Excel PivotTables and Epi Curves

Adapted from the "Introduction to MS Excel 2010 for Epidemiologists: Self-Study Guide" developed by the Public Health Agency of Canada's Centre for Emergency Preparedness and Response's Training and Development Unit, presented at the 'Epidemiology in Action 2020' course.

Table of Contents

Basics of creating a PivotTable

PivotTables are used to quickly and easily summarize, analyze, explore and present data. The following example will take you through the basics of PivotTable creation.

### Step 1: Format the data to be used in the PivotTable
1. It's good practice to use a copy of your master data just in case something goes wrong.
2. It is important to check the consistent formatting of your data (e.g., date fields, etc.).
3. All the columns must be titled. Ensure that variable names are listed in Row 1.
4. If you have a large worksheet with many variables, delete any unnecessary columns/variables.
   *Tip! Keep it simple!!*

Step 2: Create a PivotTable
Ideally, you should start by creating a table with your data so the pivot table refers to the table as a whole rather than just a range of data. By doing this, it will ensure that any new entry in the table is included in the pivot table without having the readjust the range of data.
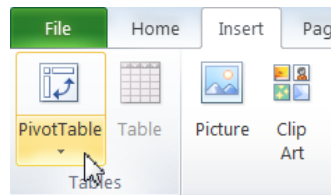
1. Highlight the data you want to analyze (or if you are starting from a table, click on any of the cell within the table)

2. Select the **PivotTable** button in the **Tables** menu under the **Insert** tab

3. Double check that your **Table/Range** of data is correct (i.e., that all data that you wish to summarize in your table has been highlighted).

4. Select **New Worksheet** (or **Existing Worksheet** if you are already working in a PivotTable). Click **OK**.
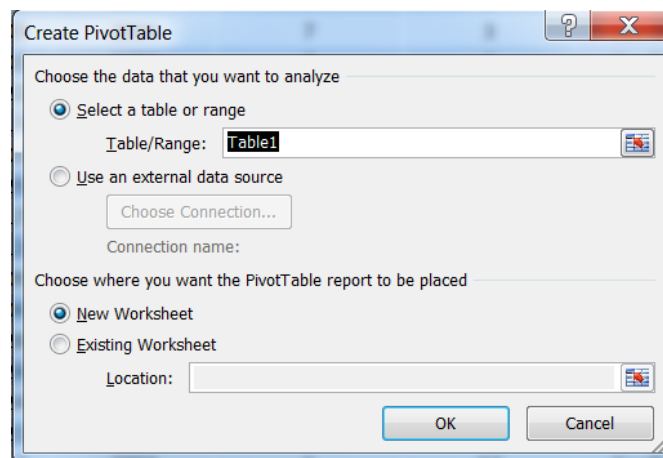
5. You will be taken to a new sheet with the PivotTable shell. There will also be a **PivotTable Field List** box. This contains all the fields for your selected dataset and an 'areas' section. You can pick any one of the items and drag it into one of the four areas in the PivotTable (**Report Filter**, **Row Labels**, **Colum Labels**, **Values**) or into the 'areas' under the field list.
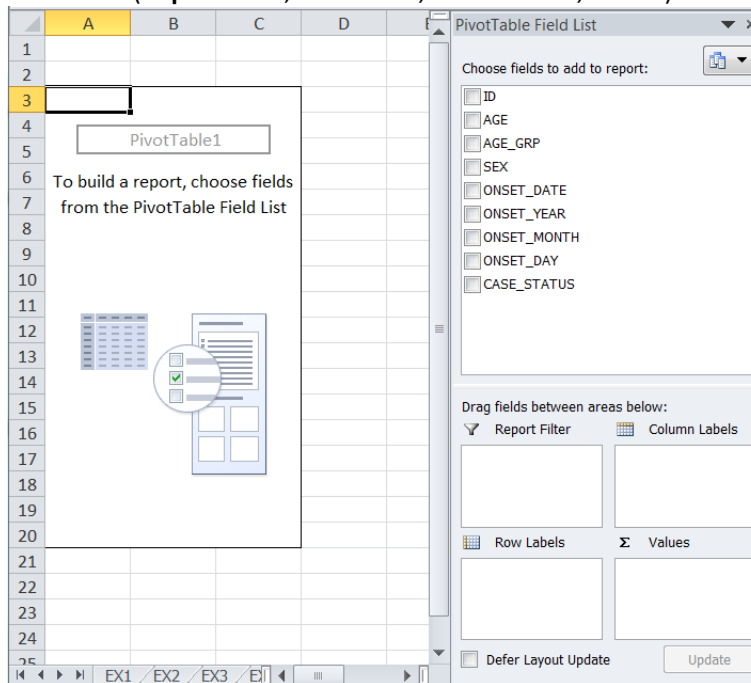

a. For your **Values** variable, use a variable that is complete in your master list (i.e., no blanks) and is unique for each case in your dataset. Case ID or Personal Health Number often works well.

b. Pick variables from the **Field List** and drag into **Row Labels** and **Column Labels** to summarize your data into columns and rows.

c. **FOR EXAMPLE:** using AGE_GRP in **Row Labels,** SEX in **Column Labels**, and ID in the **Values** field of a sample dataset provides the following PivotTable:

Count of ID	Column Labels		
Row Labels	F	M	Grand Total
15-19	1	1	2
20-24	5	5	10
25-29	1	3	4
30-34	3	3	6
35-39	3	4	7
40-44	2	5	7
45-49		7	7
50-54	1	3	4
5-9	1		1
60-64	1		1
65-69	1		1
Grand Total	19	31	50

*Note: When selecting your **Values** ensure that the data are complete and that there are no blanks because this is the data that is used to summarize the 'count'*

*Note: Make sure that you are 'counting' and 'not adding' the values (e.g., the case ID#'s). If Excel is adding the values, right-click on the **Values** area (the part of the table that contains the data), select the **value field settings** tab and ensure that 'count' is selected. You can also verify this by looking in the Σ value section in the PivotTable Field List.*
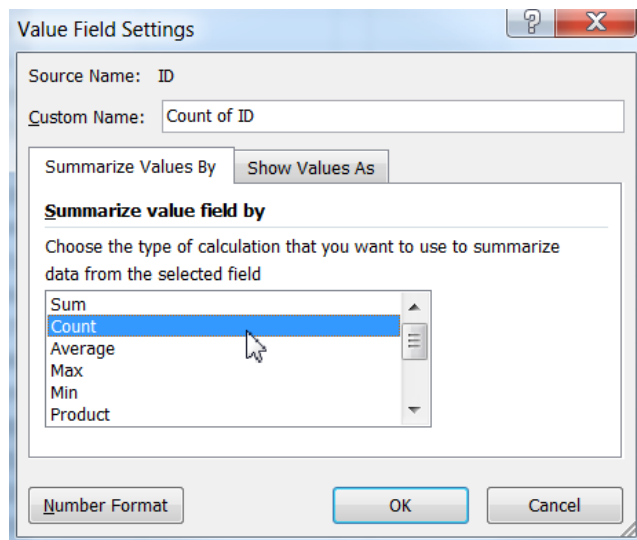
6. You can also use PivotTables to calculate proportions. Follow the same steps as above, then:
   a. Right click in the **Values** area
   b. Click on **Value Field Settings** then **Show values as** then select "% of Grand Total"
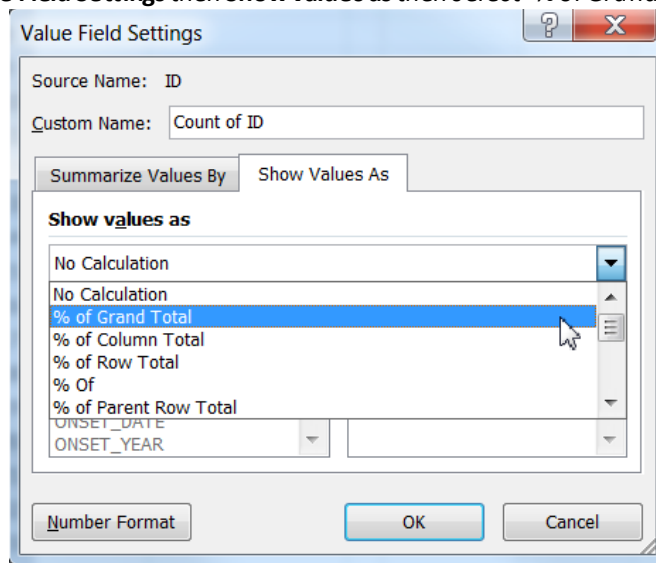
7. To remove an item from the PivotTable, simply select it and drag it out of the table.

8. If you need to reorder data in the PivotTable, you can do this very simply:
   a. Right click on row or column heading within PivotTable that you would like to move/reorder.
   b. Select **Move** > Select option that best fits with desired order.

   FOR EXAMPLE: To reorder the AGE_GRP data of the sample PivotTable:
   a. Ensure that AGE_GRP is in the **Row Labels** area PivotTable
   b. Right click on age group 5 to 9
   c. Select **Move** > **Move "5 to 9" to Beginning** (should now be in proper order and will remain in this order even if you remove it from PivotTable and re-enter)

9. If you need to add a field to a given variable in a pivot table:
   FOR EXAMPLE, in the above table there are individuals between the ages of 10-14 and therefore the 10-14 age category is not visible in the table. To add the category:
   a. Click on one of the existing age category labels in the pivot table.
   b. In the Options tab click on **Fields, Items, & Sets** in the **Calculations** menu and select **Calculated Item** in the drop box.
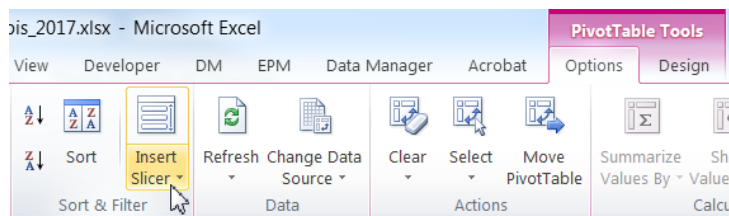
c. In the pop up box select the AGE_GRP variable from the fields list on the left hand side. In box titled name type 10 - 14 and then select **Add** and **OK**.
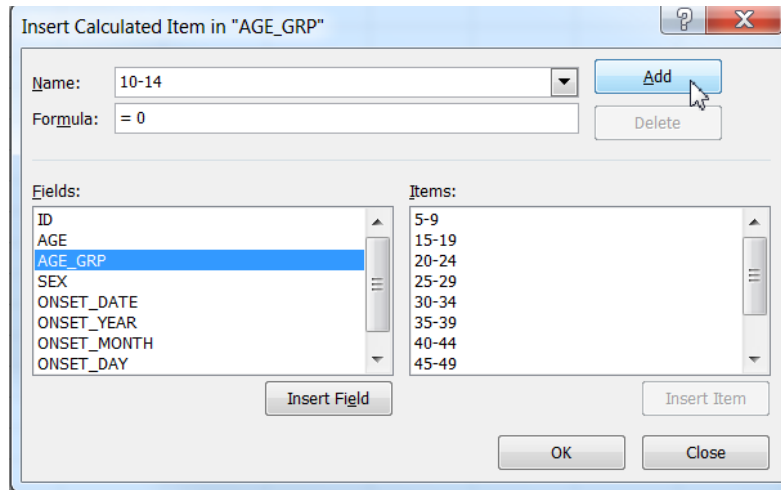
d. You can reorder the categories using **Move** option describe above.

Tip! You can also filter the data by clicking on the arrow beside the variable name within the PivotTable. By clicking on the arrows beside your variables you can add/remove options (e.g. choose to see results only for confirmed cases, only males etc.).

Slicer is a new feature of MS Excel 2010. Slicers are easy-to-use filtering components that enable you to quickly filter the data in a PivotTable report. It's an interesting alternative to "report filter" within the PivotTable.

1. Select **Insert Slicer** in the **Sort & Filter** menu of the **Options** tab.
2. In the pop-up box, select the variable the variable you want to filter. The example below show the use of the Slicer feature to quickly look the age and sex distribution based on case status.

Count of ID	Column Labels		
Row Labels	F	M	Grand Total
5-9	1		1
10-14	0	0	0
15-19	1	1	2
20-24	2	2	4
25-29		2	2
30-34	2	2	4
35-39	2	3	5
40-44	1	3	4
45-49		4	4
50-54		2	2
65-69	1		1
Grand Total	**10**	**19**	**29**

CASE_STATUS

Confirmed

Probable

Suspect

5

One of the very useful functions of PivotTables is the ability to group data. This is great for grouping individual days into weeks or ages into age categories. When grouping, you need to ensure that there is no missing data in the column being grouped.

FOR EXAMPLE: The following steps take you through the steps required to group a hypothetical AGE variable into 10-year age groups.

1. Ensure that the PivotTable has the AGE variable in the **Row Labels.**

2. Right click anywhere in the AGE field.

3. Click **Group**

4. In the dialog box, deselect the two check boxes for **Starting at** and **Ending at**.

5. You may need to adjust the starting date so that it starts at age 0 and ends at age 69.

6. For **By**, enter **10** (to create 10 year age groups). Click **OK**.
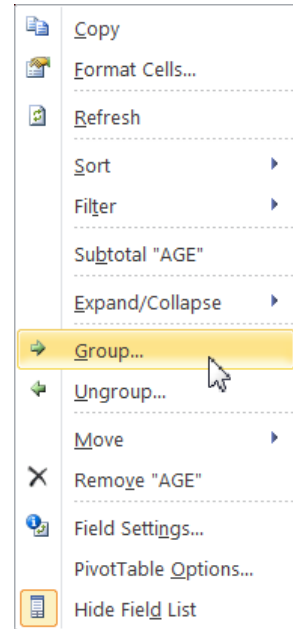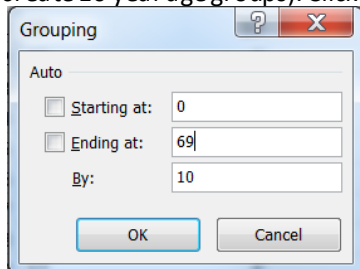
7. To ungroup, click on the **Ungroup** button under **Options** or right click on any cell and click **Ungroup**.

## Some other tips and tricks for working with PivotTables

Creating multiple PivotTables

You can have more than one PivotTable on each Excel worksheet, as well as multiple PivotTables on multiple sheets. To create a second PivotTable on your worksheet, simply select the existing PivotTable, copy and paste it into the worksheet.

Refreshing your data

When you make changes to your master linelist, you will need to refresh each of the PivotTables. **Be careful, if you don't refresh, your PivotTable will still be summarizing old data!** To refresh your PivotTable:

    a.   Make sure that you have saved your updated linelist.
    b.   Go to PivotTable toolbar under **options** and click on **Refresh** in the **Data** menu or right-click data and select **Refresh**.
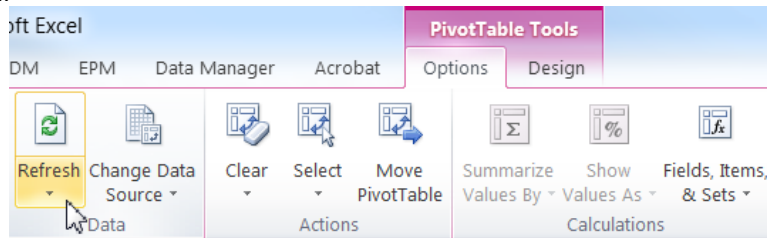
Adding rows or columns of data to your linelist

If you are working from a table, any row added to your linelist will be integrated in the Pivot Table. If you are not working from a table, you will need to expand the reference area for the PivotTable. This is why it is strongly suggested to always start from a table.

Note: the refresh button will only work for any changes made to data within your reference area.

To expand the reference area for your PivotTable:
1. Go to the **options** tab and select **Change data source.**
2. You will return to your linelist.
3. Highlight the area including the new row and click **OK.**

Generate a linelist of specific cases

Let's say you would like to see a linelist of <u>only Confirmed </u>cases:
1. Go to the PivotTable and double click on the data cell containing the count of Confirmed cases w.
2. A linelist of the specific cases will be generated in a new worksheet.

Creating Charts with PivotTables

1. Quick method
    a. Click on any cell in the PivotTable on which you wish to base the chart.
    b. Under the **Options** tab, select **PivotChart** in the **Tools** menu**.**
    c. Select the chart type you want and click **OK.**
    d. Continue formatting chart.

2. Alternate method
    a. Copy the data you want to graph from your PivotTable.
    b. Paste special (values and numbers) into a new excel worksheet.
    c. Decide what to do with unknowns (typically exclude and add a note at the bottom of table).
   d. Select your data and click on graph function - do not include totals in your data selected.

Creating Epidemic Curves with Pivot Tables

## Step 1: Create a Pivot Table containing the data you want in your epi-curve

1. Follow steps as outlined above for creating a PivotTable (pg. 2).
2. Drag Onset Date into **Row Labels** and ID into the **Values** field. Put whichever variable you would like cases to be stratified by into the **Column Labels** area (e.g., P/T, Case Classification, Symptom Status)

Step 2: Show dates between cases and regroup date

1. Ensure that the checkbox **Show items with no data** is checked. To do this, right click on any of the data in the PivotTable and select **Field Settings**. In the Layout & Print tab, check **Show items with no data.**
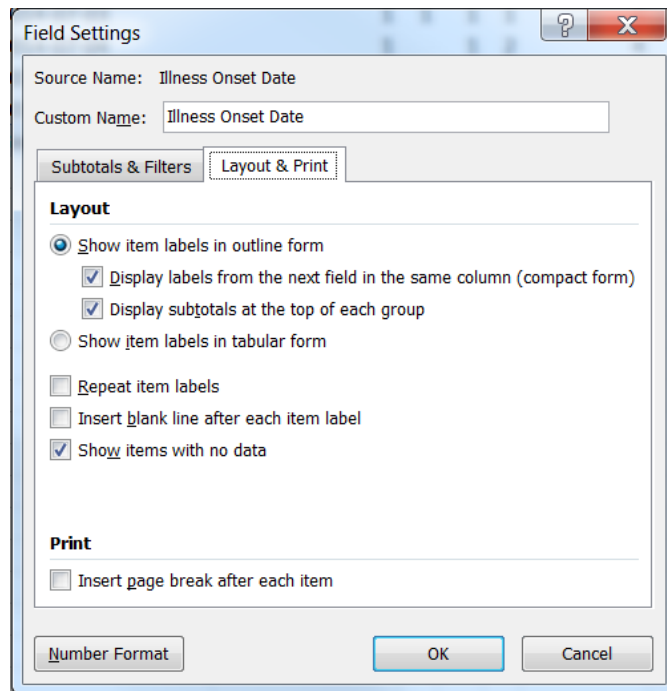
2. In order to show the dates between cases right-click on any of the dates in the PivotTable, and select **Group.**

3. In the grouping window, unclick all selected options (highlighted in navy blue) and choose **Days** and set the **Number of days: to 1**. (Optional: also select **Months** and then **Years**.) Click **OK.**
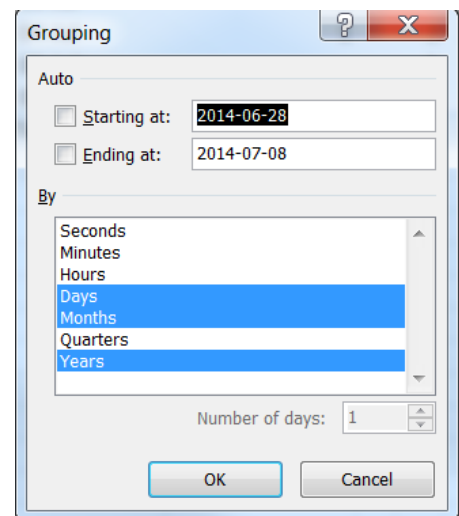
4. You will notice that all dates from January onwards are showing up. These can be hidden so that they do not show up in the Epi Curve. Scroll down to the dates where there are cases and highlight the area with data (in addition, a few days before and after the first and last illness onset dates can be highlighted). For example, if cases fell between June 28 and July 5, highlight the rows in the pivot table from June 23 and July 10. Right click the highlighted area and select **Filter > Keep Only Selected Items.** Scroll up to see the data.
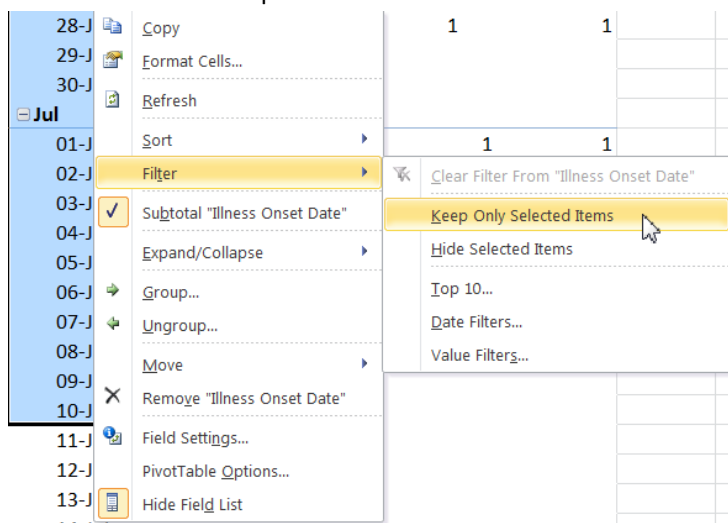
Step 3: Create the chart

1. Highlight the PivotTable. Select the **Insert** tab on the ribbon. In the **Charts** section, select **Column** and then **Stacked Column** (second option in the 2-D Column section). A graph should appear. Resize as needed.
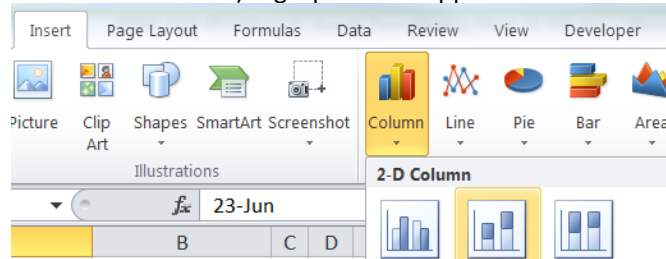
2. To remove the spaces between the bars, right click on any bar on the graph. Select **Format Data Series**. Under the **Series Options** and the **Gap Width** section, drag the slider to the very left to the **No Gap** side and click Close (alternatively, in the box below the Gap Width section, delete 150% and enter 0%).
3. In the **Analyze** tab of the Pivot Chart Tools ribbon click **Field Buttons > Hide All.**
4. If Months and Years were also selected in the Grouping window and are not desired: To remove the **year** or **month** titles from the bottom of the epi curve, drag the **Years** and **Month** variable out of the **Axis Fields** box in the **Pivot Table Field List.**

Formatting Guidelines

Suggested formatting styles and guidelines for epi-curves are listed below.

- Title
  - Use consistent capitalization from figure to figure
  - Italicize Latin names
  - If a title is more than a single line of text use line breaks
- Axes
  - Use horizontal or vertical alignment for x- and y-axis title and labels, as appropriate
  - Select appropriate scales
  - Choose a sensible periodicity for x-axis labels
  - Use actual dates on the x-axis not just numbered weeks
  - Centre x-axis labels with histogram bars
  - Use a consistent font type for all labels
- Graph Area
  - Do not include a top border line on around the graph area
  - Do not include coloured backgrounds
  - Do not include space between bars
  - Use stacked bars when using more than one category of cases. Ensure colour scheme/patterns are distinct when printed in black and white.
  - Outline bars with a white or black line so they look more defined
- Legend
  - Place within the graph area
  - The ordering of categories within the legend should be identical to the order used in the curve

A formatted version of the epidemic curve can be seen below.

Figure 1. Number of cases of E.coli O157 in Canada, June 23 to July 10, 2014