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.
Data Visualization via Conditional Formatting

- **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.
Data Visualization via Conditional Formatting

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

<table>
<thead>
<tr>
<th></th>
<th>Product A</th>
<th>Product B</th>
<th>Product C</th>
<th>Product D</th>
<th>Product E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Month</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>
Data Visualization via Conditional Formatting

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

<table>
<thead>
<tr>
<th>Month</th>
<th>Product A</th>
<th>Product B</th>
<th>Product C</th>
<th>Product D</th>
<th>Product E</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>7792</td>
<td>5554</td>
<td>3105</td>
<td>3168</td>
<td>10350</td>
</tr>
<tr>
<td>February</td>
<td>7268</td>
<td>3024</td>
<td>3228</td>
<td>3751</td>
<td>8965</td>
</tr>
<tr>
<td>March</td>
<td>7049</td>
<td>5543</td>
<td>2147</td>
<td>3319</td>
<td>6827</td>
</tr>
<tr>
<td>April</td>
<td>7560</td>
<td>5232</td>
<td>2636</td>
<td>4057</td>
<td>8544</td>
</tr>
<tr>
<td>May</td>
<td>8233</td>
<td>5450</td>
<td>2726</td>
<td>3837</td>
<td>7535</td>
</tr>
<tr>
<td>June</td>
<td>8629</td>
<td>3943</td>
<td>2705</td>
<td>4664</td>
<td>9070</td>
</tr>
<tr>
<td>July</td>
<td>8702</td>
<td>5991</td>
<td>2891</td>
<td>5418</td>
<td>8389</td>
</tr>
<tr>
<td>August</td>
<td>9215</td>
<td>3920</td>
<td>2782</td>
<td>4085</td>
<td>7367</td>
</tr>
<tr>
<td>September</td>
<td>8986</td>
<td>4753</td>
<td>2524</td>
<td>5575</td>
<td>5377</td>
</tr>
<tr>
<td>October</td>
<td>8654</td>
<td>4746</td>
<td>3258</td>
<td>5333</td>
<td>7645</td>
</tr>
<tr>
<td>November</td>
<td>8315</td>
<td>3566</td>
<td>2144</td>
<td>4924</td>
<td>8173</td>
</tr>
<tr>
<td>December</td>
<td>7978</td>
<td>5670</td>
<td>3071</td>
<td>6563</td>
<td>6088</td>
</tr>
</tbody>
</table>
Sparklines

Sparklines are graphics that summarize a row or column of data in a single cell.
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*. 
Creating an Excel Table

1. 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).

2. 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.
Table-Based Calculations

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 \(=\text{SUM}(C4:C428)\). However, with a table, we could use the formula \(=\text{SUM(Table1[Savings])}\). 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,
Sorting Data in Excel

- 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.
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.
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.
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.
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 PowerPoint presentation.

• To restore the original data file, click on the drop-down arrow again and then click Clear filter from “Item Description.”
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.
- 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.
Tabular vs. Visual Data Analysis

- it is difficult to draw big picture conclusions from table
- A visual chart provides the means to
  - compare overall sales of different products (Product C sells the least);
  - identify trends (sales of Product D are increasing)
  - find patterns (sales of Product C is relatively stable while sales of Product B fluctuates more over time)
  - See exceptions (Product E’s sales fell considerably in September)
Dashboards

- A **dashboard** is a visual representation of a set of key business measures.
  - automobile’s control panel: speed, gasoline level, temperature

- Dashboards provide summaries of key business information to help manage a business
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 are useful for comparing categorical or ordinal data, for illustrating differences between sets of values, and for showing proportions or percentages of a whole.

- vertical bar chart = *column chart*
- horizontal bar charts = *bar chart*
- 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.
Bad data?

Generations and “Selfies”

% saying they have shared a selfie

- Millennial: 55%
- Gen X: 24%
- Boomer: 9%
- Silent: 4%

Note: Based on all adults, N=1,821. Respondents who knew what a selfie was were asked if they had ever shared a selfie on a photo sharing or social networking site such as Facebook, Instagram or Snapchat.


PEW RESEARCH CENTER
Creating a Column Chart

Alabama Employment

- Officials & Managers
- Professionals
- Technicians
- Sales Workers
- Office & Clerical Workers
- Craft Workers
- Operatives
- Laborers
- Service Workers

- ALL EMPLOYEES
- Men
- Women
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>
<tr>
<td>WHITE</td>
<td>407,545</td>
<td>51,252</td>
<td>67,622</td>
<td>28,830</td>
<td>41,091</td>
<td>44,565</td>
<td>45,742</td>
<td>67,555</td>
<td>26,712</td>
</tr>
<tr>
<td>Men</td>
<td>237,516</td>
<td>36,536</td>
<td>34,842</td>
<td>16,004</td>
<td>17,756</td>
<td>7,656</td>
<td>42,699</td>
<td>50,537</td>
<td>17,802</td>
</tr>
<tr>
<td>Women</td>
<td>170,029</td>
<td>14,716</td>
<td>32,780</td>
<td>12,826</td>
<td>23,335</td>
<td>36,909</td>
<td>3,043</td>
<td>17,018</td>
<td>8,910</td>
</tr>
<tr>
<td>MINORITY</td>
<td>224,784</td>
<td>9,006</td>
<td>13,111</td>
<td>11,038</td>
<td>20,928</td>
<td>22,449</td>
<td>15,580</td>
<td>53,255</td>
<td>42,040</td>
</tr>
<tr>
<td>Men</td>
<td>111,837</td>
<td>5,241</td>
<td>4,950</td>
<td>3,844</td>
<td>5,971</td>
<td>3,637</td>
<td>13,154</td>
<td>34,187</td>
<td>26,934</td>
</tr>
<tr>
<td>Women</td>
<td>112,947</td>
<td>3,765</td>
<td>8,161</td>
<td>7,194</td>
<td>14,957</td>
<td>18,812</td>
<td>2,426</td>
<td>19,068</td>
<td>15,106</td>
</tr>
</tbody>
</table>

To add a title, click on the first icon in the Chart Layouts group. 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.
Line Charts

- Line charts provide a useful means for displaying data over time.
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.

• Restrict pie charts to small numbers of categories, always ensure that the numbers add to 100%, and use labels to display group names and percentages. Avoid rotated 3-D pie charts.
Area 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.
Scatter Charts

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

- A way to plot three variables in two dimensions.
Excel Camera Tool

• This tool allows you to create live pictures of various ranges from different worksheets that you can place on a single page, size them, and arrange them easily.

• They are simply linked pictures of the original ranges, and the advantage is that as any data are changed or updated, the camera shots are also.

  – To use the camera too, first add it to the Quick Access Toolbar (the set of buttons above the ribbon). From the File menu, choose Options and then Quick Access Toolbar. Choose Commands, and then Commands Not in the Ribbon. Select Camera and add it.
Exploring Data Using PivotTables

• 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.
Click inside your database

Insert >

Tables >

PivotTable

The wizard creates a blank PivotTable

Select and drag the fields to one of the PivotTable areas:

- Report Filter
- Column Labels
- Row Labels
- \( \Sigma \) Values
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
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

<table>
<thead>
<tr>
<th>Count of CustID</th>
<th>Column Labels</th>
<th>DVD</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td></td>
<td>56</td>
<td>42</td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>South</td>
<td></td>
<td>43</td>
<td>42</td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td>West</td>
<td></td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td>71</td>
<td>69</td>
</tr>
</tbody>
</table>

**Counts**
- East: 56, 42, 98
- South: 43, 42, 85
- West: 100, 90, 190
- Email: 13, 19, 32, 20, 27, 47
- Web: 35, 39, 74, 42, 47, 89

**Source**
- East
- South
- West
- Email
- Web

**PivotTable Fields**
- **Choose fields to add to report:**
  - Region
  - Payment
  - Transaction Code
  - Source
  - Product
  - Time of Day
- **MORE TABLES...**
- **Drag fields between areas below:**
  - **FILTERS**
  - **COLUMNS**
  - **VALUES**
  - Region
  - Count of Cust ID
  - Source
- **Defer Layout Update**
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 change the data that are displayed by filtering the data.

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

- **Slicers** “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.

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.
Confusing visualizations

Empty Classrooms
First-year law school enrollment

Source: The American Bar Association
Per capita consumption of cheese (US) correlates with Number of people who died by becoming tangled in their bed sheets

<table>
<thead>
<tr>
<th>Year</th>
<th>Per capita consumption of cheese (US)</th>
<th>Number of people who died by becoming tangled in their bed sheets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pounds (USDA)</td>
<td>(US) (CDC)</td>
</tr>
<tr>
<td>2000</td>
<td>29.8</td>
<td>327</td>
</tr>
<tr>
<td>2001</td>
<td>30.1</td>
<td>456</td>
</tr>
<tr>
<td>2002</td>
<td>30.5</td>
<td>509</td>
</tr>
<tr>
<td>2003</td>
<td>30.6</td>
<td>497</td>
</tr>
<tr>
<td>2004</td>
<td>31.3</td>
<td>596</td>
</tr>
<tr>
<td>2005</td>
<td>31.7</td>
<td>573</td>
</tr>
<tr>
<td>2006</td>
<td>32.6</td>
<td>661</td>
</tr>
<tr>
<td>2007</td>
<td>33.1</td>
<td>741</td>
</tr>
<tr>
<td>2008</td>
<td>32.7</td>
<td>809</td>
</tr>
<tr>
<td>2009</td>
<td>32.8</td>
<td>717</td>
</tr>
</tbody>
</table>

Correlation: 0.947091

http://www.tylervigen.com
Why Get Married?

Percent of general public saying each is a...

<table>
<thead>
<tr>
<th>Reason</th>
<th>Very important reason</th>
<th>Somewhat important reason</th>
<th>Not an important reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Love</td>
<td>88%</td>
<td>9%</td>
<td>2%</td>
</tr>
<tr>
<td>Making a lifelong commitment</td>
<td>81%</td>
<td>14%</td>
<td>4%</td>
</tr>
<tr>
<td>Companionship</td>
<td>76%</td>
<td>19%</td>
<td>3%</td>
</tr>
<tr>
<td>Having children</td>
<td>49%</td>
<td>30%</td>
<td>19%</td>
</tr>
<tr>
<td>A relationship recognized in a religious ceremony</td>
<td>30%</td>
<td>33%</td>
<td>35%</td>
</tr>
<tr>
<td>Financial stability</td>
<td>28%</td>
<td>48%</td>
<td>22%</td>
</tr>
<tr>
<td>For legal rights and benefits</td>
<td>23%</td>
<td>38%</td>
<td>37%</td>
</tr>
</tbody>
</table>


Note: “Don’t know/Refused” not shown