Content

- Introduction
- Types of Charts
- Data Tables
- Summarizing Data
- Cross-Tabulation
Data Visualization

- **Data visualization** - the process of displaying data (often in large quantities) in a meaningful fashion to provide insights that will support better decisions.
  - Data visualization improves decision-making, provides managers with better analysis capabilities that reduce reliance on IT professionals, and improves collaboration and information sharing.
Example 3.1: Tabular vs. Visual Data Analysis

- Tabular data can be used to determine exactly how many units of a certain product were sold in a particular month, or to compare one month to another.
  - For example, we see that sales of product A dropped in February, specifically by 6.7% (computed as $1 - \frac{B3}{B2}$). Beyond such calculations, however, it is difficult to draw big picture conclusions.
A visual chart provides the means to:

- easily compare overall sales of different products (Product C sells the least, for example);
- identify trends (sales of Product D are increasing), other patterns (sales of Product C is relatively stable while sales of Product B fluctuates more over time), and exceptions (Product E’s sales fell considerably in September).
A **dashboard** is a visual representation of a set of key business measures. It is derived from the analogy of an automobile’s control panel, which displays speed, gasoline level, temperature, and so on.

- Dashboards provide important summaries of key business information to help manage a business process or function.
Content

- Introduction
- Types of Charts
- Data Tables
- Summarizing Data
- Cross-Tabulation
Creating Charts in Microsoft Excel

- Select the *Insert* tab.
- Highlight the data.
- Click on chart type, then subtype.

- Use *Chart Tools* to customize.
Column and Bar Charts

- Excel distinguishes between vertical and horizontal bar charts, calling the former *column charts* and the latter *bar charts*.
  - A clustered column chart compares values across categories using vertical rectangles;
  - a stacked column chart displays the contribution of each value to the total by stacking the rectangles;
  - a 100% stacked column chart compares the percentage that each value contributes to a total.

- Column and bar charts are useful for comparing categorical or ordinal data, for illustrating differences between sets of values, and for showing proportions or percentages of a whole.
### Example 3.2: Creating a Column Chart

Highlight the range C3:K6, which includes the headings and data for each category. Click on the *Column Chart* button and then on the first chart type in the list (a clustered column chart).

<table>
<thead>
<tr>
<th>Racial/Ethnic Group and Gender</th>
<th>Total Employment</th>
<th>Officials &amp; Professionals</th>
<th>Technicians</th>
<th>Sales Workers</th>
<th>Office &amp; Clerical</th>
<th>Craft Workers</th>
<th>Operatives</th>
<th>Laborers</th>
<th>Service Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL EMPLOYEES</td>
<td>632,329</td>
<td>60,258</td>
<td>80,733</td>
<td>39,868</td>
<td>62,019</td>
<td>67,014</td>
<td>61,322</td>
<td>120,810</td>
<td>68,752</td>
</tr>
<tr>
<td>Men</td>
<td>349,353</td>
<td>41,777</td>
<td>39,792</td>
<td>19,848</td>
<td>23,727</td>
<td>11,293</td>
<td>55,853</td>
<td>84,724</td>
<td>44,736</td>
</tr>
<tr>
<td>Women</td>
<td>282,976</td>
<td>18,481</td>
<td>40,941</td>
<td>20,020</td>
<td>38,292</td>
<td>55,721</td>
<td>5,469</td>
<td>36,086</td>
<td>24,016</td>
</tr>
</tbody>
</table>

| WHITE                         | 407,545         | 51,252                   | 67,622      | 28,830        | 41,091           | 44,565        | 45,742     | 67,555   | 26,712         | 34,176         |
| Men                           | 237,516         | 36,536                   | 34,842      | 16,004        | 17,756           | 7,656         | 42,699     | 50,537   | 17,802         | 13,684         |
| Women                         | 170,029         | 14,716                   | 32,780      | 12,826        | 23,335           | 36,909        | 3,043      | 17,018   | 8,910          | 20,492         |

| MINORITY                      | 224,784         | 9,006                    | 13,111      | 11,038        | 20,928           | 22,449        | 15,580     | 53,255   | 42,040         | 37,377         |
| Men                           | 111,837         | 5,241                    | 4,950       | 3,844         | 5,971            | 3,637         | 13,154     | 34,187   | 26,934         | 13,919         |
| Women                         | 112,947         | 3,765                    | 8,161       | 7,194         | 14,957           | 18,812        | 2,426      | 19,068   | 15,106         | 23,458         |
Example 3.2: Creating a Column Chart

To add a title, click on the first icon in the Chart Layouts group. Click on “Chart Title” in the chart and change it to “EEO Employment Report—Alabama.” The names of the data series can be changed by clicking on the Select Data button in the Data group of the Design tab. In the Select Data Source dialog (see below), click on “Series1” and then the Edit button. Enter the name of the data series, in this case “All Employees.” Change the names of the other data series to “Men” and “Women” in a similar fashion.
Line charts provide a useful means for displaying data over time.

- You may plot multiple data series in line charts; however, they can be difficult to interpret if the magnitude of the data values differs greatly. In that case, it would be advisable to create separate charts for each data series.

Example 3.3: A Line Chart for China Export Data
Pie Charts

- A pie chart displays this by partitioning a circle into pie-shaped areas showing the relative proportion.

Example 3.4: A Pie Chart for Census Data

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Education Data</td>
<td>Not a High School Grad</td>
<td>High School Graduate</td>
<td>Some College No Degree</td>
<td>Associate's Degree</td>
<td>Bachelor's Degree</td>
<td>Advanced Degree</td>
</tr>
<tr>
<td>18 Marital Status</td>
<td>Never Married</td>
<td>4,120,320</td>
<td>7,777,104</td>
<td>4,799,872</td>
<td>1,828,392</td>
<td>5,124,648</td>
</tr>
<tr>
<td>20 Married, spouse present</td>
<td>15,516,160</td>
<td>35,382,720</td>
<td>18,084,352</td>
<td>8,346,624</td>
<td>19,154,432</td>
<td>9,523,712</td>
</tr>
<tr>
<td>21 Married, spouse absent</td>
<td>1,647,880</td>
<td>2,368,024</td>
<td>1,154,012</td>
<td>465,352</td>
<td>670,712</td>
<td>301,135</td>
</tr>
<tr>
<td>22 Separated</td>
<td>1,188,090</td>
<td>1,657,010</td>
<td>842,715</td>
<td>336,165</td>
<td>405,240</td>
<td>165,780</td>
</tr>
<tr>
<td>23 Widowed</td>
<td>5,145,883</td>
<td>4,670,488</td>
<td>1,765,010</td>
<td>556,657</td>
<td>977,544</td>
<td>475,195</td>
</tr>
<tr>
<td>24 Divorced</td>
<td>2,968,660</td>
<td>7,003,040</td>
<td>3,806,000</td>
<td>1,674,640</td>
<td>2,340,690</td>
<td>1,217,920</td>
</tr>
</tbody>
</table>
Data visualization professionals don't recommend using pie charts. In a pie chart, it is difficult to compare the relative sizes of areas; however, the bars in the column chart can easily be compared to determine relative ratios of the data.

- If you do use pie charts, restrict them to small numbers of categories, always ensure that the numbers add to 100%, and use labels to display the group names and actual percentages. Avoid three-dimensional (3-D) pie charts—especially those that are rotated—and keep them simple.
Area Charts

- An area chart combines the features of a pie chart with those of line charts.
  - Area charts present more information than pie or line charts alone but may clutter the observer’s mind with too many details if too many data series are used; thus, they should be used with care.

Example 3.5: An Area Chart for Energy Consumption
Scatter Charts

- Scatter charts show the relationship between two variables. To construct a scatter chart, we need observations that consist of *pairs* of variables.

Example 3.6: A Scatter Chart for Real Estate Data
A bubble chart is a type of scatter chart in which the size of the data marker corresponds to the value of a third variable; consequently, it is a way to plot three variables in two dimensions.

Example 3.7: A Bubble Chart for Stock Comparisons
Miscellaneous Excel Charts

- Stock chart
- Surface chart
- Doughnut chart
- Radar chart
Many applications of business analytics involve geographic data. Visualizing geographic data can highlight key data relationships, identify trends, and uncover business opportunities. In addition, it can often help to spot data errors and help end users understand solutions, thus increasing the likelihood of acceptance of decision models.

Companies like Nike use geographic data and information systems for visualizing where products are being distributed and how that relates to demographic and sales information. This information is vital to marketing strategies.

Geographic mapping capabilities were introduced in Excel 2000 but were not available in Excel 2002 and later versions. These capabilities are now available through Microsoft MapPoint 2010, which must be purchased separately.
Other Excel Data Visualization Tools

- Data bars
- Color scales
- Icon sets
- Sparklines
- Camera tool
Data bars display colored bars that are scaled to the magnitude of the data values (similar to a bar chart) but placed directly within the cells of a range.

- Highlight the data in each column, click the *Conditional Formatting* button in the *Styles* group within the *Home* tab, select *Data Bars*, and choose the fill option and color.
Example 3.8: Data Visualization through Conditional Formatting

- **Color scales** shade cells based on their numerical value using a color palette.
  - Color-coding of quantitative data is commonly called a **heatmap**.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>Product A</td>
<td>Product B</td>
<td>Product C</td>
<td>Product D</td>
<td>Product E</td>
</tr>
<tr>
<td>2</td>
<td>January</td>
<td>7792</td>
<td>5554</td>
<td>3105</td>
<td>3168</td>
</tr>
<tr>
<td>3</td>
<td>February</td>
<td>7268</td>
<td>3024</td>
<td>3228</td>
<td>3751</td>
</tr>
<tr>
<td>4</td>
<td>March</td>
<td>7049</td>
<td>5543</td>
<td>2147</td>
<td>3319</td>
</tr>
<tr>
<td>5</td>
<td>April</td>
<td>7560</td>
<td>5232</td>
<td>2636</td>
<td>4057</td>
</tr>
<tr>
<td>6</td>
<td>May</td>
<td>8233</td>
<td>5450</td>
<td>2726</td>
<td>3837</td>
</tr>
<tr>
<td>7</td>
<td>June</td>
<td>8629</td>
<td>3943</td>
<td>2705</td>
<td>4664</td>
</tr>
<tr>
<td>8</td>
<td>July</td>
<td>8702</td>
<td>5991</td>
<td>2891</td>
<td>5418</td>
</tr>
<tr>
<td>9</td>
<td>August</td>
<td>9215</td>
<td>3920</td>
<td>2782</td>
<td>4085</td>
</tr>
<tr>
<td>10</td>
<td>September</td>
<td>8986</td>
<td>4753</td>
<td>2524</td>
<td>5575</td>
</tr>
<tr>
<td>11</td>
<td>October</td>
<td>8654</td>
<td>4746</td>
<td>3258</td>
<td>5333</td>
</tr>
<tr>
<td>12</td>
<td>November</td>
<td>8315</td>
<td>3566</td>
<td>2144</td>
<td>4924</td>
</tr>
<tr>
<td>13</td>
<td>December</td>
<td>7978</td>
<td>5670</td>
<td>3071</td>
<td>6563</td>
</tr>
</tbody>
</table>
Example 3.8: Data Visualization through Conditional Formatting

- **Icon sets** provide similar information using various symbols such as arrows or stoplight colors.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>Product A</td>
<td>Product B</td>
<td>Product C</td>
<td>Product D</td>
<td>Product E</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>January</td>
<td>7792</td>
<td>5554</td>
<td>3105</td>
<td>3168</td>
</tr>
<tr>
<td>3</td>
<td>February</td>
<td>7268</td>
<td>3024</td>
<td>3228</td>
<td>3751</td>
</tr>
<tr>
<td>4</td>
<td>March</td>
<td>7049</td>
<td>5543</td>
<td>2147</td>
<td>3319</td>
</tr>
<tr>
<td>5</td>
<td>April</td>
<td>7560</td>
<td>5232</td>
<td>2636</td>
<td>4057</td>
</tr>
<tr>
<td>6</td>
<td>May</td>
<td>8233</td>
<td>5450</td>
<td>2726</td>
<td>3837</td>
</tr>
<tr>
<td>7</td>
<td>June</td>
<td>8629</td>
<td>3943</td>
<td>2705</td>
<td>4664</td>
</tr>
<tr>
<td>8</td>
<td>July</td>
<td>8702</td>
<td>5991</td>
<td>2891</td>
<td>5418</td>
</tr>
<tr>
<td>9</td>
<td>August</td>
<td>9215</td>
<td>3920</td>
<td>2782</td>
<td>4085</td>
</tr>
<tr>
<td>10</td>
<td>September</td>
<td>8986</td>
<td>4753</td>
<td>2524</td>
<td>5575</td>
</tr>
<tr>
<td>11</td>
<td>October</td>
<td>8654</td>
<td>4746</td>
<td>3258</td>
<td>5333</td>
</tr>
<tr>
<td>12</td>
<td>November</td>
<td>8315</td>
<td>3566</td>
<td>2144</td>
<td>4924</td>
</tr>
<tr>
<td>13</td>
<td>December</td>
<td>7978</td>
<td>5670</td>
<td>3071</td>
<td>6563</td>
</tr>
</tbody>
</table>
Sparklines are graphics that summarize a row or column of data in a single cell.

Excel has three types of sparklines: line, column, and win/loss.

- **Line sparklines** are clearly useful for time-series data.
- **Column sparklines** are more appropriate for categorical data.
- **Win-loss sparklines** are useful for data that move up or down over time.
Generally you need to expand the row or column widths to display them effectively. Notice, however, that the lengths of the bars are not scaled properly to the data; for example, in the first one, products D and E are roughly one-third the value of Product E yet the bars are not scaled correctly. So be careful when using them.
Content

- Introduction
- Types of Charts
- Data Tables
- Summarizing Data
- Cross-Tabulation
Data Queries: Tables, Sorting, and Filtering

- Managers often need to sort and filter data.
  - **Filtering** means extracting a set of records having certain characteristics.
- Excel provides a convenient way of formatting databases to facilitate analysis using sorting and filtering, called *Tables*. 
Example 3.10: Creating an Excel Table

- First, select the range of the data, including headers (a useful shortcut is to select the first cell in the upper left corner, then click Ctrl+Shift+down arrow, and then Ctrl+Shift+right arrow).

- Next, click Table from the Tables group on the Insert tab and make sure that the box for My Table Has Headers is checked. (You may also just select a cell within the table and then click on Table from the Insert menu.)

- The table range will now be formatted and will continue automatically when new data are entered.

- If you click within a table, the Table Tools Design tab will appear in the ribbon, allowing you to do a variety of things, such as change the color scheme, remove duplicates, change the formatting, and so on.
Suppose that in the *Credit Risk Data* table, we wish to calculate the total amount of savings in column C. We could, of course, simply use the function =SUM(C4:C428). However, with a table, we could use the formula =SUM(Table1[Savings]). The table name, Table1, can be found (and changed) in the Properties group of the Table Tools Design tab. Note that Savings is the name of the header in column C. One of the advantages of doing this is that if we add new records to the table, the calculation will be updated automatically,
The sort buttons in Excel can be found under the Data tab in the Sort & Filter group. Select a single cell in the column you want to sort on and click the “AZ down arrow” button to sort from smallest to largest or the “AZ up arrow” button to sort from largest to smallest. You may also click the Sort button to specify criteria for more advanced sorting capabilities.
Example 3.12  Sorting Data in the Purchase Orders Database

- Suppose we wish to sort the data by supplier. Click on any cell in column A of the data (but not the header cell A3) and then the “AZ down” button in the Data tab. Excel will select the entire range of the data and sort by name of supplier in column A.
Pareto Analysis

- An Italian economist, Vilfredo Pareto, observed in 1906 that a large proportion of the wealth in Italy was owned by a small proportion of the people.
- Similarly, businesses often find that a large proportion of sales come from a small percentage of customers, a large percentage of quality defects stems from just a couple of sources, or a large percentage of inventory value corresponds to a small percentage of items.
- A Pareto analysis involves sorting data and calculating cumulative proportions.
Example 3.13: Applying the Pareto Principle

75% of the bicycle inventory value comes from 40% (9/24) of items.
Filtering Data

- For large data files, finding a particular subset of records that meet certain characteristics by sorting can be tedious.
- Excel provides two filtering tools:
  - AutoFilter for simple criteria, and
  - Advanced Filter for more complex criteria.
Example 3.14: Filtering Records by Item Description

In the *Purchase Orders* database, suppose we are interested in extracting all records corresponding to the item Bolt-nut package.

Select any cell in the database

*Data > Sort & Filter > Filter*

Click on the dropdown arrow in cell D3.

Select Bolt-nut package to filter out all other items.
Example 3.14: Filter Results

- The filter tool does not extract the records; it simply hides the records that don’t match the criteria. However, you can copy and paste the data to another Excel worksheet, Microsoft Word document, or a Power-Point presentation.
- To restore the original data file, click on the drop-down arrow again and then click Clear filter from “Item Description.”

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Order N</th>
<th>Item N</th>
<th>Item Description</th>
<th>Item Cost</th>
<th>Quantity</th>
<th>Cost per ord</th>
<th>A/P Terms (Months)</th>
<th>Order Date</th>
<th>Arrival Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steelpin Inc.</td>
<td>A0123</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,250</td>
<td>$15,937.50</td>
<td>30</td>
<td>08/25/11</td>
<td>09/01/11</td>
</tr>
<tr>
<td>Steelpin Inc.</td>
<td>A0207</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,200</td>
<td>$15,750.00</td>
<td>30</td>
<td>09/01/11</td>
<td>09/10/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0223</td>
<td>4224</td>
<td>Bolt-nut package</td>
<td>$3.95</td>
<td>4,500</td>
<td>$17,775.00</td>
<td>30</td>
<td>10/15/11</td>
<td>10/20/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A1222</td>
<td>4111</td>
<td>Bolt-nut package</td>
<td>$3.55</td>
<td>4,250</td>
<td>$14,910.00</td>
<td>25</td>
<td>09/15/11</td>
<td>10/15/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A1444</td>
<td>4111</td>
<td>Bolt-nut package</td>
<td>$3.55</td>
<td>4,200</td>
<td>$14,910.00</td>
<td>25</td>
<td>09/20/11</td>
<td>10/10/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A1445</td>
<td>4111</td>
<td>Bolt-nut package</td>
<td>$3.55</td>
<td>4,200</td>
<td>$14,910.00</td>
<td>25</td>
<td>09/25/11</td>
<td>10/25/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A1449</td>
<td>4111</td>
<td>Bolt-nut package</td>
<td>$3.55</td>
<td>4,600</td>
<td>$16,330.00</td>
<td>25</td>
<td>10/05/11</td>
<td>10/19/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A1457</td>
<td>4569</td>
<td>Bolt-nut package</td>
<td>$3.50</td>
<td>3,900</td>
<td>$13,650.00</td>
<td>45</td>
<td>10/05/11</td>
<td>10/10/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A3487</td>
<td>4111</td>
<td>Bolt-nut package</td>
<td>$3.55</td>
<td>4,800</td>
<td>$17,040.00</td>
<td>25</td>
<td>09/05/11</td>
<td>09/20/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A5869</td>
<td>4111</td>
<td>Bolt-nut package</td>
<td>$3.55</td>
<td>4,585</td>
<td>$16,276.75</td>
<td>25</td>
<td>09/10/11</td>
<td>09/30/11</td>
</tr>
<tr>
<td>Steelpin Inc.</td>
<td>B0445</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,150</td>
<td>$15,562.50</td>
<td>30</td>
<td>09/03/11</td>
<td>09/11/11</td>
</tr>
</tbody>
</table>
Example 3.15: Filtering Records by Item Cost

- Suppose we wish to identify all records in the *Purchase Orders* database whose item cost is at least $200. First, click on the drop-down arrow in the Item Cost column and position the cursor over *Numbers Filter*. This displays a list of options. Select *Greater Than Or Equal To* ... from the list.
Example 3.15: Filtering Records by Item Cost

- The Custom AutoFilter dialog allows you to specify up to two specific criteria using “and” and “or” logic. Enter 200 in the box as shown; the tool will display all records having an item cost of $200 or more.
About the AutoFilter

- *AutoFilter* creates filtering criteria based on the type of data being filtered. If you choose to filter on Order Date or Arrival Date, the *AutoFilter* tools will display a different *Date Filters* menu list for filtering that includes “tomorrow,” “next week,” “year to date,” and so on.
- *AutoFilter* can be used sequentially to “drill down” into the data.
  - For example, after filtering the results by Bolt-nut package, we could then filter by order date and select all orders processed in September.
Content

- Introduction
- Types of Charts
- Data Tables
- Summarizing Data
- Cross-Tabulation
Statistics is both the science of uncertainty and the technology of extracting information from data.

A statistic is a summary measure of data.

Descriptive statistics are methods that describe and summarize data.

Microsoft Excel supports statistical analysis in two ways:
1. Statistical functions
2. Analysis Toolpak add-in
A frequency distribution is a table that shows the number of observations in each of several nonoverlapping groups.

- Categorical variables naturally define the groups in a frequency distribution.

To construct a frequency distribution, we need only count the number of observations that appear in each category.

- This can be done using the Excel COUNTIF function.
### Example 3.16: Constructing a Frequency Distribution for Items in the *Purchase Orders* Database

- List the item names in a column on the spreadsheet.
- Use the function `=COUNTIF($D$4:$D$97,cell_reference)`, where `cell_reference` is the cell containing the item name.

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airframe fasteners</td>
<td></td>
</tr>
<tr>
<td>Bolt-nut package</td>
<td></td>
</tr>
<tr>
<td>Control Panel</td>
<td></td>
</tr>
<tr>
<td>Door Decal</td>
<td></td>
</tr>
<tr>
<td>Electrical Connector</td>
<td></td>
</tr>
<tr>
<td>Gasket</td>
<td></td>
</tr>
<tr>
<td>Hatch Decal</td>
<td></td>
</tr>
<tr>
<td>Machined Valve</td>
<td></td>
</tr>
<tr>
<td>O-Ring</td>
<td></td>
</tr>
<tr>
<td>Panel Decal</td>
<td></td>
</tr>
<tr>
<td>Pressure Gauge</td>
<td></td>
</tr>
<tr>
<td>Shielded Cable/ft.</td>
<td></td>
</tr>
<tr>
<td>Side Panel</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airframe fasteners</td>
<td>14</td>
</tr>
<tr>
<td>Bolt-nut package</td>
<td>11</td>
</tr>
<tr>
<td>Control Panel</td>
<td>4</td>
</tr>
<tr>
<td>Door Decal</td>
<td>2</td>
</tr>
<tr>
<td>Electrical Connector</td>
<td>8</td>
</tr>
<tr>
<td>Gasket</td>
<td>10</td>
</tr>
<tr>
<td>Hatch Decal</td>
<td>2</td>
</tr>
<tr>
<td>Machined Valve</td>
<td>4</td>
</tr>
<tr>
<td>O-Ring</td>
<td>12</td>
</tr>
<tr>
<td>Panel Decal</td>
<td>1</td>
</tr>
<tr>
<td>Pressure Gauge</td>
<td>7</td>
</tr>
<tr>
<td>Shielded Cable/ft.</td>
<td>11</td>
</tr>
<tr>
<td>Side Panel</td>
<td>8</td>
</tr>
</tbody>
</table>
Example 3.16: Constructing a Frequency Distribution for Items in the *Purchase Orders* Database

- Construct a column chart to visualize the frequencies.
Relative Frequency Distributions

- Relative frequency is the fraction, or proportion, of the total.
- If a data set has \( n \) observations, the relative frequency of category \( i \) is:

\[
\text{relative frequency of category } i = \frac{\text{frequency of category } i}{n}
\]  

(3.1)

- We often multiply the relative frequencies by 100 to express them as percentages.
- A relative frequency distribution is a tabular summary of the relative frequencies of all categories.
Example 3.17: Constructing a Relative Frequency Distribution for Items in the *Purchase Orders* Database

- First, sum the frequencies to find the total number (note that the sum of the frequencies must be the same as the total number of observations, \( n \)).
- Then divide the frequency of each category by this value.

![Relative Frequency Table](image)
Frequency Distributions for Numerical Data

- For numerical data that consist of a small number of discrete values, we may construct a frequency distribution similar to the way we did for categorical data; that is, we simply use COUNTIF to count the frequencies of each discrete value.
Example 3.18: Frequency and Relative Frequency Distribution for A/P Terms

- In the *Purchase Orders* data, the A/P terms are all whole numbers 15, 25, 30, and 45.
Excel Histogram Tool

- A graphical depiction of a frequency distribution for numerical data in the form of a column chart is called a **histogram**.
- Frequency distributions and histograms can be created using the **Analysis Toolpak** in Excel.
  - Click the **Data Analysis** tools button in the **Analysis** group under the **Data** tab in the Excel menu bar and select **Histogram** from the list.
Specify the *Input Range* corresponding to the data. If you include the column header, then also check the *Labels* box so Excel knows that the range contains a label. The *Bin Range* defines the groups (Excel calls these “bins”) used for the frequency distribution.
Using Bin Ranges

- If you do not specify a Bin Range, Excel will automatically determine bin values for the frequency distribution and histogram, which often results in a rather poor choice.
- If you have discrete values, set up a column of these values in your spreadsheet for the bin range and specify this range in the Bin Range field.
Example 3.19: Using the *Histogram* Tool

- We will create a frequency distribution and histogram for the A/P Terms variable in the *Purchase Orders* database.
- We defined the bin range below the data in cells H99:H103 as follows:
  - **Month**
    - 15
    - 25
    - 30
    - 45
Example 3.19: Using the *Histogram* Tool

- *Histogram* tool results:

<table>
<thead>
<tr>
<th>Months</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>45</td>
<td>13</td>
</tr>
<tr>
<td>More</td>
<td>0</td>
</tr>
</tbody>
</table>

![Histogram chart](image-url)
For numerical data that have many different discrete values with little repetition or are continuous, a frequency distribution requires that we define by specifying

1. the number of groups,
2. the width of each group, and
3. the upper and lower limits of each group.

Choose between 5 to 15 groups, and the range of each should be equal.

Choose the lower limit of the first group (LL) as a whole number smaller than the minimum data value and the upper limit of the last group (UL) as a whole number larger than the maximum data value.

\[
\text{group width} = \frac{\text{UL} - \text{LL}}{\text{number of groups}}
\] (3.2)
Example 3.20: Constructing a Frequency Distribution and Histogram for Cost per Order

- The data range from a minimum of $68.75 to a maximum of $127,500; set the lower limit of the first group to $0 and the upper limit of the last group to $130,000.
- If we select 5 groups, using equation (3.2) the width of each group is $(130,000 - 0) / 5 = $26,000
Example 3.20: Constructing a Frequency Distribution and Histogram for Cost per Order

- Ten-group histogram
Example 3.21 Computing Cumulative Relative Frequencies

- Set the cumulative relative frequency of the first group equal to its relative frequency. Then add the relative frequency of the next group to the cumulative relative frequency.

- For example, the cumulative relative frequency in cell D3 is computed as \( \text{D2+C3} = 0.000 + 0.447 = 0.447 \).
Percentiles

- The $k^\text{th}$ percentile is a value at or below which at least $k$ percent of the observations lie. The most common way to compute the $k^\text{th}$ percentile is to order the data values from smallest to largest and calculate the rank of the $k^\text{th}$ percentile using the formula:

$$\frac{nk}{100} + 0.5$$  \hspace{1cm} (3.3)

- Statistical software use different methods that often involve interpolating between ranks instead of rounding, thus producing different results.
  - The Excel function `PERCENTILE.INC(array, k)` computes the $k^\text{th}$ percentile of data in the range specified in the array field, where $k$ is in the range 0 to 1, inclusive (i.e., including 0 and 1).
Examples 3.22 and 3.23: Computing Percentiles

- Compute the 90\textsuperscript{th} percentile for Cost per order in the Purchase Orders data.
  - Rank of \( k \textsuperscript{th} \) percentile = \( nk/100 + 0.5 \)
  - \( n = 94; \ k = 90 \)
  - For the 90\textsuperscript{th} percentile, the rank is
    \[ = 94(90)/100+0.5 = 85.1 \text{ (round to 85)} \]
  - Value of the 85\textsuperscript{th} observation = $74,375

- Using the Excel function \( \text{PERCENTILE.INC(G4:G97,0.9)} \), the 90th percentile is $73,737.50, which is different from using formula (3.3).
Example 3.24 Excel Rank and Percentile Tool

Data >
Data Analysis >
Rank and Percentile

90.3rd percentile
= $74,375
(same result as manually computing the 90th percentile)

The Excel value of the 90th percentile that was computed in Example 3.23 as $74,375 is the 90.3rd percentile value.
Quartiles break the data into four parts.
- The 25th percentile is called the first quartile, Q1;
- the 50th percentile is called the second quartile, Q2;
- the 75th percentile is called the third quartile, Q3; and
- the 100th percentile is the fourth quartile, Q4.

One-fourth of the data fall below the first quartile, one-half are below the second quartile, and three-fourths are below the third quartile.

Excel function QUARTILE. INC(array, quart), where array specifies the range of the data and quart is a whole number between 1 and 4, designating the desired quartile.
Example 3.25 Computing Quartiles in Excel

- Compute the Quartiles of the Cost per Order data
  - First quartile: =QUARTILE.INC(G4:G97,1) = $6,757.81
  - Second quartile: =QUARTILE.INC(G4:G97,2) = $15,656.25
  - Third quartile: =QUARTILE.INC(G4:G97,3) = $27,593.75
  - Fourth quartile: =QUARTILE.INC(G4:G97,4) = $127,500.00
Content

- Introduction
- Types of Charts
- Data Tables
- Summarizing Data
- Cross-Tabulation
Cross-Tabulations

- A **cross-tabulation** is a tabular method that displays the number of observations in a data set for different subcategories of two categorical variables.
  - A cross-tabulation table is often called a **contingency table**.
- The subcategories of the variables must be mutually exclusive and exhaustive, meaning that each observation can be classified into only one subcategory, and, taken together over all subcategories, they must constitute the complete data set.
Example 3.26: Constructing a Cross-Tabulation

- **Sales Transactions** database

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Cust ID</td>
<td>Region</td>
<td>Payment</td>
<td>Code</td>
<td>Source</td>
<td>Amount</td>
<td>Product</td>
<td>Time</td>
</tr>
<tr>
<td>4</td>
<td>10001</td>
<td>East</td>
<td>Paypal</td>
<td>93816545</td>
<td>Web</td>
<td>$20.19</td>
<td>DVD</td>
<td>22:19</td>
</tr>
<tr>
<td>5</td>
<td>10002</td>
<td>West</td>
<td>Credit</td>
<td>74093490</td>
<td>Web</td>
<td>$17.85</td>
<td>DVD</td>
<td>13:27</td>
</tr>
<tr>
<td>6</td>
<td>10003</td>
<td>North</td>
<td>Credit</td>
<td>64942368</td>
<td>Web</td>
<td>$23.98</td>
<td>DVD</td>
<td>14:27</td>
</tr>
<tr>
<td>7</td>
<td>10004</td>
<td>West</td>
<td>Paypal</td>
<td>70550957</td>
<td>Email</td>
<td>$23.51</td>
<td>Book</td>
<td>15:38</td>
</tr>
<tr>
<td>8</td>
<td>10005</td>
<td>South</td>
<td>Credit</td>
<td>35208817</td>
<td>Web</td>
<td>$15.33</td>
<td>Book</td>
<td>15:21</td>
</tr>
<tr>
<td>9</td>
<td>10006</td>
<td>West</td>
<td>Paypal</td>
<td>20978903</td>
<td>Email</td>
<td>$17.20</td>
<td>DVD</td>
<td>13:11</td>
</tr>
<tr>
<td>10</td>
<td>10007</td>
<td>East</td>
<td>Credit</td>
<td>80103311</td>
<td>Web</td>
<td>$177.72</td>
<td>Book</td>
<td>21:59</td>
</tr>
<tr>
<td>11</td>
<td>10008</td>
<td>West</td>
<td>Credit</td>
<td>14132583</td>
<td>Web</td>
<td>$21.76</td>
<td>Book</td>
<td>4:04</td>
</tr>
<tr>
<td>12</td>
<td>10009</td>
<td>West</td>
<td>Paypal</td>
<td>40128225</td>
<td>Web</td>
<td>$15.92</td>
<td>DVD</td>
<td>19:35</td>
</tr>
<tr>
<td>13</td>
<td>10010</td>
<td>South</td>
<td>Paypal</td>
<td>49073721</td>
<td>Web</td>
<td>$23.39</td>
<td>DVD</td>
<td>13:23</td>
</tr>
</tbody>
</table>

- Count the number (and compute the percentage) of books and DVDs ordered by region.
Cross-Tabulation Visualization: Chart of Regional Sales by Product

Regional Sales by Product

- East
- North
- South
- West

Legend:
- Book
- DVD
Excel provides a powerful tool for distilling a complex data set into meaningful information: **PivotTables**.

PivotTables allows you to create custom summaries and charts of key information in the data.

PivotTables can be used to quickly create cross-tabulations and to drill down into a large set of data in numerous ways.
Constructing PivotTables

Click inside your database

*Insert > Tables > PivotTable*

The wizard creates a blank PivotTable as shown.
Select and drag the fields to one of the PivotTable areas:

- Report Filter
- Column Labels
- Row Labels
- $\Sigma$ Values
Example 3.27 Creating a PivotTable

Initial PivotTable for Regional Sales by Product

The PivotTable defaults to a sum of the field in the Values area.

We seek a count of the number of records in each category.
Changing Value Field Settings

Active Field > Analyze > Field Settings

- Change summarization method in Value Field Settings dialog box
- Select Count
## Final Pivot Table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Count of Cust ID</strong></td>
<td><strong>Column Labels</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Row Labels</strong></td>
<td><strong>Book</strong></td>
<td><strong>DVD</strong></td>
<td><strong>Grand Total</strong></td>
</tr>
<tr>
<td>East</td>
<td>56</td>
<td>42</td>
<td>98</td>
</tr>
<tr>
<td>North</td>
<td>43</td>
<td>42</td>
<td>85</td>
</tr>
<tr>
<td>South</td>
<td>62</td>
<td>37</td>
<td>99</td>
</tr>
<tr>
<td>West</td>
<td>100</td>
<td>90</td>
<td>190</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td></td>
<td><strong>261</strong></td>
<td><strong>211</strong></td>
</tr>
</tbody>
</table>

Total: 472
## Modifying PivotTables

- Uncheck the boxes in the *PivotTable Field List* or drag the field names to different areas.
- You may easily add multiple variables in the fields to create different views of the data.
  - Example: drag the *Source* field into the *Row Labels* area
Example 3.28: Using the PivotTable Report Filter

- Dragging a field into the Report Filter area in the PivotTable Field list allows you to add a third dimension to your analysis.

Click the drop down arrow in cell B1; choose Credit:
PivotCharts visualize data in PivotTables.

They can be created in a simple one-click fashion.

- Select the PivotTable
- From the analyze tab, click *PivotChart*.
- Excel will display an *Insert Chart* dialog that allows you to choose the type of chart you wish to display.
By clicking on the drop-down buttons, you can easily change the data that are displayed. By filtering the data. Also, by clicking on the chart and selecting the *PivotChart Tools Design* tab, you can switch the rows and columns to display an alternate view of the chart or change the chart type entirely.
Slicers

- Excel 2010 introduced **slicers** — tools for drilling down to “slice” a PivotTable and display a subset of data.
- To create a slicer for any of the columns in the database, click on the PivotTable and choose *Insert Slicer* from the *Analyze* tab in the *PivotTable Tools* ribbon.
Example 3.30 Using Slicers

Cross-tabulation “sliced” by E-mail
PivotTable Dashboards

- The camera tool is useful for creating PivotTable-based dashboards.
- If you create several different PivotTables and charts, you can easily use the camera tool to take pictures of them and consolidate them onto one worksheet.
- In this fashion, you can still make changes to the PivotTables and they will automatically be reflected in the camera shots.