Introduction to Excel for Data Visualization

TechChange
Basic data cleaning troubleshooting .................................................. 04

Three useful excel formulas ........................................................................ 05

INDEX MATCH
Why use INDEX MATCH? ............... 06
INDEX MATCH Example ............... 07

VLOOKUP
Why use VLOOKUP? ............... 08
VLOOKUP Example ............... 09

COUNTIF
Why use COUNTIF? ............ 10
COUNTIF Example ............ 11

Types of Data Visualization in Excel .................................................. 13

Applying Data Visualization Principles in Excel .............................. 15
Basic data cleaning troubleshooting
# Basic data cleaning troubleshooting

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>You can’t run a function because your dataset contains merged cells.</td>
<td>Merge &amp; Center</td>
<td>Use the Merge &amp; Center button to unmerge cells in your dataset.</td>
</tr>
<tr>
<td>You have too much data contained in a single cell.</td>
<td>Text to Columns</td>
<td>Use the Text to Columns button to spread out data from one column across multiple columns. For example, you can separate a column of full names into separate first and last name columns.</td>
</tr>
<tr>
<td>You have to perform a series of actions in Excel over and over again.</td>
<td>Macros</td>
<td>Use macros to record a series of actions in Excel. Assign the series of actions (now a “macro”) a keyboard shortcut to run the series of actions automatically when you use that shortcut.</td>
</tr>
<tr>
<td>You want to automatically identify specific data in your dataset.</td>
<td>Conditional Formatting</td>
<td>Use Conditional Formatting to uniquely format specific data, such as blanks, duplicate values, etc.</td>
</tr>
<tr>
<td>You want to automatically select all of the data in your dataset that match a certain criteria.</td>
<td>Find &amp; Select</td>
<td>Use the Go to Special feature of Find &amp; Select to automatically select certain data.</td>
</tr>
<tr>
<td>You want to enter in a value into multiple cells at once.</td>
<td>Ctrl + Enter</td>
<td>If you have multiple cells selected, you can fill them all with a single value by typing that value in the formula bar and clicking “ctrl” and “enter”.</td>
</tr>
</tbody>
</table>
Three Useful Excel Formulas
Three Useful Excel Formulas

1. **INDEX MATCH**

\[ = \text{INDEX}( \text{array}, \text{MATCH}(\text{lookup_value}, \text{lookup_array}, [\text{match_type}]) \) \]

- **array**: The column that contains the value you want the function to return.
- **lookup_value**: The value you want Excel to lookup.
- **lookup_array**: The column that contains the value you want to lookup.
- **[match_type]**: The number that specifies how Excel looks up data.

**[match_type] can be one of three values:**

1. **1**: Excel will find the largest value that is less than or equal to the lookup_value. If you don’t enter anything, 1 is the default value.
2. **0**: Excel will find the first value that is exactly equal to the lookup_value.
3. **-1**: Excel will find the smallest value that is greater than or equal to the lookup_value.
Three Useful Excel Formulas

Why use INDEX MATCH?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The value you want to Excel to identify using INDEX MATCH</td>
<td>The value you know (the lookup_value)</td>
</tr>
<tr>
<td>2</td>
<td>--</td>
<td>XX</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>XX</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
<td>XX</td>
</tr>
<tr>
<td>5</td>
<td>--</td>
<td>XX</td>
</tr>
<tr>
<td>6</td>
<td>--</td>
<td>XX</td>
</tr>
</tbody>
</table>

INDEX MATCH allows you to identify a value by providing Excel with a known value to the right or left of it. Using INDEX MATCH to identify data is handy when you have a very large dataset, and manually identifying corresponding data would be an onerous process.

INDEX MATCH is unique from VLOOKUP because INDEX MATCH can identify values to the left of the known, corresponding value. VLOOKUP can only identify values that are to the right of the known, corresponding value.

INDEX MATCH Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>New Zealand</strong></td>
<td>209</td>
</tr>
<tr>
<td>2</td>
<td>Australia</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>Sri Lanka</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>South Africa</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>USA</td>
<td>48</td>
</tr>
</tbody>
</table>

Assume you have identified the value 209, in the second column, to be significant. Instead of manually looking up what the corresponding value is in column A, you want Excel to automatically look up this value using INDEX MATCH.

In this case, the array is A1:A5 because that column contains the value you want to identify (in this case, New Zealand). The lookup_value in this case is 209, and since this value is in column B, B1:B5 is our lookup_array. Finally, since we want an exact match to 209, the [match_type] is 0.

=INDEX(A1:A5, MATCH(209, B1:B5), 0)
Three Useful Excel Formulas

2. VLOOKUP

\[=VLOOKUP(\text{lookup\_value}, \text{table\_array}, \text{col\_index\_num}, [\text{range\_lookup}])\]

- **lookup\_value**: The value you would like Excel to search for in the first column of the array.
- **table\_array**: The table of information in which data is looked up.
- **col\_index\_num**: The column number that contains the value you'd like Excel to return.
- **[range\_lookup]**: A logical value that specifies how Excel looks up data.

**[range\_lookup] can be one of two values:**

- **FALSE**: Excel will find an exact match of the lookup\_value. If an exact match is not found, the #N/A error value is returned.
- **TRUE**: Excel will find an approximate match of the lookup\_value (the next largest value that is less than the lookup\_value). If you leave this field blank, TRUE is the default value.
Three Useful Excel Formulas

Why use VLOOKUP?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The value you know (the lookup_value)</td>
<td>xx</td>
</tr>
<tr>
<td>2</td>
<td>--</td>
<td>xx</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>xx</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
<td>xx</td>
</tr>
<tr>
<td>5</td>
<td>--</td>
<td>xx</td>
</tr>
</tbody>
</table>

VLOOKUP allows you to identify the value in any column to the right of first column simply by providing excel with the value in the first column and the corresponding range of data.

INDEX MATCH is unique from VLOOKUP because VLOOKUP can only identify values to the right of the known, corresponding value. As such, VLOOKUP is most useful when you have a large dataset in which a value in the first column is the known value.

Say you want to identify the value in column C that corresponds to “New Zealand” in column A. In this case, the lookup_value is A1 because that cell contains the known value (in this case, New Zealand). The table array in this case in A1:C5. Since we want to identify the corresponding value in column C, the col_index_num is 3 (since column C is the third column). Finally, since we want an exact match to 209, the [range_lookup] is FALSE.

=VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])
=VLOOKUP(A1, A1:C5, 3, FALSE)
3. COUNTIF

=COUNTIF (range, criteria)

- The group of cells you want to count
- The rule for which cells will be counted

The criteria field can be a number, expression, cell reference, or text string. Use quotation marks for criteria that reference a text value.
COUNTIF allows you to count the number of entries in your dataset that meet a certain criteria. It is most useful when you have a large data set and manually counting data that meet a certain criteria would be onerous.

**COUNTIF Example**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jordan</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Madagascar</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Bahrain</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Belarus</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Bulgaria</td>
<td>0</td>
</tr>
</tbody>
</table>

Say you want to count the number of countries that have 0 associated with them. Since the 0 values are in column B, in order to count them the range must be B1:B5. The criteria for being counted is that the value is equal to 0. However, when we type = without quotation marks, Excel assumes we are trying to enter a new formula. We need to put quotation marks around the equals sign in order to let Excel know that it is a part of the same COUNTIF formula. Then we’ll use an ampersand to let Excel know what the value should be equal to (in this case, 0).

```
=COUNTIF ( range, criteria )
=COUNTIF ( B1:B5      “”&0    )
  =3
```
TYPES OF DATA VISUALIZATIONS IN EXCEL
Type of Data Visualization in Excel

**Sparklines**

Sparklines are mini graphs in a single cell that represent a selected row of data. They are best used when you would like to **quickly compare** several entries to one another.

**Bar/column charts**

Bar/column charts are best used to compare **values across categories**.

**Dot plots**

Dot plots are best used to show the relationship between **sets** of values.

**Pie charts**

Pie charts are best used when attempting to show a **part-to-whole** relationship.

**Line graphs**

Line graphs are best used to show **trends over time**.

**PivotCharts**

PivotCharts work in tandem with **PivotTables**. PivotCharts are a good option for working with data that you want to be able to adjust and automatically update (through PivotTables).

**Slicers**

Slicers make it easy to visually filter PivotTables (and associated PivotCharts). Slicers are a good option for creating an **interactive dashboard**, as they visually display what information a PivotChart is displaying.
APPLYING DATA VISUALIZATION PRINCIPLES IN EXCEL
Oftentimes, auto-generated charts in Excel do not adequately highlight the important aspects of the data. A common issue is the fact that Excel automatically sets the numeric range of charts, and this automatic range does not always make sense within the greater context of the dataset.

To **manually adjust a numeric axis**, right click on the axis. A “Format Axis” window should pop up. Under “Axis Options” you can change the minimum and maximum values of the axis as well as the units.

Another option to keep in mind is creating **multiple graphs** to convey a single key takeaway. While a single graph may be sufficient in some cases, if creating an additional graph will make the key takeaway more clear, it is an option worth considering.

Excel automatically sets the gridlines, colors, and borders of generated charts. Often this adds unnecessary clutter.

To change the **color scheme** of a chart, click on the chart. A “Chart Tools” ribbon should appear. In the “Design” tab, click on the “Change Colors” button, and select a scheme that makes sense with the data.

To adjust the **gridlines**, of a chart, simply right click on them. A formatting window should pop-up with an option to change the color of the gridlines.

To adjust the **border** of a chart, click on the chart. A “Chart Tools” ribbon should appear. In the “Format” tab, click on “Shape Outline” to change the color of the border, or choose “No Outline”.

An easy way to integrate text in Excel is to add **data labels**. This often eliminates the need for a legend.

Here is one way to quickly add data labels to a chart:

1. Click the chart.
2. A “Chart Tools” ribbon should appear. In the “Design” tab, click the “Add Chart Element” button.
3. Hover over “Data Labels”.
4. Click “More Data Label Options”.
5. A “Format Data Labels” window should pop-up. Under “Label Options”, select what the labels should contain (e.g. the value, the series name, etc.)
take the full course at
www.course.tc/011a

for more information e-mail:
info@techchange.org