example.

- **The third type** of data comparison is a **Time Series Comparison**, or changes over time. The goal is to illustrate changes or fluctuations over time. For example, you might want to simply analyze your sales over time to determine if sales continued to grow or decline after fluctuations seen in the marketplace.
- **The fourth type** of data comparison is a **Frequency Distribution**. The goal is to illustrate how many items fall into a series of progressive ranges. For example, you may want to analyze the average income of your employees compared with that in the industry over time. Or maybe you want to analyze over time where the largest growth in sales occurred.
- Finally, **the fifth type** of data comparison is **Correlation Comparison**. The goal is to analyze if the relationship between variables is relevant. For example, if you're in HR, you may notice that your services sector is experiencing high turnover which you suspect is a result of excessive travel. Therefore, you might want to compare the amount each service representative travels and number of months until the employee quits. Or maybe you want to analyze the impact of lowering prices on inventory sold. …."

Now that you have a better idea of each type of comparison, let's see how each graph corresponds to each type of comparison.



According to Zelazny, "it's not the data or measure that indicate which chart to use; it's what you say about it."

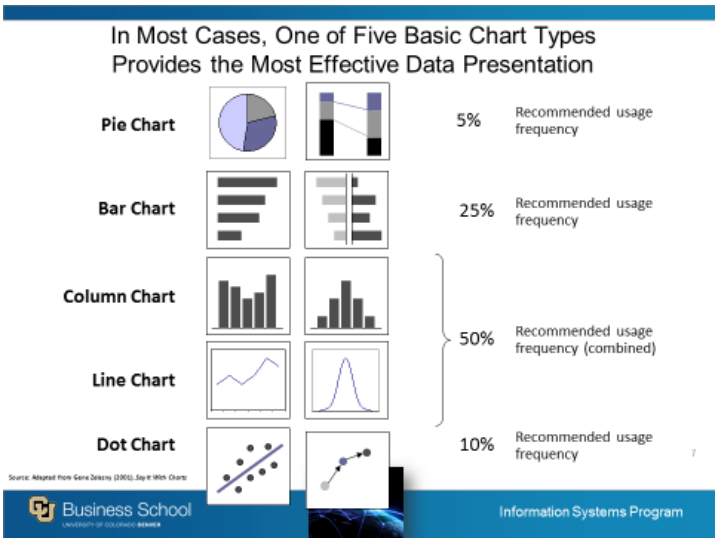
It is recommended that you use some form of these charts when performing a specific type of data comparison. You need to understand the message you're trying to communicate when you select your visualizations. You must let your message dictate your chart type. Different types of charts are appropriate for conveying different types of information.

- Line graphs are good for time-series data.
- Bar charts are good for depicting nominal or numerical data that can be easily categorized. **The bar chart** makes sense for an item comparison.
- Pie charts should be used for depicting proportions. You shouldn't use pie charts if the number of categories is very large.

- Scatter plots and bubble charts are good for illustrating relationships between two or three variables (bubble charts add a dimension via the size of the dot). Histograms are like bar charts, except they depict frequency distributions.

So you can see how these basic chart types can be extended to advanced visualizations and enable users to more easily analyze different dimensions of their data for a more extensive analysis. All of these visualizations are significantly more compact. In any instance where a graph needs to be used, you need to consider the type of comparison that you are making. When making a component comparison he recommends using a pie chart form.

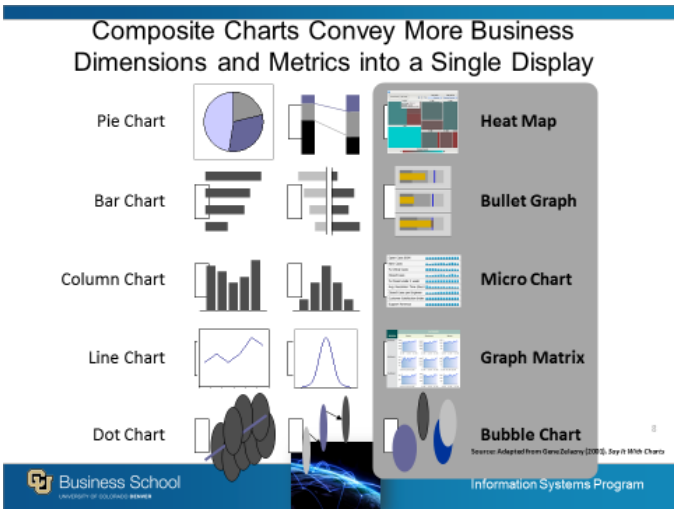
Both a Line and Column Chart are recommended for a time-series or frequency comparison. Finally, a bar or dot chart are recommended for a correlation comparison.



Excerpts from SHARDA, RAMESH; DELEN, DURSUN; TURBAN, EFRAIM, BUSINESS INTELLIGENCE AND ANALYTICS: SYSTEMS FOR DECISION SUPPORT, 10th Edition, © 2015. Used by permission of Pearson Education, Inc., New York, NY. All Rights Reserved.

- **Pie charts** are great for understanding a contribution to the whole. In general, pie charts are the least practical of the five chart forms. They should only be used 5% of the time. However, they are typically the most misused and abused. Pie charts are visually appealing, as the name implies, pie-looking charts. Because they are so visually attractive, they are often incorrectly used. Pie charts should only be used to illustrate relative proportions of a specific measure. For instance, they can be used to show relative percentage of advertising budget spent on different product lines or they can show relative proportions of majors declared by college students in their sophomore year. If the number of categories to show are more than just a few (say, more than 4), one should seriously consider using a bar chart instead of a pie chart.
- **Bar charts** are used to show a ranking of items from high to low or low to high. It's recommended you use these about 25% of the time. This is one of the most underutilized bar type and is often inaccurately replaced by a pie chart. Bar charts are among the most basic visuals used for data representation. Bar charts are effective when you have nominal data or numerical data that splits nicely into different categories so you can quickly see comparative results and trends within your data. Bar charts are often used to compare data across multiple categories such as percent advertising spending by departments or by product categories. Bar charts can be vertically or horizontally oriented. They can also be stacked on top of each other to show multiple dimensions in a single chart.

- **Column and line charts** are your work-horse and are used about 50% of the time. These are great for showing a distribution or time-series analysis of data. These are fairly easy to understand and make sense for the more common types of business analysis, such as trend analysis.
- Line charts are perhaps the most frequently used graphical visuals for time- series data. Line charts (or line graphs) show the relationship between two variables; they most often are used to track changes or trends over time (having one of the variables set to time on the *x*-axis). Line charts sequentially connect individual data points to help infer changing trends over a period of time. Line charts are often used to show time-dependent changes in the values of some measure such as changes on a specific stock price over a 5-year period or changes in the number of daily customer service calls over a month.
- **A dot chart** is used to show a relationship between two variables of data. Dot charts can be somewhat intimating and bar charts are usually overlooked and replaced with a column chart.
- The remaining 10% of the chart types are left for combination charts, such as a line chart with a column chart or a pie chart with a bar chart.
- However, when you use a chart, you want to ask yourself "does it makes sense" and "is the message clear"?



- **Heat maps** and highlight tables illustrate the comparison of continuous values across two categories using color.
- **The heat map** enables you to perform multiple component comparisons for numerous dimensions of data. Much like the pie chart, you are going use the Heat Map for component comparison to analyze a piece of data to the whole. They enable you to quickly grasp the state and impact of a large number of variables at once. The heat map enables users to quickly recognize complicated data relationships that may not be so obvious. The heat map is quite similar to a pie chart, but adds additional dimensionality that enables you to evaluate multiple variables simultaneously based on the size and color of each square. You would need to use multiple pie charts in order to convey this message. For instance, heat maps can be used to show segmentation analysis of the target market where the measure (color gradient would be the purchase amount) and the dimensions would be age and income distribution. Nowadays, one can find many other specialized graphs and charts that serve a specific purpose.
- **Bullet graphs** are ideal for evaluating an item comparison because they immediately show you a comparison to your target and whether or not you were below average, average, or above average. The VP of Sales might want to analyze which regions are doing well and which are doing poorly. Very quickly, with a bullet chart, they can compare a single value, sales and see which regions are meeting their targets, who is doing above average and who is doing poorly.
- **Micro charts:** The Microchart consists of compact representations of data that allow analysts to quickly visualize trends in data. A column Microcharts convey information so that a user can, at a

glance, determine the trend of a metric over time or how a metric is performing compared to forecasted figures.

- **Graph matrix:** A Graph Matrix visualization allows you to view your data in a chart containing multiple graphs. You can display the data in a chart containing one graph for every combination of the data you specify, allowing users to examine the data for each combination individually. With a graph matrix, you can slice your data, by displaying a graph for each combination of attribute elements in the rows and columns of the Graph Matrix visualization. For example, you can display the revenue data for each Region as a separate line graph, or display a bar graph containing store sales for each year
- Finally, **a bubble chart** is a type of Dot Chart. Much like the dot chart, you are making a comparison between two data elements. However, the bubble chart enables users to add a third dimension, which is based on the bubble size. A fourth dimension – time - can be added to fully animate the bubble chart, which we'll see in more detail in the next slide. Bubbles might be used for SKU analysis or trying to determine the impact price increases of your products have on consumer purchases.

- **Gantt charts** and **PERT** charts are good at illustrating project timelines and task dependencies.
- **Geographic maps**, of course, show geographic information.
- **Bullet graphs** show progress toward a goal. Bullet graph is essentially a variation of a bar chart.
- **Tree maps** are good for showing hierarchical information.
- Even though these charts and graphs cover a major part of what is commonly used in information visualization, they by no means cover it all.

See more information in the web resources listed below

- Trending Now: The Science of Data Visualization
- The Art Of Visualization: Squeeze Your Data Viz To The Ultimate
- Visualization: A Picture Tells A Thousand Words
- How P&G Presents Data to Decision-Makers
- When Data Visualization Works — And When It Doesn't
- Finding the Right Color Palettes for Data Visualizations