**The Data Ribbon**

About Importing Data from Other Types of Files

Excel 2007 incorporates wizards for importing data from many types of files. View the Data Ribbon and choose the appropriate button in the *Get External Data* group on the left end of the ribbon. Excel will launch the appropriate Import Wizard; follow the prompts to locate the file to import and complete the process.
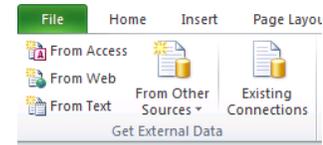
Text File Definitions

Text files fall into several classifications. When importing from a text file, you need to specify the correct type for the import to go smoothly.

Fixed Width data is separated into columns by "counting off" a certain number of characters. Shorter entries are padded to fill the same space as longer ones; the dots below represent the padding spaces.

```
Barb····Lawson··9011 Williams Drive···········Miami········OH
Warren··Lilley··P.O. Box 121···············Jackson······NY
Sue·····Mason···33975 E. Mississippi··········Aurora·······IL
Paul····Masters·2377 South Olde Bridge Parkway·Redwood······NY
Eduardo·Mendoza·377 SE Ridgeview··············Littleton····CO
Stephen·Miller··P.O. Box 3122················Youngstown···MI
Rebecca·Nelson··201 S. Tulip·················Grand Rapids·MI
```

Delimited data is separated into columns by "delimiters", which are special characters such as tabs, commas, etc.

Robert L. Smith→677 E. Madison Grove→Downsville→NY
William Smith→200 E. Minolta→San Diego→CA
Patricia Snyder→4001 SE Mason→Mission→CA
Amanda Stevens→3072 S. Temple→Salt Lake City→UT

Ivan Petersen,1211 Moscow,Duane,CA
Janet Peterson,12122 S. Marion Way,Georgian,MI
Tasha Rogers,4 Courtside #311,Albans,NY

Note: Comma delimited text files are opened automatically in Excel, without the need for an input wizard.
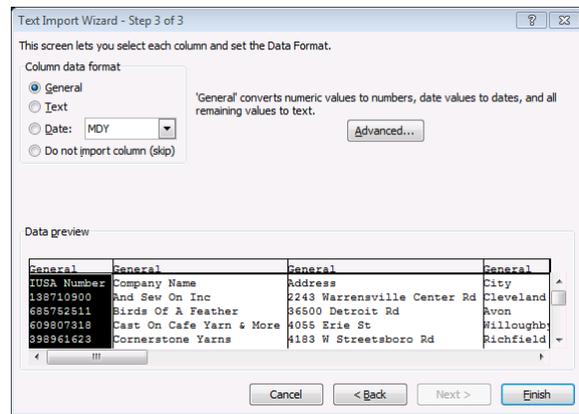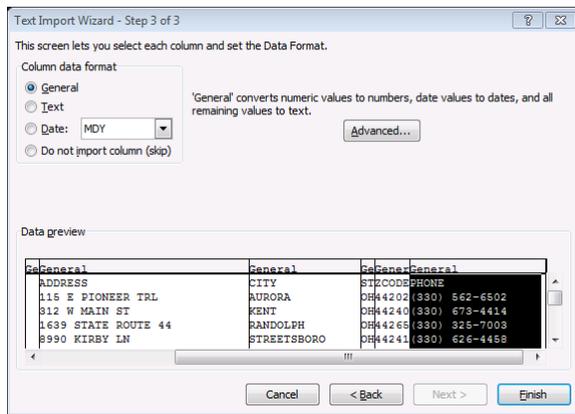
Import Wizard Steps
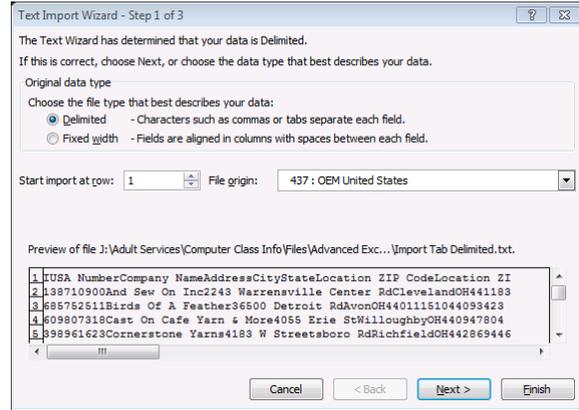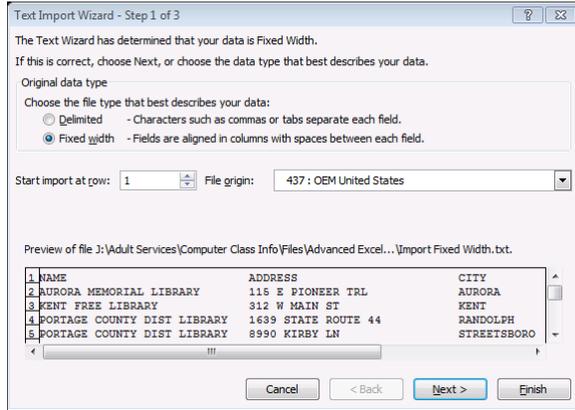
In Step 1, tell Excel whether the file you are importing is fixed width or delimited.
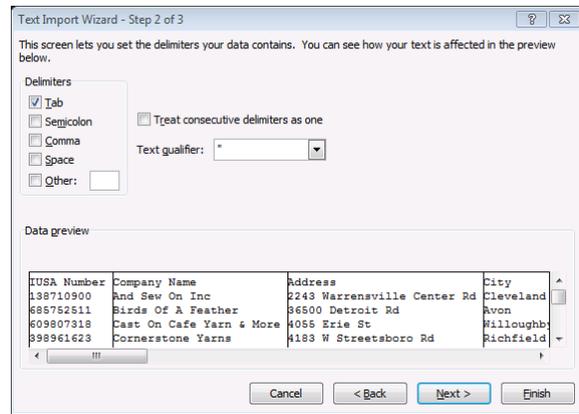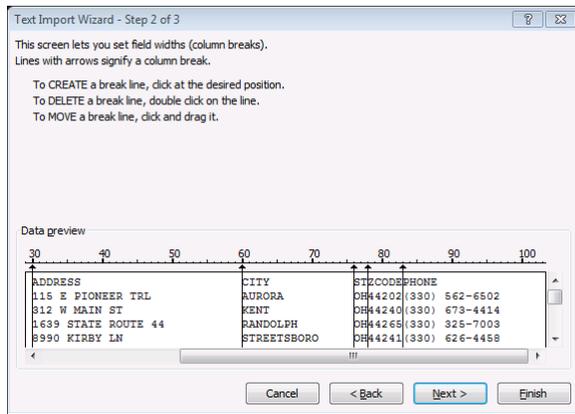
In Step 2, tell Excel where the breaks between
fixed-width columns should occur, or tell Excel what kind of delimiter(s) is present.

Step 3 lets you define data types for each column.

Text to Columns

Use this tool to separate the contents of one cell into several cells, such as separating a column of names into separate first and last name columns.

Add blank columns to hold the data to be separated. For instance, if you will be dividing one column into three columns, insert two blank columns to the right of the original one.
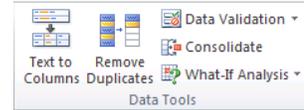
Select the original column and click the Text to Columns button in the *Data Tools* group toward the right of the Data Ribbon. Excel will launch a wizard to guide you through the process. This wizard is virtually identical to the Text Import Wizard above.

Remove Duplicates

This tool lets you remove duplicate values in a data set. Select the data set you are working with and click the Remove Duplicates button in the *Data Tools* group on the Data Ribbon.

Click check boxes to select the columns you want to search for duplicates, then click OK to remove the duplicate rows.

Be aware that if you select fewer than all the columns you may accidentally remove rows that aren't really duplicates. For instance, if you only searched the First Name and Last Name columns, and there were several people with the same name but different addresses, all but the first person would be deleted.
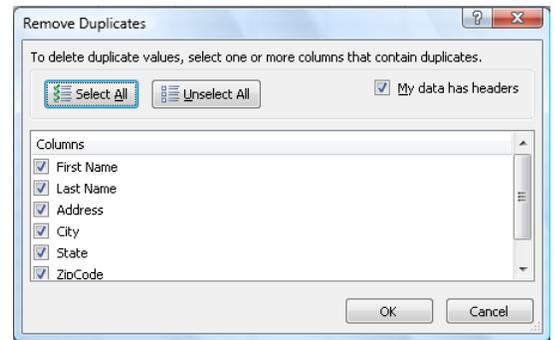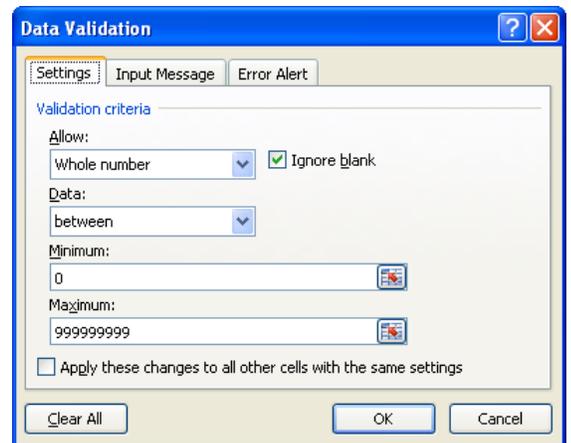
Data Validation

Data validation limits the type of data the user can type into particular cells. (For example, if a cell will be used in a formula, you may want to restrict it to contain only numbers, not text.)

Select one or more cells to validate, then click the Data Validation button in the *Data Tools* group.

- On the Settings tab, choose your criteria for valid data.
- On the Input Message tab, type information that people entering data in these cells should see.
- On the Error Alert tab, type an error message for your users to see if they try to enter invalid data.
- Click OK.

If you set data validation on cells that already contain invalid values, no error messages will appear. Click the Data Validation button and choose Circle Invalid Data to locate these values.

Filter Data with AutoFilter

Filtering "pulls data out" of an Excel data table by selectively displaying rows containing particular data. (The data table must contain a header row.) When you use print, copy and paste commands, you print, copy and paste only the filtered (visible) data.

On the Data Ribbon, click the Filter button in the *Sort & Filter* group. Drop-down arrows will appear by each category in the header row. You can also turn on filtering by clicking the Sort and Filter button in the *Editing* group on the Home Ribbon.

<u>Note</u>: when you format a data list as a table, filtering is automatically turned on.

Click the drop-down button beside the category you wish to filter by. You can filter by any individual entry in that category; click the check box beside any value to see only rows containing that value.

Excel also gives you specific filters that vary depending on the type of data in the column you are filtering by. There are separate lists of Number Filters, Text Filters and Date Filters:
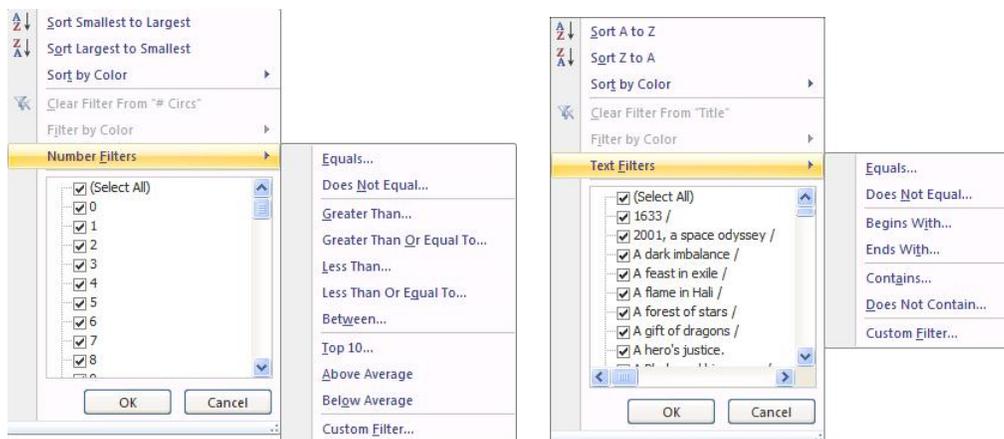
In each category, Custom Filter allows you to refine your search, for example by selecting two different criteria within the same category (e.g. to display records where the state is either OH or NY).

You can filter by more than one field, by choosing another category and repeating the process. You will then see only data that matches both criteria.

Remove Filters

Remove all filters by clicking the Filter button again. Clear individual filters by clicking that column's drop down button and choosing "Clear filter from…".

Copying and Pasting Filtered Data

- When you *copy* and paste filtered data, only the visible, filtered cells are copied and pasted.
- Do not use cut and paste to move filtered data. All the rows between the first and last visible rows are cut but only the visible ones are pasted; the rest are lost.
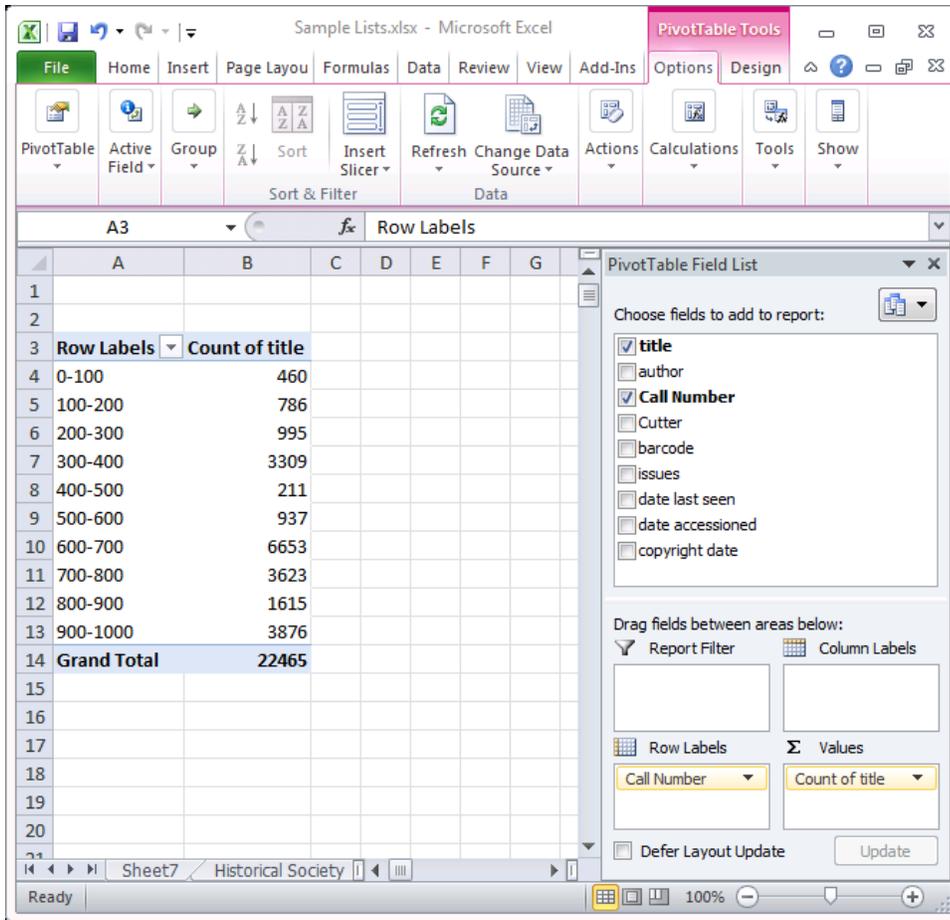
About PivotTables

PivotTables are one of Excel's most flexible tools for analyzing and summarizing your data. They let you examine relationships between fields in your data tables, quickly expanding and collapsing the data to look at details and summaries.

Set Up a PivotTable

1. Prepare your source data. Data for a PivotTable should conform to the normal Excel table requirements: it should not contain any blank columns (fields) or blank rows (records); each column must have a field name at the top, and the various columns/fields should contain one type of data only (e.g. text, numbers or dates.) You may want to give the source data a name.
2. Select the data, or at least place the active cell inside the data range.
3. Click to view the Insert Ribbon and click the PivotTable button in the *Tables* group at the far left. The Create PivotTable dialog box will appear. By default, Excel will use your selected range and place the PivotTable on a new worksheet. Click OK.
4. On the new sheet, you will see the PivotTable layout area on the left, and the field list on the right. At the lower right corner, you see four areas defined: Report Filter, Column Labels, Row Labels and Values.
5. Drag and drop fields into the four areas of your PivotTable:
   a. Fields in the Column Labels area appear across the top of the report.
   b. Fields in the Row Labels area appear down the left hand side of the report.
   c. Fields placed in the Values area contain the data that will be summarized.
   d. Fields placed in the Report Filter area are not displayed within the PivotTable. Use this area to isolate and focus on subsets of your data.
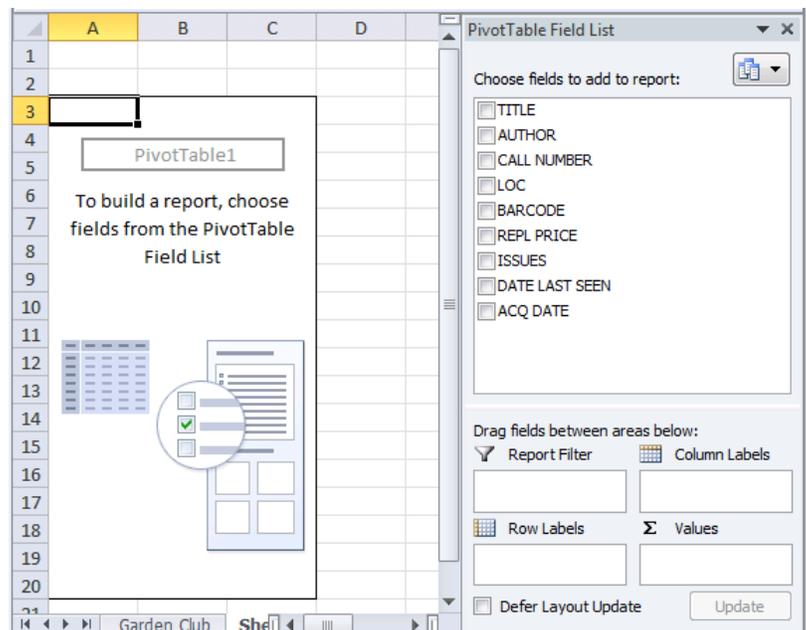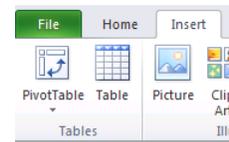
As you add each field, the PivotTable starts assembling itself; when you drag a field from one area to another it is instantly redrawn.

To see the raw data underlying any summarized number in a PivotTable, double click that number. A new sheet will open, displaying those records. Delete the sheet when done. (This sheet cannot be used to change the source data; it is for observation only.)

Fields in the Values Area
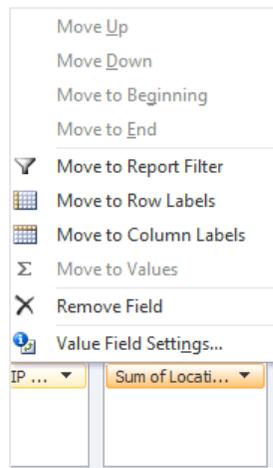
Excel selects standard summarizing functions for the fields in the Values area: Sum for numerical fields and Count for text fields, dates, etc. Select other functions by clicking the drop down arrow beside the field name and choosing Value Field Settings from the menu.

On the Value Field Settings dialog box, you can choose from summarizing functions such as Average, Max and Min. You can also format your numbers using the Number Format button in the lower left of the dialog box. (You need to change the settings here rather than formatting numbers manually, since manual changes are reset every time you refresh the pivot table.)
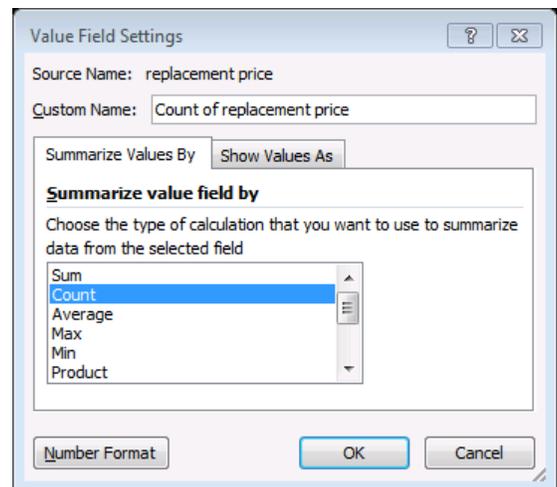
Fine-tune your PivotTable with tools from the Contextual Ribbons
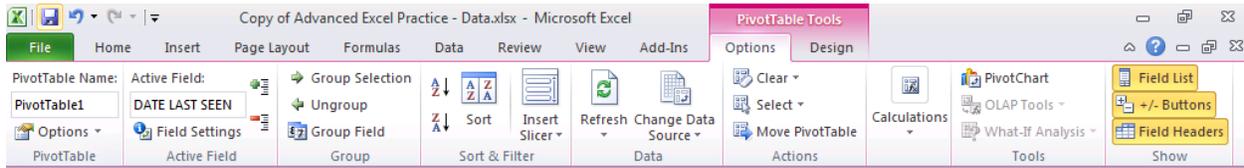
1. On the Options Ribbon, the Group Field button in the *Group* group lets you summarize your PivotTable by numerical or date fields. You can also manually group text fields to achieve the same effect.
2. The Refresh button in the *Data* group updates the PivotTable with any changes that have been made in the source data. This is most important if your data source is an external database.
3. The PivotChart button in the *Tools* group lets you add a PivotChart to graphically display your data.
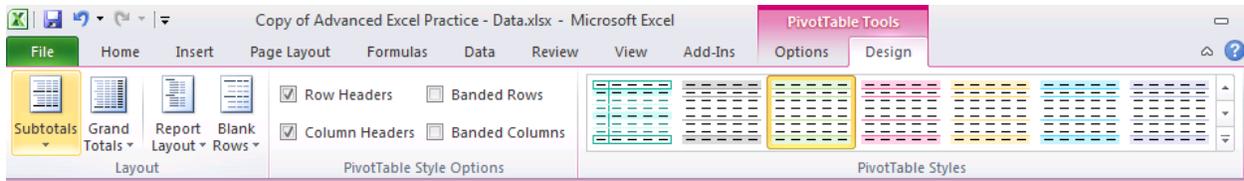
4. On the Design Ribbon, select styles from the *PivotTable Styles* group and add features like banded (striped) rows or columns from the *PivotTable Style Options* group. Add or remove automatic totals and alter the layout in the *Layout* group.

Practice Exercises:

Go to this website: *http://smfpl.org/advanced_classes/advanced_ms_excel_3_data_analysis* and download the following files: Yarn Shops (a tab-delimited text file) and Advanced Excel Practice – Data (an Excel 2010 or 2003 workbook).

1. Open Excel and import the Yarn Shops file. While importing, define the "Employee Size" field as a text field (otherwise Excel will try to interpret it as a date field.)
   a. Format the data as a table.
   b. Using filters:
      i. Display only stores with a sales volume between $500,000 and $1 million. How many are there?
      ii. Use a second filter to display only those stores with five to nine employees. How many stores remain?
      iii. Create a new blank worksheet. Copy and paste the filtered records to the new worksheet. What is the result?
      iv. Clear all the filters from the original data table. Save your workbook as Yarn Shops. (Note: This file can have the same name as the text file because it is a different file type – an Excel workbook.)

2. Open the Advanced Excel Practice – Data workbook.
   a. View the Garden Club sheet.
        i. On this sheet, the members' names occupy only one column. Split the names into two separate columns for First Name and Last Name. Rename the columns appropriately.
        ii. Set up validation rules to limit the State column to two characters, and to limit the Zip Code column to whole numbers between 44200 and 44300. Write appropriate Input Messages for the users of the database. What happens if you try to enter a zip code of 14617?
   b. View the F&SF sheet.
        i. Search for duplicate books. Consider a book a duplicate if it has the same title and author as another book. You started with 1337 records; how many are left after removing duplicates? Use Undo to restore the deleted records.
        ii. Use pivot tables to find the following information:
            1. How many different authors are represented in this collection? Who are the top three most popular authors? In other words, which authors have the largest total number of issues (book checkouts)?
            2. Who are the authors with the largest number of books in our collection?
            3. How many titles were purchased in January? February? The rest of the months? (Hint: use grouping.)
            4. How many books do we have in each of the four categories: FANT, FANTPB, SF and SFPB? Which section has the largest total number of issues? What is each section's average number of issues (how many times, on average, was each title checked out)? Change the labels of your columns to "Number of Titles", "Total Checkouts" and "Average # of Checkouts"; format the Average field to show two decimal places.