

# Excel for Data Cleaning and Management



## **Background Information**

This workshop is designed to teach skills in Excel that will help you manage data from large imports and save them for further use in SPSS.

## **Required Skills**

### **Agenda**

- Importing data sets
- Create a copy of a worksheet
  - Renaming worksheets
  - Hide columns and rows
- Eliminating unwanted data
- Sorting data
- Quick Analysis
  - Sparklines
- Split Screens
- Saving data for further use in SPSS
- What we learned

## **Creating a copy of your worksheet**

Before analyzing or cleaning your data in Excel, you should always create a copy of your raw data worksheet before making any changes.

To create a copy of a worksheet:

1. Right click on your worksheet tab
2. Select *Move or Copy*
3. The *Move or Copy* dialogue box will appear
4. Select what workbook you want to move your worksheet to (Excel will list only workbooks that are currently open)
5. Select the location where you want your copy worksheet to appear
6. Make sure to click *Create a Copy*, otherwise it will just move the selected worksheet
7. Click *OK*

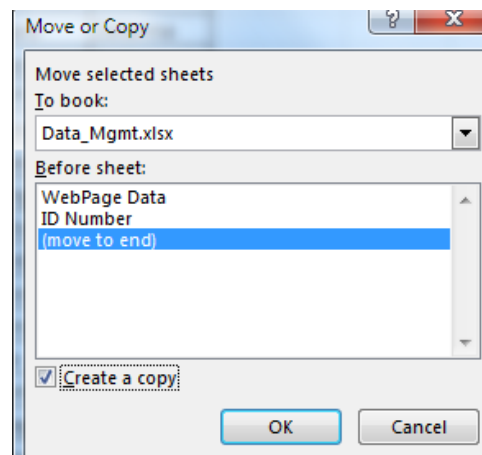


Figure 1 - Move or Copy box

## Renaming a Worksheet

Renaming a worksheet helps make your data easier to navigate.

To do so:

1. Double click on the worksheet tab you wish to change
2. Type in the new name for your worksheet
3. Press the *Enter* key when finished

**Activity:** Create a copy of the worksheet and rename it to: *Working Copy*

## Hiding columns and rows

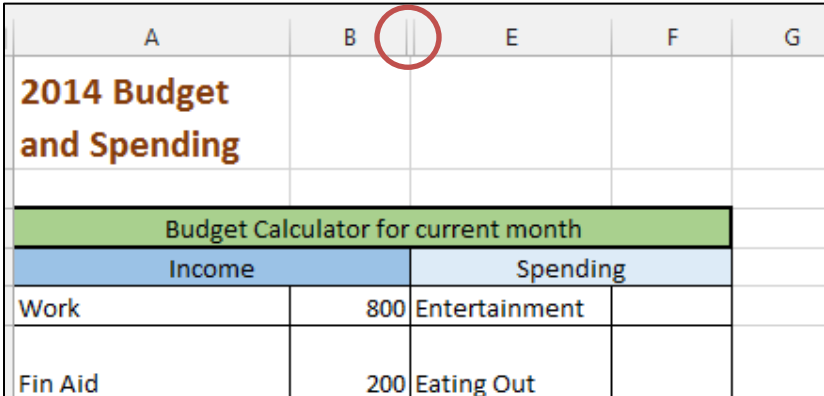
Hiding whole columns and/or rows can help make your data easier to comprehend by focusing your data in your worksheet.

### To hide columns and rows

1. Right click the label of the row or column you wish to hide
2. A menu will appear, select *Hide*
3. The row or column is now hidden

### To unhide columns and rows

1. Right click the hidden row/column indicator you wish to unhide
2. A menu will appear, select *Unhide*
3. The row or column is now visible



A	B	E	F	G
<b>2014 Budget and Spending</b>				
Budget Calculator for current month				
Income		Spending		
Work	800	Entertainment		
Fin Aid	200	Eating Out		


Figure 2 - A hidden column

## Importing Data Sets

Excel 2013 makes it simple to acquire data from a variety of sources, including Microsoft Access database files, web pages, text files, and other data sources

### Retrieving data from Access database tables

To import data from Microsoft Access:

1. Go to the **DATA** tab on the command ribbon
2. In the *Get External Data* group select  From Access
3. The *Select Data Source* window will appear where you can browse your computer to select a file
4. Make sure that an Access file is selected and hit the *Open*
5. The *Select Table* dialogue box will appear, from here you can select the data table that you want to import into the worksheet.

- To import multiple tables into the worksheet, select *Enable selection of multiple tables* and select the tables you want to import
6. Select *Okay*, and your data is imported
  7. The *Import Data* dialogue box will appear. This dialogue box lets you select how you want your data to be viewed in your worksheet
  8. The first options allow you to choose what will be used to display your data: A table, pivot table, pivot chart, and only as a connection are your options (creates a connection that you can use to later import your data).
  9. There are two options to select where you want to put the data: selecting a cell in the existing worksheet or a new worksheet
  10. Check the *Add this data the Data Model* add the imported data to the Data Model already defined in the Excel Workbook
  11. Click *OK* when finished

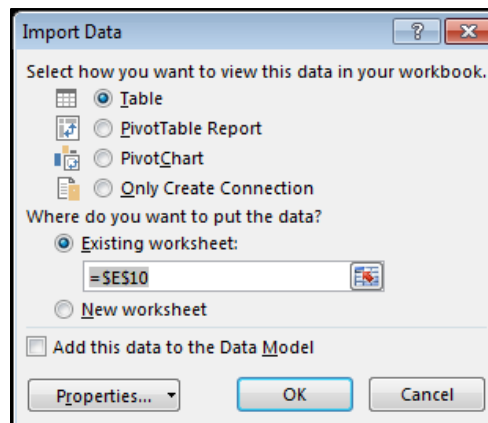


Figure 3 - Importing Data Window

## Retrieving data from a web source

To import data from a webpage:

1. Go to the **DATA** tab on the command ribbon
2. In the *Get External Data* group select **From Web**
3. The *New Web Query* dialogue window will appear that contains your homepage for your default browser
4. Type the name of your desired URL into the address box
5. Click *Go*
6. Excel indicates what information can be imported by adding a **+** to the left of the table
7. Click on the **+** to select the table to import, when selected the icon will change to a **✓**
8. After selecting the data to be imported click *Import*
9. The *Import Data* dialogue box will appear, select where in your current worksheet you would like to insert the data or select to add the information to a new worksheet
10. Click *OK* when finished

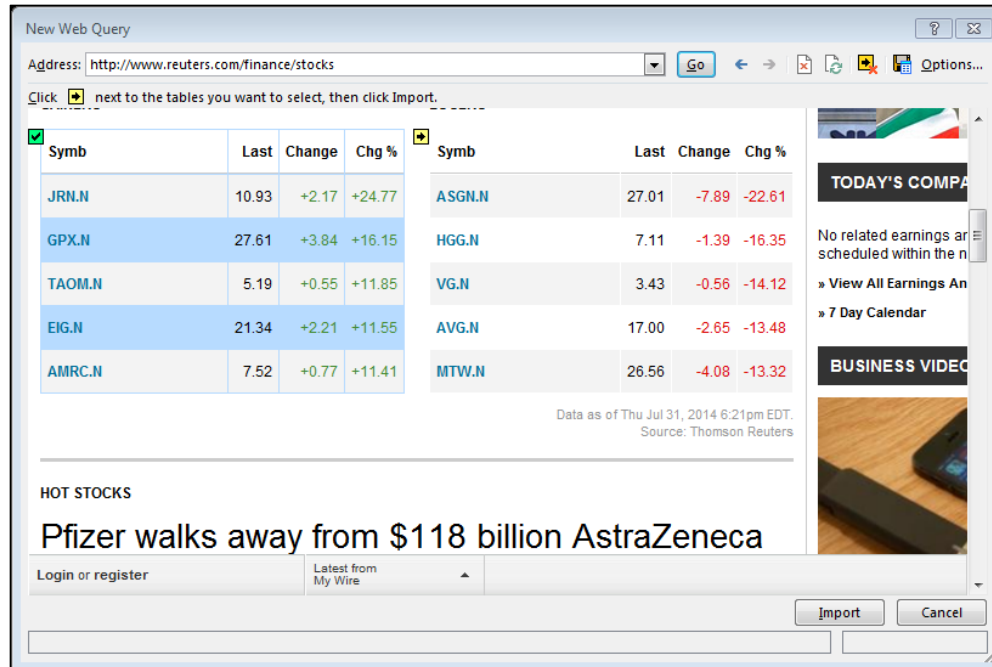


Figure 4 - New Web Query window

## **Eliminating Unwanted Data**

After importing large amounts of data you may find that you don't need all of the information in your worksheet. There are several ways to eliminate unwanted data from Excel:

### **Eliminating Columns and Rows**

To remove an entire column or row:

1. Right click on the column or row label
2. Select *Delete*
3. Your data will move over to replace the blank row or column

### **Eliminating Duplicated Data**

In Excel 2013 you can simply highlight duplicated data or locate and remove it

#### **Highlighting Duplicated Data**

1. Select the data that you want to locate duplicates in
2. Go to the **HOME** tab, locate the *Styles* group
3. Click on *Conditional Formatting*, a drop down menu will appear go to *Highlight Cell Rules > Duplicate Values*
4. A duplicate values dialogue box will appear, here you can change the style that duplicates are highlighted or change it from duplicate values to unique
5. Select *OK* and your duplicates will be highlighted

#### **Locating and Eliminating Data**

1. Select the data that you want to remove duplicates from
2. Go to the **DATA** tab in the ribbon, locate the *Data tools* group and click on *remove duplicates*
  - If you do not select all of the data a warning box will appear asking if you want to expand your selection
3. The *Remove Duplicates* dialogue box will appear, here you can select which columns you want to delete duplicates from
4. Select *OK* when finished

5. A dialogue box will appear telling you how many duplicates were found and removed and how many unique entries remain

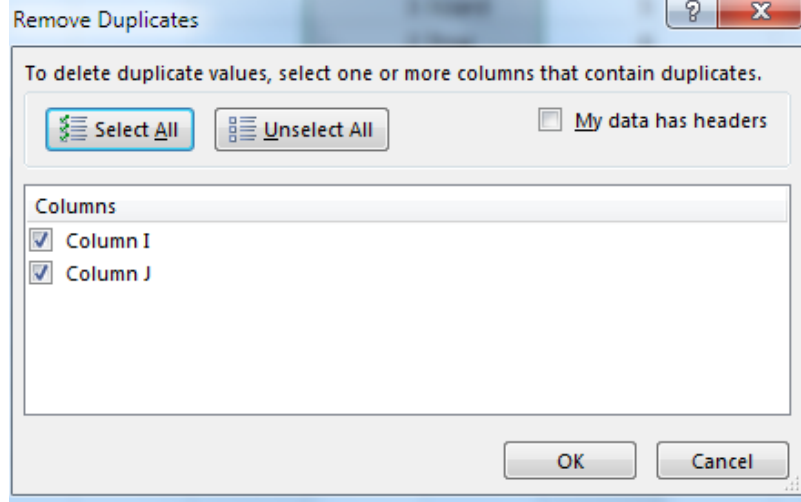


Figure 5 - Duplicate Removal box

## **Sorting Data**

Sorting your data can make it easier to find individual cases or to simply scan for points of interest.



### **Please Note:**

- Excel will automatically detect the first row as a header and exclude those cells from sorting, if you want the first cell to be counted you can do so in the *Sort* dialogue box discussed in sorting for multiple fields

## **Sorting data on a single field**

Sorting data on one field (like ID number, last name, location) is a simple process where you can sort data in ascending or descending order (Like A to Z, Z to A, Smallest to Largest, Largest to Smallest, Oldest to Newest and Newest to Oldest)


To sort data on a single field:

1. Select the first cell in the column or row you wish to sort (Selecting a title or header will not affect this)
2. Go to the **DATA** tab in the ribbon
3. Go to the *Sort & Filter* group and select either the ascending  or descending  sorting function
4. Your data is now sorted

## **Sorting data on multiple fields**

After sorting your data on a single field you may find that you have some duplicated data (this can be same last names or dates for example), Excel will automatically sort this data in ascending order from which their information was originally recorded, adding another sorting field can further sort the duplicated data (ex: duplicated last names are sorted alphabetically by first name).

To sort on multiple fields:

1. Select a cell in your data sheet that is part of the set
2. Go to the **DATA** tab in the ribbon
3. In the *Sort & Filter* group select *sort* 
4. The *Sort* dialogue box will open displaying your first sort level, if you added one before opening the *Sort* box

5. To add a level select: *Add Level*
6. Use the drop down menu underneath the header *Column* to select the column to sort
7. The other sorting properties have default settings that you can change
  - Under the *Sort On* header you can change the property to be sorted, such as: values, cell color, font color, and cell icon
  - Under the *Order* header you can choose to sort the data by ascending, descending or a custom order
8. Click *OK* when finished

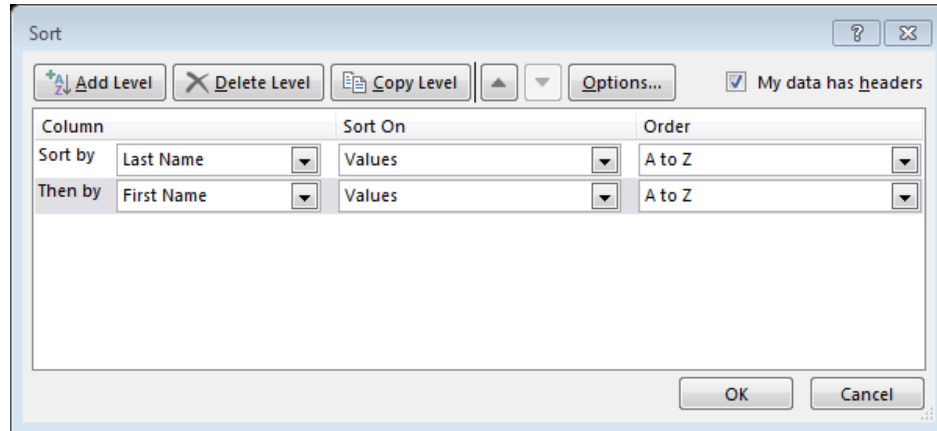


Figure 6 - Sort dialogue box

**Please Note:**

- The order that you place your sort levels matter

## **Freeze panes and split screens**

### **Freeze panes**

Freeze panes allow you to look through your document while keeping your headers in sight.

To enable freeze panes:

1. Go to the **VIEW** tab in the ribbon
2. In the *Windows* group and select *Freeze Panes*
3. Select the option that you would like to use either freeze the first row, column, or a whole area above a selection

**Activity:** Create multiple freeze panes and remove them

### **Splitting your worksheet**

Splitting a worksheet allows you to scroll and view multiple parts of your worksheet at once so you can analyze your data with ease

1. Select where you want to split
  - Selecting a column will create a split on the left side of the column
  - Selecting a row will create a split above the row
  - Selecting a single cell will create a 4-way split centered on that cell
2. Go to the **VIEW** tab in the ribbon
3. In the *Windows* group and select *Split*

4. To remove splits, select the command again

**Activity:** Split a worksheet all three ways

## Arrange All


The arrange all command allows you easily switch back and forth between multiple workbooks for an easy comparison.

1. Go to the **VIEW** tab in the ribbon
2. In the *Windows* group and select *Arrange All*
3. The *Arrange Windows* dialogue box will appear, here you can select how you want to view your windows
4. Select *OK* when finished

**Activity:** Arrange two workbooks vertically

## Quick Analysis

The quick analysis toolbar gives easy access to formatting, charts, functions, pivot tables and spark lines.

To access the toolbar select your group of cells that you wish to analyze and a  symbol will appear in the lower right hand corner of your selected cells, by clicking on it you will open the quick analysis toolbar. There are five categories in the quick access toolbar: formatting, charts, totals, tables, and sparklines, for this workshop we will cover formatting, totals, and sparklines. The options under these categories change depending on what kind of data is selected (i.e. dates, numerical, labels etc.). You can hover over an option for a preview or click on one to apply it.

## Formatting

The formatting menu allows quick access to some commonly used formats to allow the reader to quickly scan for data that is exceptional, by highlighting the commands in the menu you can preview your formatting.

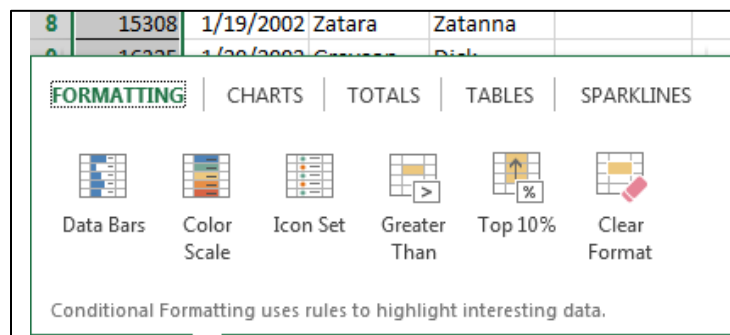


Figure 7- Formatting Quick Analysis

## Totals

The totals tab allows quick access to commonly used formulas and choose where the equation will appear. While the options are limited they allow quick previews of computations.



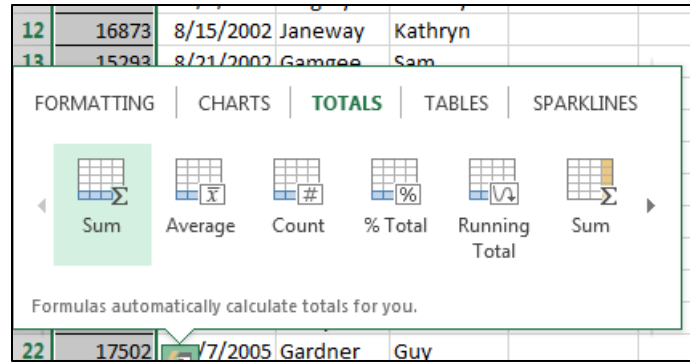


Figure 8 - Totals tab

## Sparklines

The sparklines tab gives quick access to creating small single cell graphs quickly. You can always go back the design tab in the ribbon to further edit your sparkline.

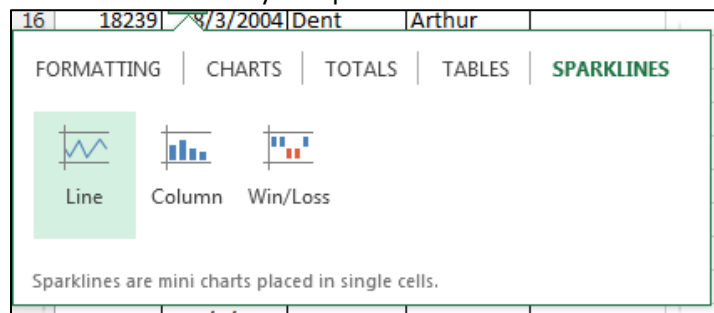


Figure 9 - Sparkline tab

**Activity:** Go through the different options of the *Quick Analysis* tool

## Creating new Variables

Creating new variables out of existing data is easy to do with Excel 2013, creating new variables is helpful if you want an average variable of sets of data.

To create new variables in excel:

1. Decide what type of variable you want to make
2. Choose the data you want to use for your new variable
3. Create a new column or row for your data to go in
4. Create a header for your new variable
5. Enter your formula into the top cell
6. Use auto fill to drag down your formula
7. Your new variable is complete

**Activity:** Create a new variable titled: “Montgomery GI Bill” in the first cell create a formula adding the first entry of *Montgomery GI Bill: Active Duty* and *Montgomery GI Bill: Selective Reserve* or =D4+E4 use autofill to drag the data down to complete the variable. Hide columns D and E

## **Saving for Further Use in SPSS or Excel**

### **Saving for Excel**

1. Select *File* from the top menu
2. Select *Save As* in the left column group
3. Select *Computer > Browse*
4. Choose a location to save your file
5. Create a save file name and press *Save*

### **Saving for SPSS**

Saving your data in the proper format can save you a lot of time when importing data into SPSS from Excel.

- Make sure that you have column headers that will transfer into SPSS correctly (Short headers with no spaces or symbols outside of “\_”)
- Excel files must be saved with a .xlsm .xls or .xlsx extension to be readable by SPSS