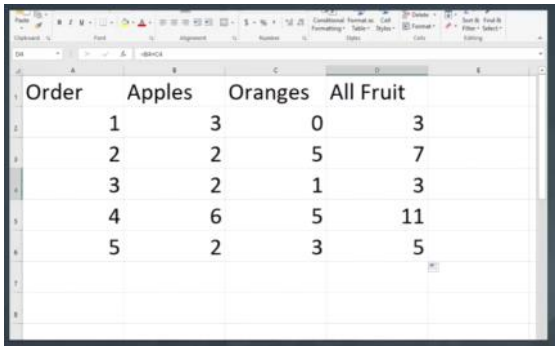
Spreadsheets

Saturday, November 30, 2024 7:09 PM

Why Spreadsheets?

Before spreadsheets, it was a major chore to extract meaningful information and gain insight about the bigger picture in a company or process.
With a spreadsheet, the data could be moved around, corrected or inserted.



Order	Apples	Oranges	All Fruit
1	3	0	3
2	2	5	7
3	2	1	3
4	6	5	11
5	2	3	5

We can add formulas of all kinds to columns or aggregate information from a list, maybe a sum or an average, we could compare information groups within our dataset.
There are limitations though.
If you have a lot of users at the same time or if the data set is very large your performance may suffer.
The capacity of a spreadsheet application, the maximum size of the sheets and how many there can be varies with the application.
But spreadsheets are easy to obtain and use. The ability to quickly load, modify and compare to see your data and what it means is powerful.
This is why you need a good understanding of what you can accomplish with spreadsheets.
With a little practice, the process is fast and easy.

Quiz Question

What benefits do spreadsheets provide? Check all that apply.

- [ ] they're perfect for massively large data sets
- [x] you can quickly load and visualize data ✓
- [x] you can add formulas to analyze data ✓
- [x] they make it easy to manipulate data ✓

Commas vs. Periods in Different Countries

A quick note before we get started with spreadsheet applications.
Depending on your location, decimals and commas may be treated differently in spreadsheet applications (mostly in European countries) than what you will see in this course.

For this course, we will be using a convention where commas separate large values - like 1,000.00 or 1,000,000.00. On the other hand, the period will be used as the decimal separator, as you can see from these previous numbers.

You may want to assure that you are following along with these same conventions, or at least be aware of the differences you may see should you be using a different convention for commas and decimals.

Potential Solutions

Periods and Commas
https://www.officetooltips.com/excel_2016/tips/change_the_decimal_point_to_a_comma_or_vice_versa.html(opens in a new tab)

- Set Decimal Separator to .
- Set Thousands Separator to ,

Commas and Semicolons

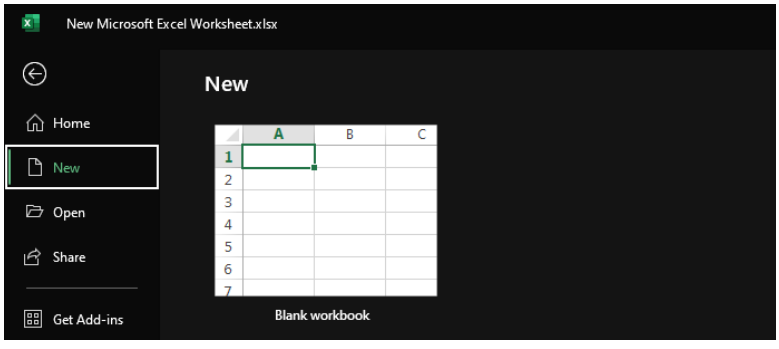
https://lockone.wordpress.com/2015/08/06/excel-now-using-semicolons-instead-of-commas-in-separating-formulas-fix-it-now/(opens in a new tab)
Note you should Select English (United States) to have your formatting match what is shown in the course

After making any changes to the settings of Excel, you may need to restart your computer for the changes to actually go into effect

Navigation: Worksheet

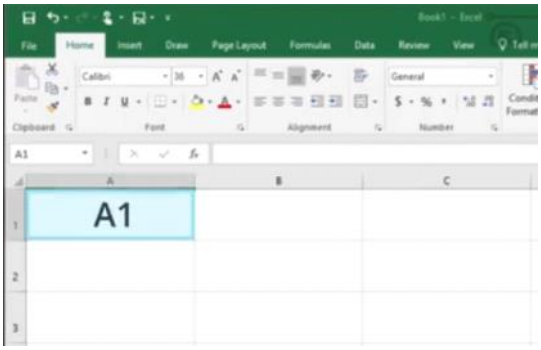
Opening a Blank Spreadsheet:

Open a new Excel spreadsheet by navigating to File > New > Blank spreadsheet.



Structure of a Spreadsheet:

Columns are labeled alphabetically (A, B, C, etc.). After Z, labels continue as AA, AB, AC, etc.
Rows are labeled numerically (1, 2, 3, etc.). Each box is called a cell, and each cell has an address based on its column and row (e.g., A1, E8).

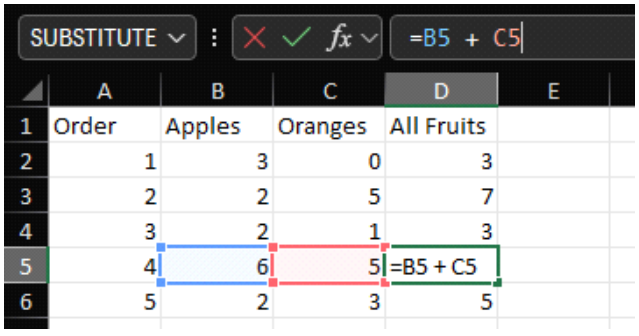


Working with Sheets:

Spreadsheets can have multiple sheets. Use the Add Sheets button to create new sheets and rename them for clarity. **When working with multiple sheets, the sheet name becomes part of the cell address.**

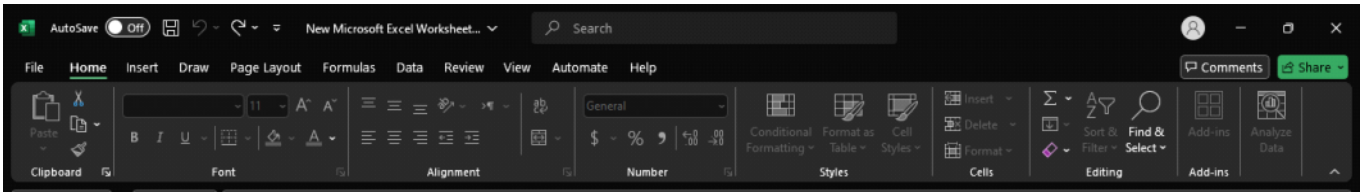
Formula Bar:

The formula bar shows the contents of a selected cell, including formulas. You can type directly into a cell or the formula bar. Formulas start with an equal sign (=), e.g., adding two cells: =A1+B1.
Clicking the formula bar highlights all cells referenced in the formula.



Main Menu:

The main menu offers customization features that vary across spreadsheet applications (e.g., Google Sheets, Excel, Apple Numbers). Our focus will be on common features across these platforms.



Navigation: Menu Bar

Main Menu Bar Overview

Structure and Customization:

The menu bar contains submenus (tabs) with tools and operations specific to the spreadsheet application.

In Excel, menus can be customized as needed. Features may vary in location and naming across Google Sheets, Apple Numbers, and Excel.

Help Features:

Help is available on the Home menu or by pressing F1.

Useful for finding features or learning more about operations.

Menu Highlights

File Menu

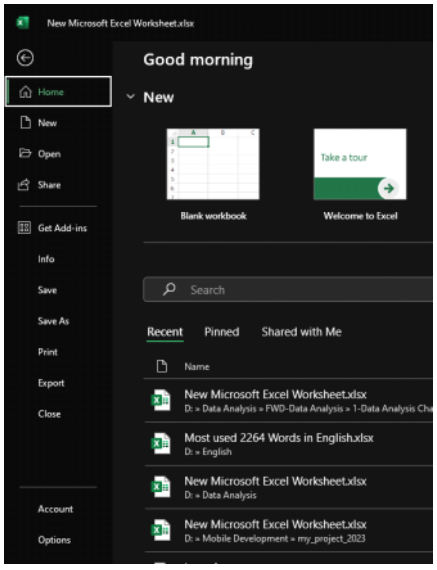
Used for file operations between the spreadsheet and the computer or online environment.

Key operations:

New: Create a blank spreadsheet.

Open: Open existing spreadsheets.

Save As: Save spreadsheets with a new name or format.



Home Menu

Core editing and formatting functions:

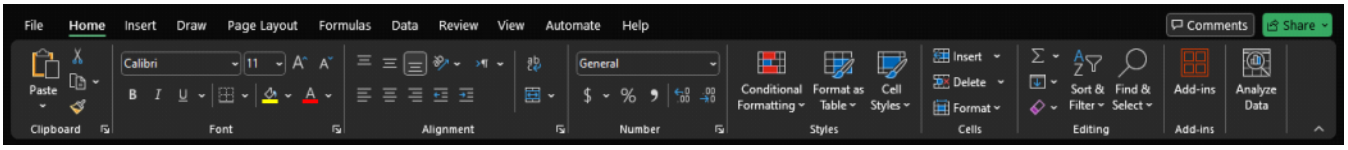
Clipboard operations: Cut, Copy, Paste.

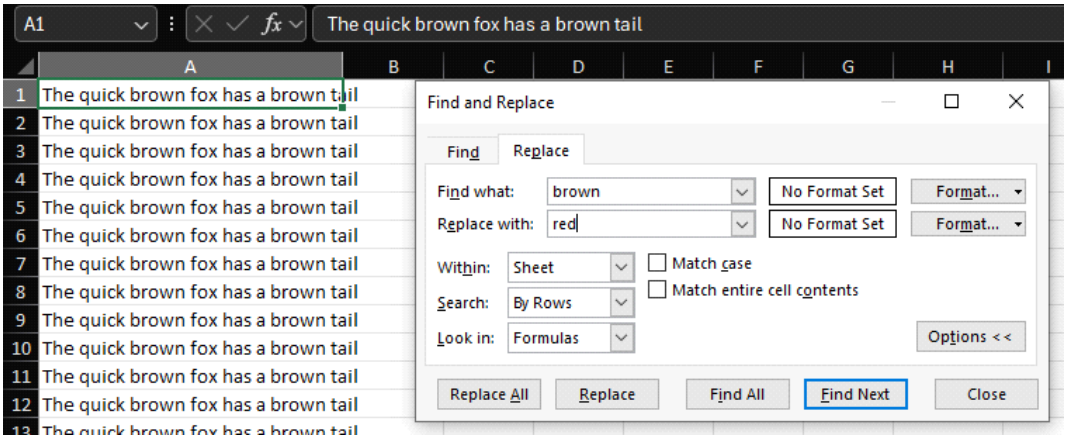
Undo/Redo: Found in the Quick Access Toolbar (customizable).

Formatting options: Fonts, data types, table styles, and row/column management.

Search and Replace: Use Find and Select to locate or replace data.

Tip: Specify the scope (selected cells, entire sheet, or multiple sheets) when searching or replacing.

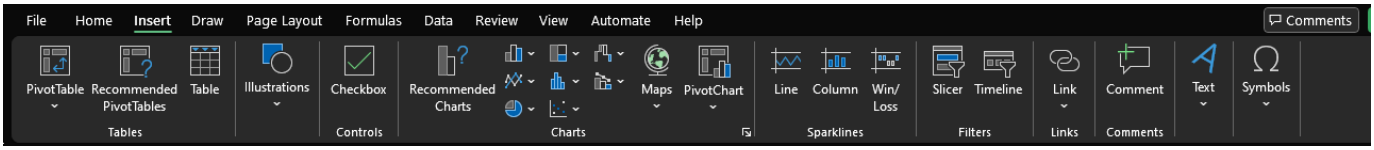




Insert Menu

Special items and features:

Add hyperlinks or charts.
Charts default to being inserted into the active sheet.
Pivot Tables: Will be discussed in-depth later.



Data Menu

Essential tools for data manipulation:

Filter data files.
Sort lists.
Split columns.
These tools will be utilized in upcoming lessons.

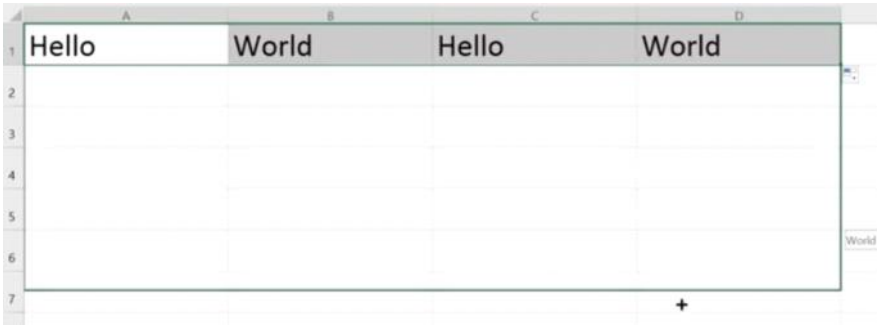
General Tips

Formatting data helps document and communicate your analysis effectively.
Regular use and exploration make features easier to find and utilize.

Navigation: Shortcuts

Spreadsheets have unique gestures that enhance productivity. One such feature is the fill handle, which allows users to copy or extend patterns across cells. To use it:

- Select a cell or range.
- Hover over the lower-right corner until a plus sign appears (the fill handle).
- Drag the mouse to fill adjacent cells.



The fill handle can copy values, create sequences, or replicate formulas, making it a powerful tool for repetitive tasks.

Context Menu

The context menu offers quick access to relevant functions for a selected cell or range. You can open it by right-clicking on your selection. This menu is a convenient way to perform common operations without navigating the main menu.

Keyboard Shortcuts

Many standard keyboard shortcuts work in spreadsheets, such as:

- Undo (Ctrl + Z)
- Redo (Ctrl + Y)

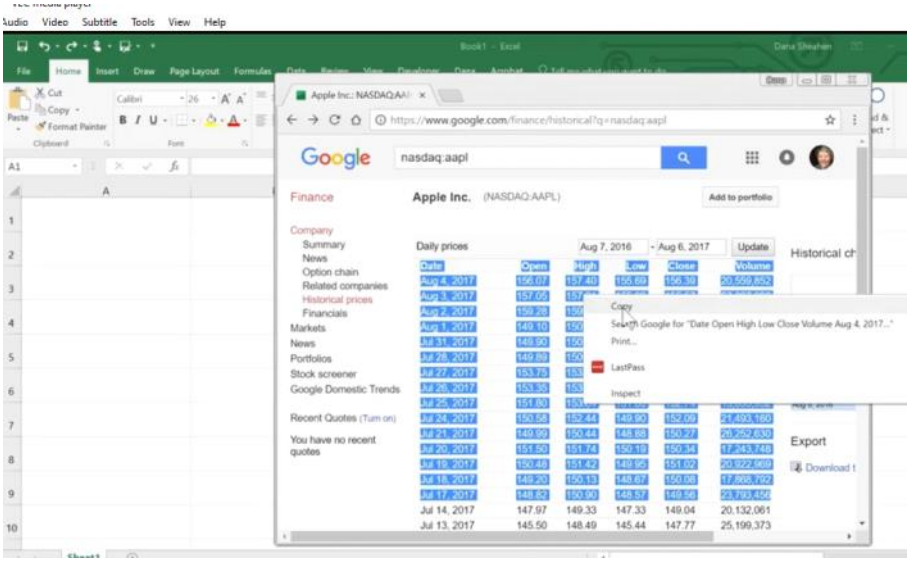
These shortcuts save time and streamline workflow, helping you focus on analysis rather than repetitive actions.

Purpose of Spreadsheet Gestures

These features and shortcuts simplify working with spreadsheets, allowing users to move beyond calculations and concentrate on analyzing and interpreting data effectively.

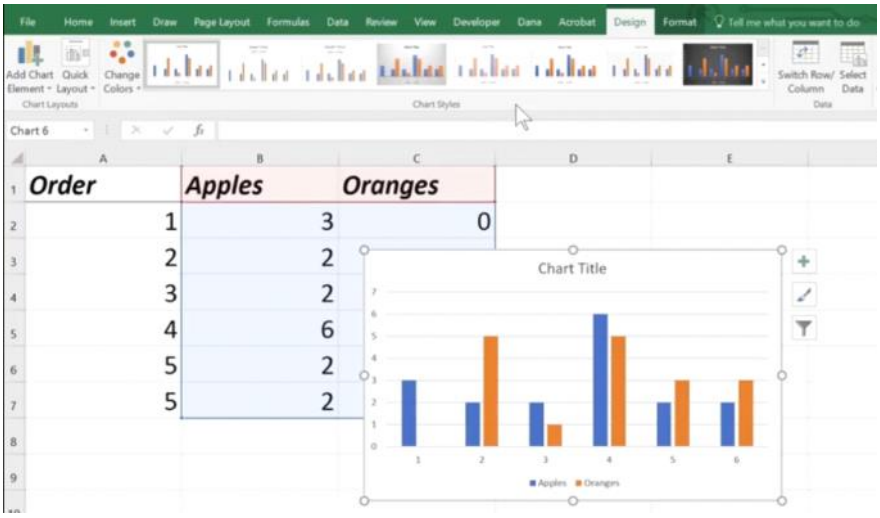
Copy Data

Let's work through some specific manipulation skills.
First step, copy data and open a Google Finance web page of Apple historical stock prices.
With spreadsheets, we can copy whole tables. This works well when a website is set up in a format that spreadsheets recognize such as this financial data. When I copy and paste this table from the financial website, it's copied and pasted as cells. Each cell has a separate value.



Range Addressing

You've already figured out by now that a group of cells is called a range. The address of a cell or a range is required when we want to reference that cell or range in a formula or chart. So how do we reference a particular range in Excel? Select a range by holding down your mouse button and swiping or dragging. This selects a group of cells that are next to each other which can be used to make a chart which we will explore in the visualize data lesson.

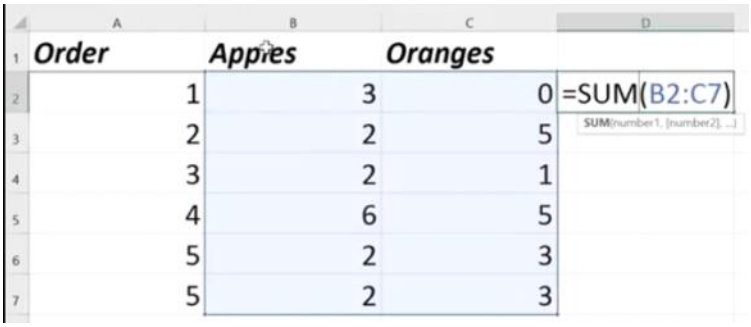


Order	Apples	Oranges
1	3	0
2	2	
3	2	
4	6	
5	2	
5	2	

Or to add formatting such as cell highlighting or font color, many choices for formatting.

Order	Apples	Oranges
1	3	0
2	2	5
3	2	1
4	6	5
5	2	3
5	2	3

Or for inclusion in a Formula. One way to see this is to look at how range is referenced in a simple formula using sum. In this example, we'll create a formula that's just equal sign then type sum and then select a range.



The range address shows up in the formula. A range is defined by the upper left and lower right corners of the area selected.
In this example, we can see that the range is from B2 in the upper left to C7 in the lower right. The range corners are separated by a colon to indicate, and everything in between.
In addition to selecting ranges by column and row labels, it is also possible to name ranges to make them easier to read and identify. You'll learn about that in a later lesson.

Next, I want to show you something else that's special about addressing with spreadsheets. That's the idea of relative versus absolute addressing.

Quiz Question

Which of the following is a valid range address? Check all that apply.

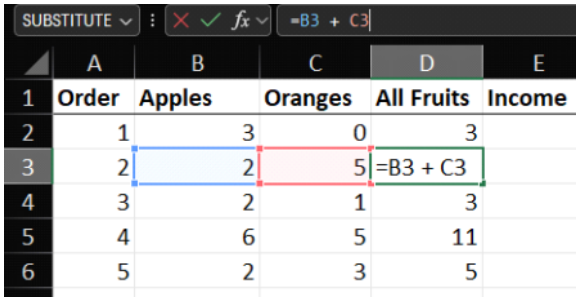
- [x] A1:A2
- [ ] C8:A9
- [x] AA3:BB5
- [ ] 9F:9G
- [x] E:G
- [x] 16:18

Correct! A1:A2 and AA3:BB5 are valid because they both start with an upper left cell address and end with a lower right cell address. E:G and 16:18 are also valid... this was a bit of a trick question. It is possible to define whole columns, such as columns E thru G, and whole rows, such as rows 16 thru 18, as ranges.

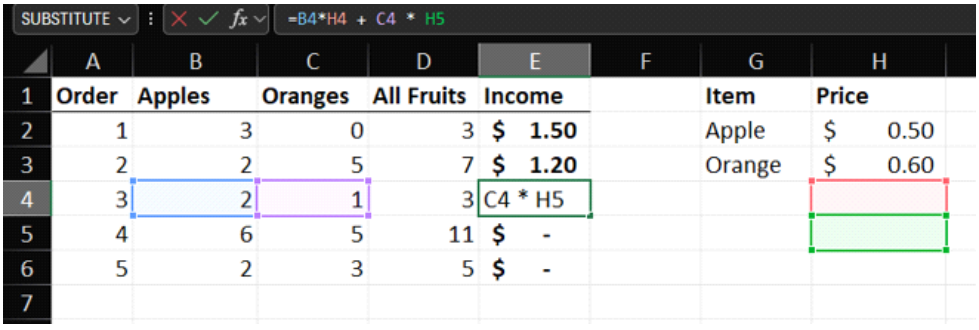
Relative vs Absolute Addressing

In spreadsheets, addressing cells in formulas can be either relative or absolute, depending on how the references behave when copied or filled into other cells. **Relative addressing** references cells based on their

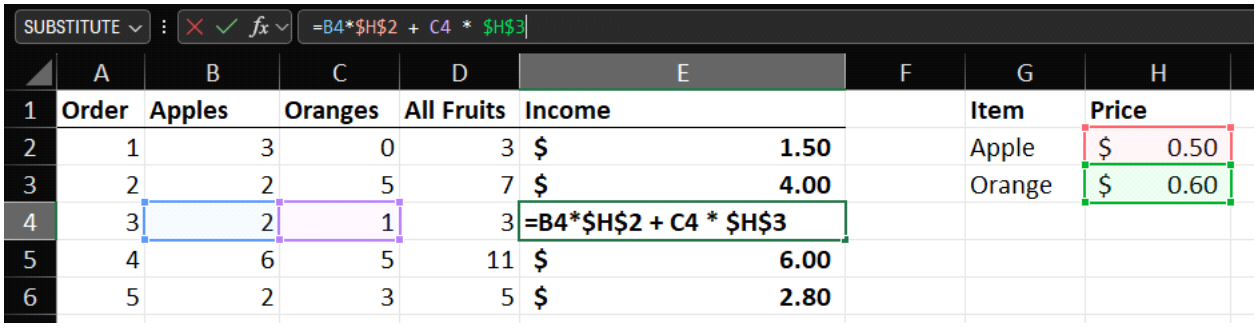
position relative to the formula's location. For example, when summing apples and oranges in a "Fruits" column, the formula =B2+C2 adjusts automatically as you fill it down, changing to =B3+C3 and so on. This is the default behavior and works well for patterns that need to adapt row by row.



However, relative addressing can lead to errors when fixed references are needed. For instance, when calculating total income with a formula like (Apples × Apple Price) + (Oranges × Orange Price), copying or filling down the formula causes the price references to shift, resulting in incorrect values.



To address this, **absolute addressing** is used. Absolute addressing fixes the cell reference, ensuring it doesn't change regardless of where the formula is copied. This is achieved by adding a dollar sign ($) before the column letter and row number in the cell address. For example, H2 keeps the reference to the Apple Price fixed at cell H2, while H3 does the same for the Orange Price. Updating the formula to include absolute references ensures that only the sales quantities adjust for each row, while the price references remain constant, providing accurate calculations.



This distinction between relative and absolute addressing is key to ensuring that formulas behave as expected, especially in more complex spreadsheet operations.

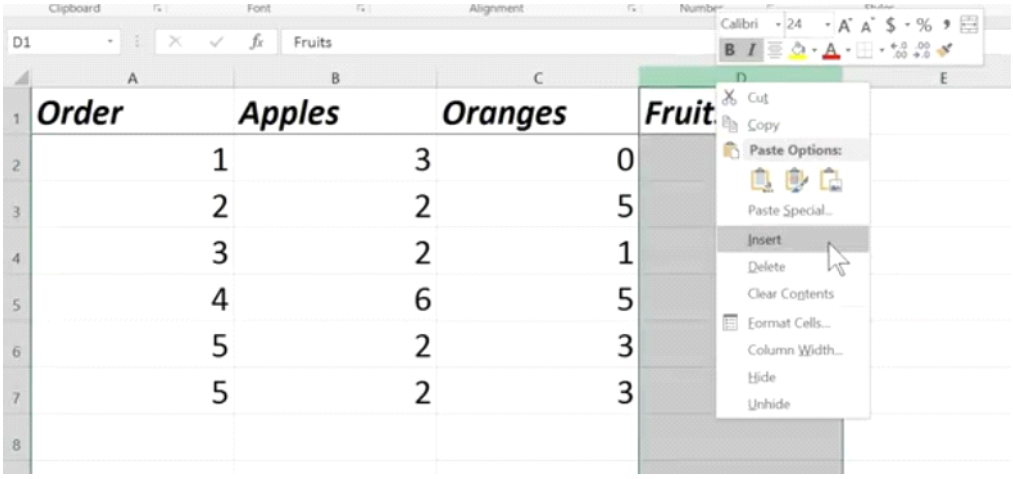
Quiz Question

Which of the following are valid cell addresses?

- [x] A1
- [x] A1
- [x] $A1
- [x] A$1

Insert and Delete

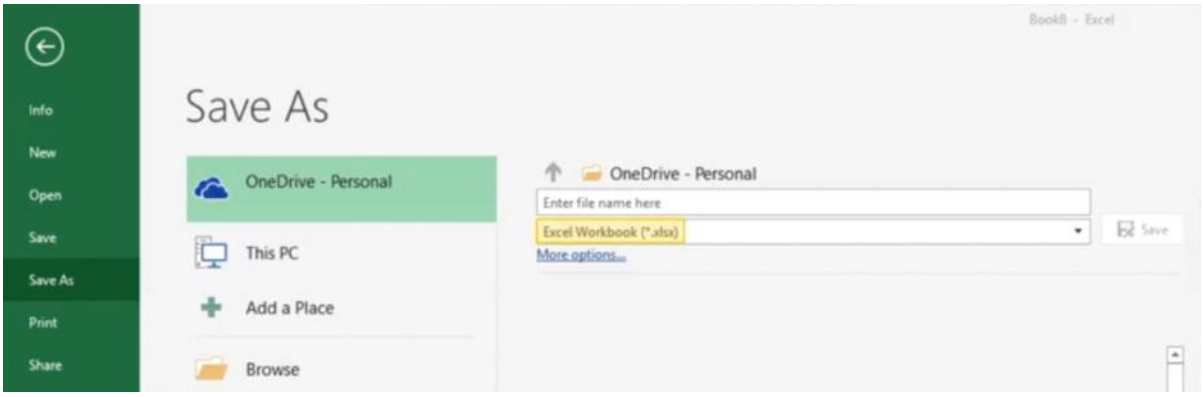
Both the home and context menus have options for inserting and deleting rows, columns, and cells. Recall that the context menu appears when you right-click.
Suppose I have another kind of fruit at my fruit stand and I want to add a column for it, select a column, right-click and insert, add your label and you're ready to go.



Save Data

Always save your work. There's nothing worse than spending hours on a project only to have it disappear because of a power glitch.

If you're using Google Sheets, your work is saved in the cloud every time you make a change. If you're using a desktop spreadsheet like Excel, you must be more diligent. Get in the habit of clicking save frequently or use the shortcut Ctrl+S.
The first time you save your sheet, the Save As dialog will appear.



After that, Excel will know where to go to save your work.
The xlsx format is the default Excel format but there are a number of formats you can save your work in with the pull down menu. Google Sheets, Apple numbers and Apache Open Office can all open xlsx files as well as the older Excel format xls.
Now, it's your turn to save your work.

Quiz Question

Which of the following file types can be usefully read by Excel and most other spreadsheet applications?

- [x] Excel Workbook (*.xlsx)
- [x] Comma Delimited (*.csv)
- [x] Excel 2003 (*.xls)
- [ ] JSON (*.json)

Introduction to Manipulate Data

In this lesson, you'll learn to manipulate data in your spreadsheet. You already know how to paste data into your spreadsheet, and delete and insert rows and columns. But in this lesson, we'll go farther.
We'll start at the cell level with cell formulas. A typical use case in working with spreadsheets is to put together a formula in a cell and then duplicated in more rows. It all starts with getting the first cell formula right, and you're going to learn how to do that.

- You'll learn to manipulate text data with functions and combine and split text in various ways.
- Then, you'll learn to manipulate numerical data using basic math and statistical functions.

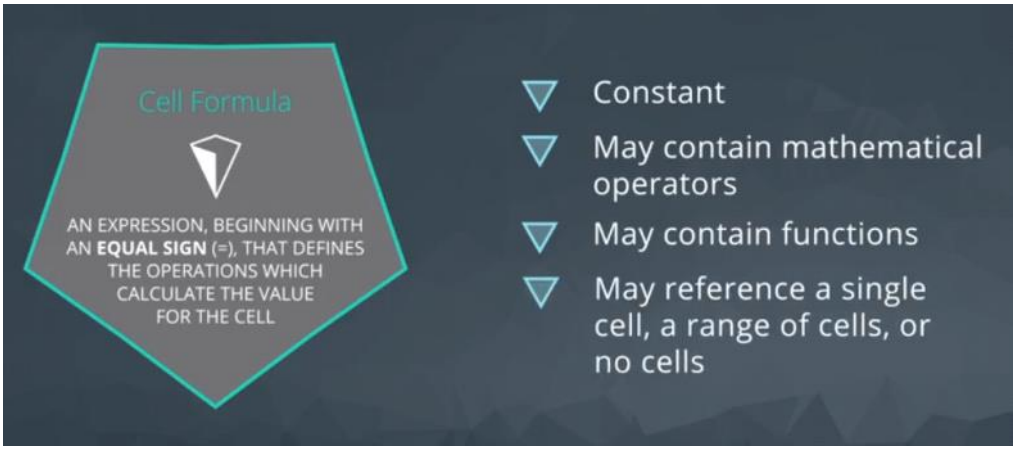
After working at the cell level, we'll move on to table level data manipulations.

- You'll learn how to remove duplicate rows and how to split columns. These procedures are useful for cleaning your data.
- You'll learn how to sort and filter your data. The filtering tools have a number of powerful information features built in. So, we'll take time to explore these features.

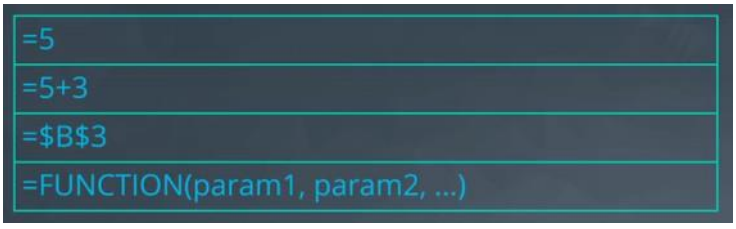
Cell Formulas

It is common when working with spreadsheets, to create a single cell formula and then duplicate it with a fill to create a column or row. So, it's important to get that first formula right.

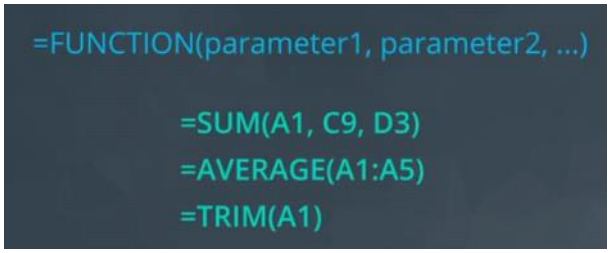
A cell formula is an expression that defines the operations for calculating the value of a cell.
It might be a constant or contain mathematical operators or functions, and it can reference a single cell, a range of cells or none at all.



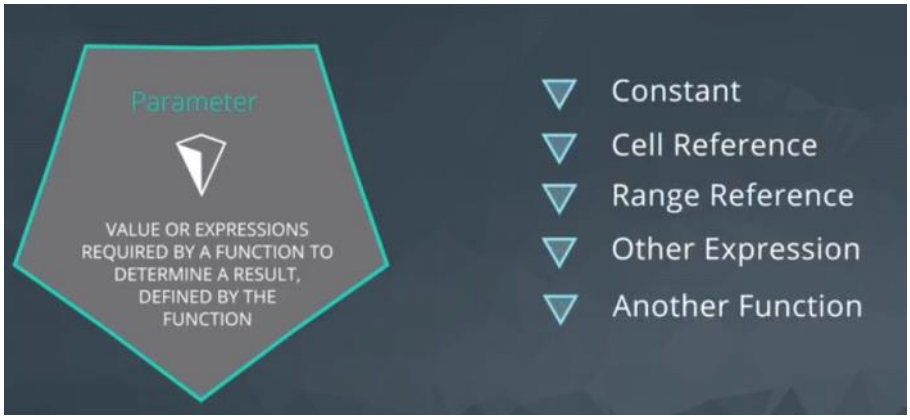
A cell formula begins with an equal sign. This cell is now equal to whatever I type to the right of the equal sign, that can be a number. It could be a math operation like addition or multiplication of two constants. It could be a reference to some other cell. If the original cell is referenced with a formula, then a change to the original cell will automatically change the one that references it. Besides constants and references, functions can be included in formulas and that's where the true power lies.



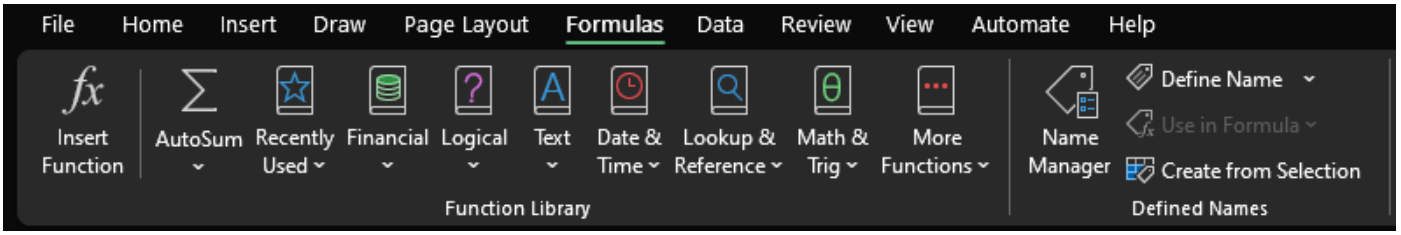
A function is a built-in routine that calculates a value from inputs provided. Functions always start with some key words such as sum or average or trim and so on, followed by a set of parentheses with parameters in between.



A parameter might be a constant or a cell reference or range reference or other expression that is required for the function calculation. It's even possible to nest functions. They can become quite complex.



You have to look at the function definition to know what kind of parameters are expected and how they will be used by the function. A function allows us to use data we already have in a new way.



There are many types, and you can take a closer look at them in the function library on the formula menus tab.
You'll learn some useful ones in these introductory lessons. There are over 400 though, so keep in mind that there are more to explore.

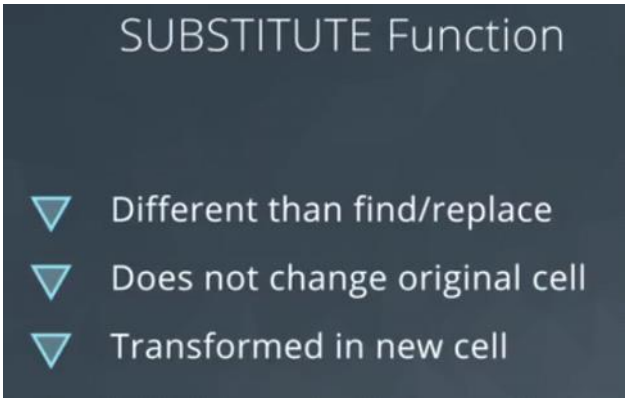
Definition	Function
Converts a text string to upper case letters.	UPPER(text)
Returns the logical value TRUE.	TRUE()
Returns the number of characters in a text string.	LEN(text)
Estimates standard deviation based on a sample (ignores logical values and text).	STDEV(n1, n2, ...)
Adds all the numbers in a range of cells.	SUM(n1, n2, ...)
Removes all spaces from a text string except single spaces between words.	TRIM(text)

SUBSTITUTE

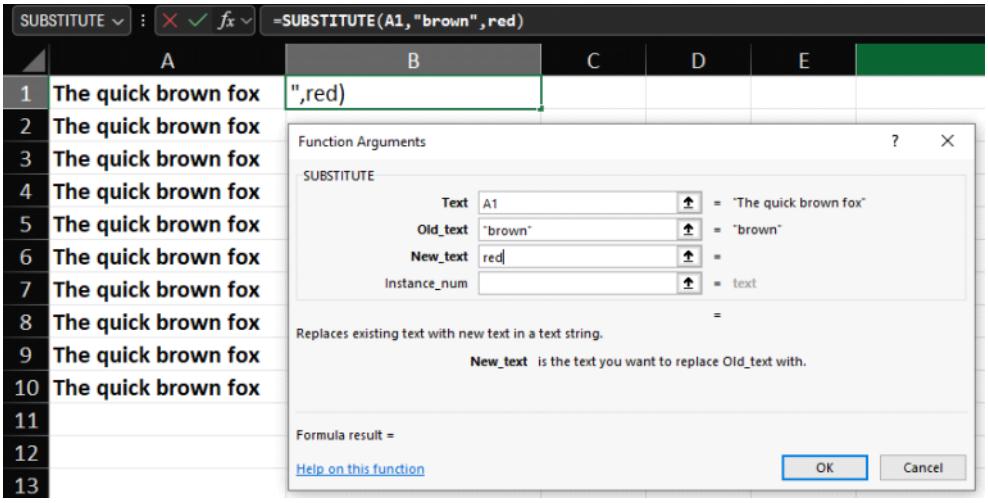
Text functions modify or provide information about text strings that we find in data.
A text string is a series of characters such as letters, numbers, and punctuation that is not treated like a

number.
Examples of text or text strings are; names, phrases, cities, countries, not a date or a number.
There are many text functions in spreadsheets.

We're going to go through a useful subset of these starting with the **Substitute** Function. This is different from the find and replace tools we looked at in the previous lesson. Those were used to make fundamental changes to values that exist in cells. In that case, the original information was destroyed and replaced.
When using a substitute function, the original data remains intact. It's just transformed in a new location that we create.



Let's take a closer look at how to do this. I have a string, the quick brown fox, and I want to replace brown with red.
There are a couple of ways that you can enter a formula in Excel and I'll go over both.
One way is to use the formula bar. Select your cell, then click the f_x button on the formula bar.



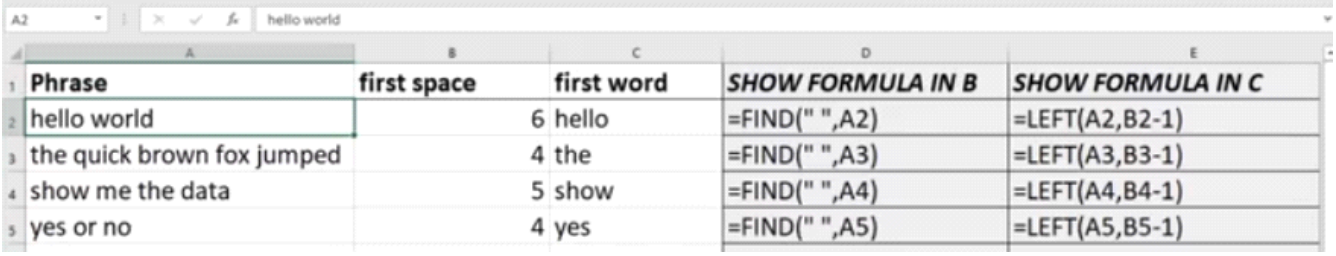
The second way to do this is more direct and a little faster. Select the cell you want and just start typing the formula, starting with the equal sign.

Extract Text

Let's suppose we have a list of phrases and we want to extract the first word from each. How would we do that?

Thinking this through, for the starting word, we want to extract everything from the left up until the first space character.

To do that, we need two functions: Find and Left. I've exposed the formulas so that we can see what's happening more clearly.

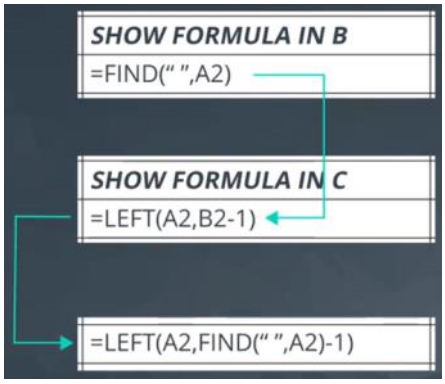


	Phrase	first space	first word	SHOW FORMULA IN B	SHOW FORMULA IN C
2	hello world	6	hello	=FIND(" ",A2)	=LEFT(A2,B2-1)
3	the quick brown fox jumped	4	the	=FIND(" ",A3)	=LEFT(A3,B3-1)
4	show me the data	5	show	=FIND(" ",A4)	=LEFT(A4,B4-1)
5	yes or no	4	yes	=FIND(" ",A5)	=LEFT(A5,B5-1)

The FIND function searches the phrase string for whatever sub-string we give it, in this case, a space, and returns the character position where that sub-string was found. So, for Hello world, the first space is at character number six in the string.

The LEFT function extracts the number of characters we tell it to starting from the left. For Hello world, we're telling the function to extract one less than the first space count that find determined were five characters on the left. This results in the final text of, Hello, which is five characters long.

This could have been done in one step, instead of two, by nesting the find function inside the left function. Once we do this, our equation is only dependent on the original data.



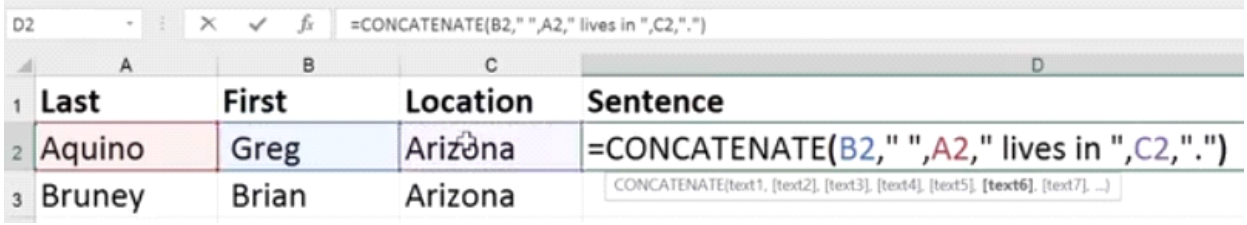
The RIGHT function works in a similar way except that right selects a number of characters from the right. In this case, the last three letters of World which are 'rld'.

MID extracts from anywhere in a phrase based on some starting point, in this case, starting with the second letter, there are three letters extracted, 'ell'.

Reformat Text

In addition to extracting text, there are some useful functions for changing how your text is displayed. One of these is concatenate. Concatenate gives us a way to combine a list of text items from multiple cells easily.
Let's see how that looks with some very simple examples.

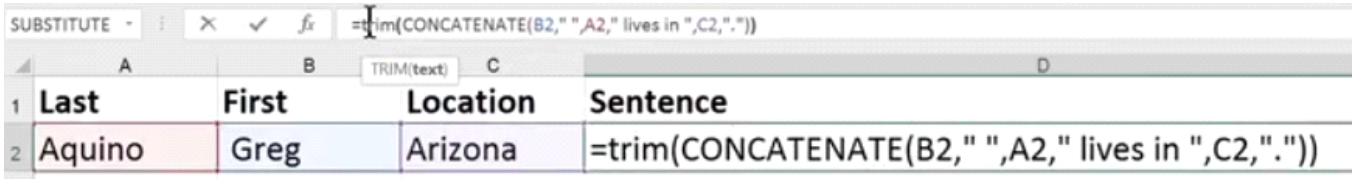
In this example, I have some tabular information: first name, last name and location. I can use concatenate to create sentences with this information. The first name, then a space, the last name, then a phrase, lives in, with spaces, then the location and a period. Fill it down and I have my sentences.



They look a little funny though, I think there may have been some spaces I didn't see in the data. That happens a lot, which is why the trim function is so helpful.

	A	B	C	D
1	Last	First	Location	Sentence
2	Aquino	Greg	Arizona	Greg Aquino lives in Arizona.
3	Bruney	Brian	Arizona	Brian Bruney lives in Arizona.
4	Choate	Randy	Arizona	Randy Choate lives in Arizona.
5	Cintron	Alex	Arizona	Alex Cintron lives in Arizona.

I can just put it right at the front of my other function, with a closing parenthesis on the end, fill it down. There, that fixed it.

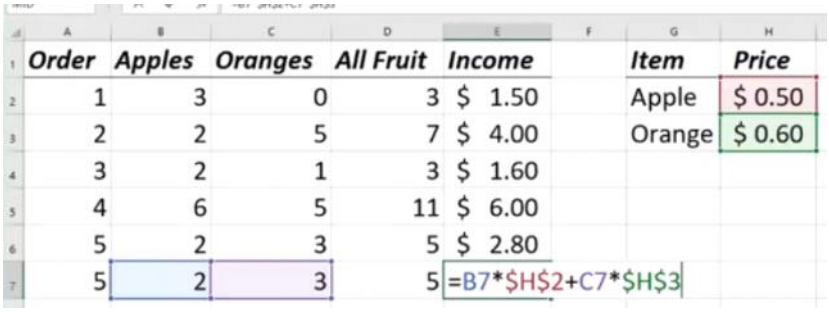


	A	B	C	D
1	Last	First	Location	Sentence
2	Aquino	Greg	Arizona	Greg Aquino lives in Arizona.
3	Bruney	Brian	Arizona	Brian Bruney lives in Arizona.
4	Choate	Randy	Arizona	Randy Choate lives in Arizona.
5	Cintron	Alex	Arizona	Alex Cintron lives in Arizona.
6	Clark	Tony	Arizona	Tony Clark lives in Arizona.

A few more easy functions that can clean up data. To ensure that names and places are properly capitalized, the proper function works great. Upper makes all letters upper case, and lower does the opposite, makes all letters lowercase.

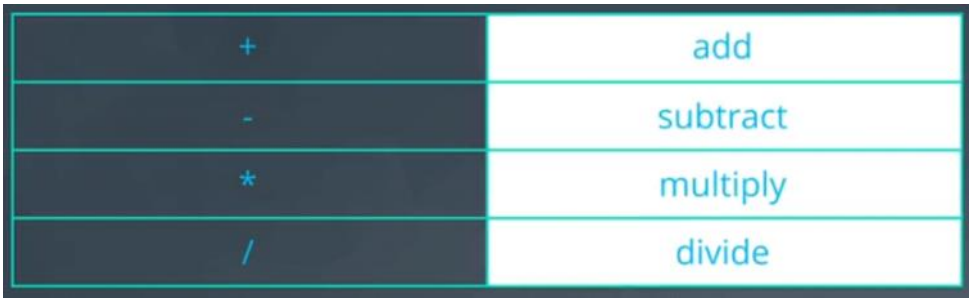
Math Functions

When we think of spreadsheets, we probably think of math operations first.
So it's pretty easy to imagine some applications for them, perhaps adding lists of money or items, determining average prices or maybe multiplying unit cost times the number of units as we did with the fruit stand example.



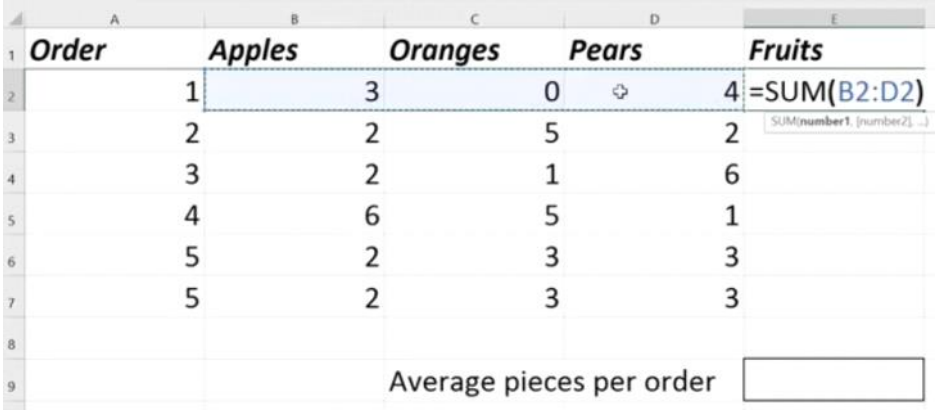
Order	Apples	Oranges	All Fruit	Income	Item	Price
1	3	0	3	$ 1.50	Apple	$ 0.50
2	2	5	7	$ 4.00	Orange	$ 0.60
3	2	1	3	$ 1.60		
4	6	5	11	$ 6.00		
5	2	3	5	$ 2.80		
5	2	3	5	=B7*H2+C7*H3		

There are two kinds of arithmetic operations in spreadsheets, those done with arithmetic operators and those that use functions. Arithmetic operators include addition, subtraction, multiplication and division using simple plus, minus, asterisk and slash symbols in the formula.



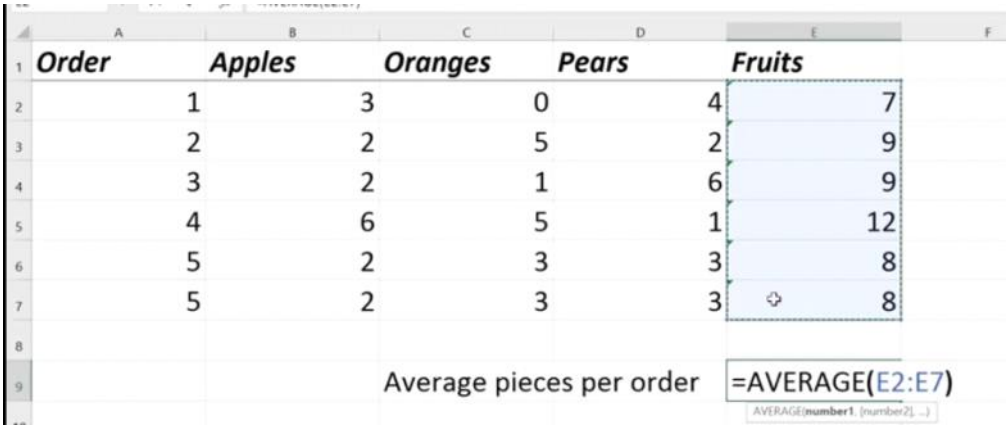
+	add
-	subtract
*	multiply
/	divide

Pretty much as you would actually write these out, as with all spreadsheet formulas, these formulas always begin with an equal sign.
The second type of math operation uses functions. A couple of very useful statistical functions are sum and average. Here I'm using sum to add up a range of numbers in the fruit stand example.



Order	Apples	Oranges	Pears	Fruits
1	3	0	4	=SUM(B2:D2)
2	2	5	2	
3	2	1	6	
4	6	5	1	
5	2	3	3	
5	2	3	3	
		Average pieces per order		

Previously, we did this with arithmetic operators adding each individual cell. It's much easier to just identify a range to sum. Average works the same way, just select the range to get the average of it.



	A	B	C	D	E	F
1	Order	Apples	Oranges	Pears	Fruits	
2	1	3	0	4	7	
3	2	2	5	2	9	
4	3	2	1	6	9	
5	4	6	5	1	12	
6	5	2	3	3	8	
7	5	2	3	3	8	
8						
9			Average pieces per order		=AVERAGE(E2:E7)	

AVERAGE(number1, [number2], ...)

Duplicate Rows

Table level data operations are useful for cleaning and manipulating lists of data. Suppose you import some data that has a lot of good information, but is not what we call clean data. When we clean our data, we're trying to rid it of corrupt and inaccurate data items.

For example, in this little data set, I see some duplicate rows and there are some text anomalies that might throw off my ability to group things later.

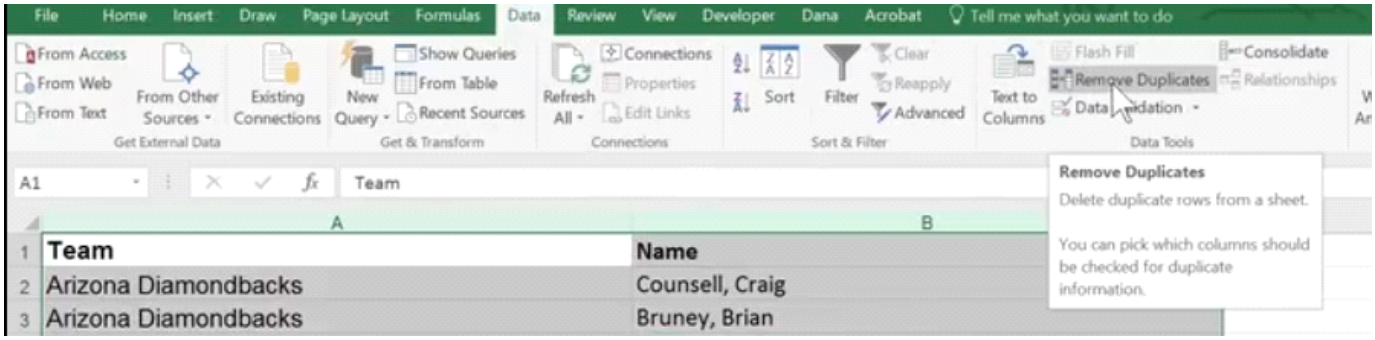
	A	B
1	Team	Name
2	Arizona Diamondbacks	Counsell, Craig
3	Arizona Diamondbacks	Bruney, Brian
4	Arizona Diamondbacks	cHOATE,rYAN
5	Atlanta Braves	Cintron, Alex
6	Atlanta Braves	Glaus, Troy
7	Atlanta Braves	Aquino, Greg
8	Chicago Bears	Cintron, Alex
9	Chicago Cubs	Gonzalez, Luis
10	Chicago Cubs	Clark, Tony
11	Chicago Cubs	Clayton, Royce
12	Arizona Diamondbacks	Cruz Jr, Jose
13	Atlanta Braves	Cruz Jr, Jose
14	Atlanta Braves	Estes, Shawn
15	Atlanta Braves	Gil, Jerry
16	Chicago Cubs	Gonzalez, Luis
17	Chicago Cubs	Gosling, Mike

4	Arizona Diamondbacks	cHOATE,rYAN
5	Atlanta Braves	Cintron, Alex
6	Atlanta Braves	Glaus, Troy

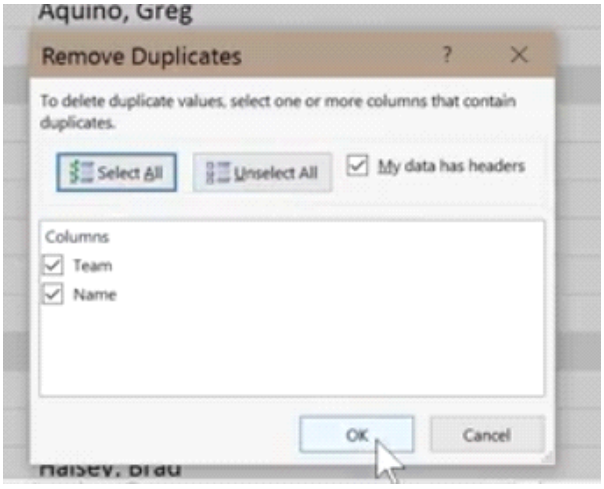
We've already learned that some text and formatting functions can be used to help clean individual cells such as how to fix this name, which seems to have its capitalization completely backwards.

4	Arizona Diamondbacks	cHOATE,rYAN

However, those cell level functions won't help us delete duplicate rows. There is a particular Excel operation for that. Select your data set range. Go to the Data menu and find the icon that looks like two columns with an arrow in between.



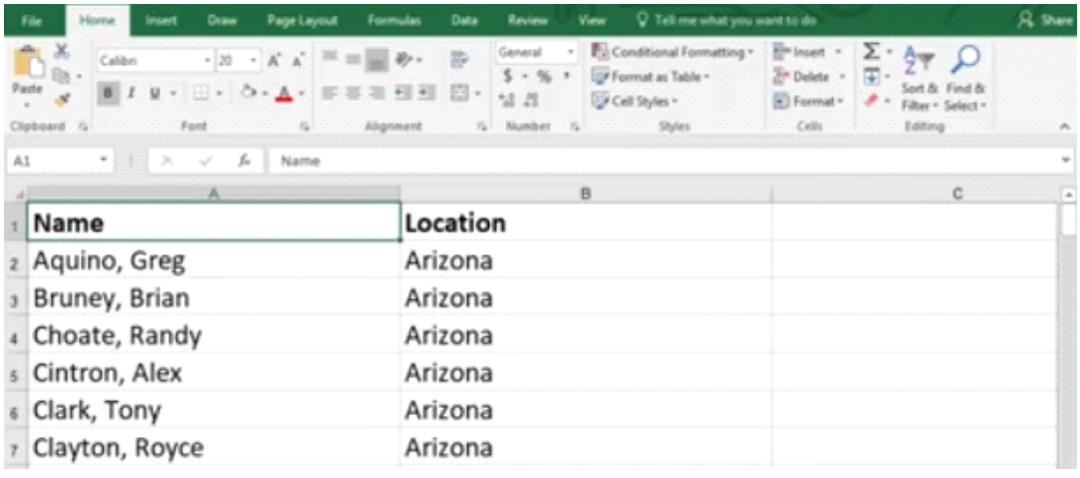
You have some choices to make in the dialog that opens. If you select My data has headers, the operation will ignore the top row of your selection. If you only want to look for duplicates in certain columns, choose those columns.



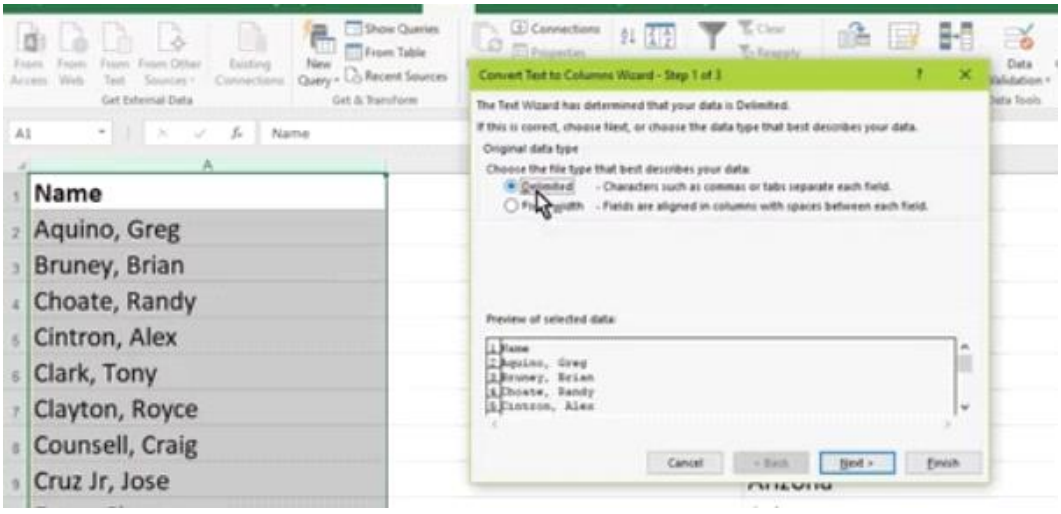
In this example, I'm looking for duplications that are exact matches in both columns. Click okay and the duplicate row is removed.

Split Columns

Another common data cleaning operation, is splitting data, that's combined in one cell. For example, suppose we have first and last names in the cells but are going to want to sort or filter by just the first or just the last name later. In that case, we're going to need a column for each. So, it would be nice to have a first and last names split into two separate columns.



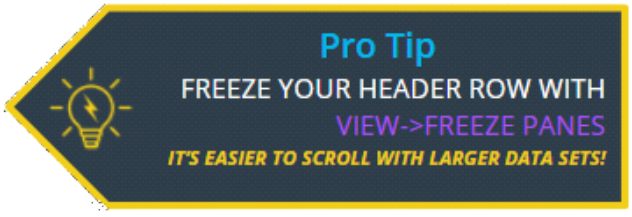
We can do this with the text to columns tool. Before you start, think about how many columns you're going to end up with. In this case, it's one column turning into two. So, insert a blank column to make room. Select your data set range. It's okay to select whole columns. Go to the **Data menu** and find, **text to columns**. Once again, you'll have some choices to make in the dialog that opens.



Delimited means that the text you want to separate has some character where the split should be. After clicking next, choose which delimiter to split on. We'll check the comma because our data is separated with commas.
Clicking next, there are some more advanced options but for our purposes, we can just click Finish.

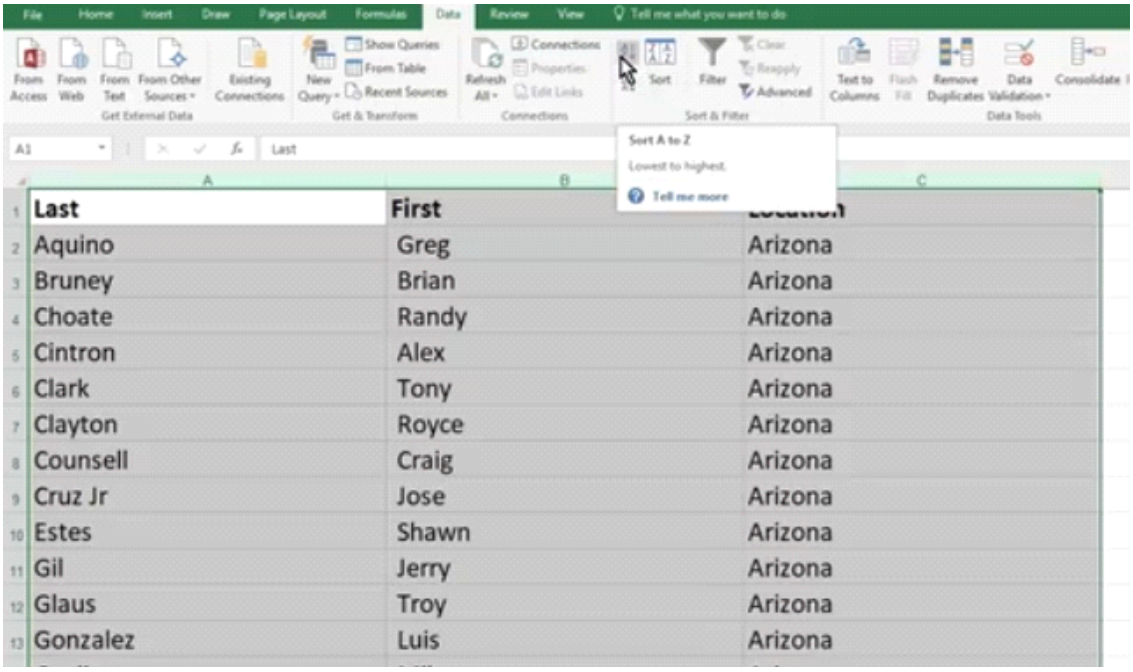
	Last	First	Location
2	Aquino	Greg	Arizona
3	Bruney	Brian	Arizona
4	Choate	Randy	Arizona
5	Cintron	Alex	Arizona
6	Clark	Tony	Arizona
7	Clayton	Royce	Arizona
8	Counsell	Craig	Arizona
9	Cruz Jr	Jose	Arizona
10	Estes	Shawn	Arizona

As you work with larger data sets, its can be helpful to "freeze" the header row or far left column as you scroll through data. To do this, use the **View->Freeze Panes** operation in Excel or **View->Freeze** in Google Sheets to select the rows or columns you want to always be visible.

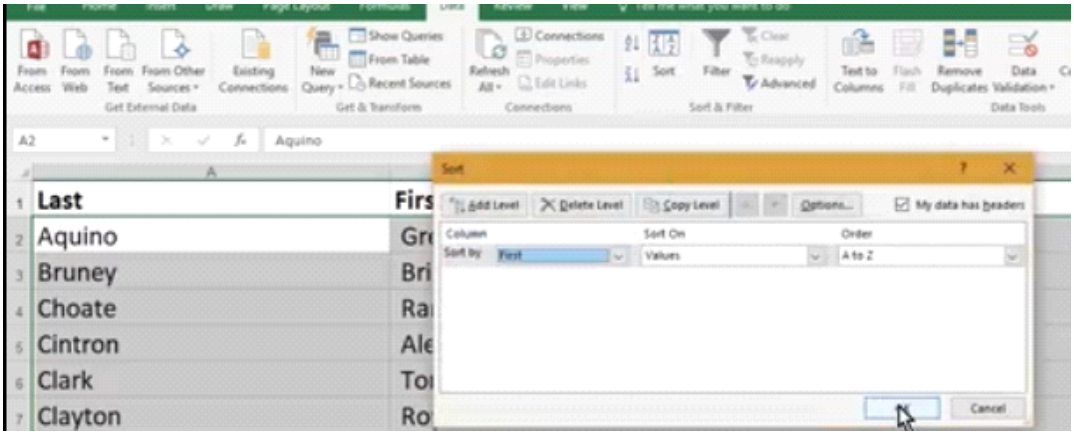


Sort Data

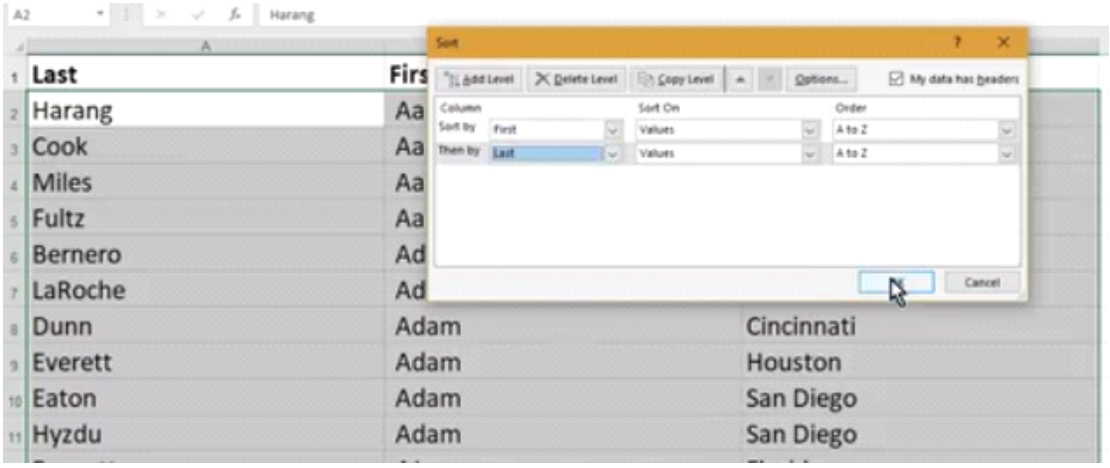
Now that the first and last names are separated, let's try sorting the list by first name.
Select all the data. It's easiest to select whole columns but not required. The sort functions are in the **Data menu**. There's a little A to Z icon.



By default, the sort will be on the first column. Well, this didn't work very well in our case because we want to sort by first name which happens to be in the second column. Let's try again. Choose the sort option and then there's a dialog to specify columns and other details. Here we can choose to sort by the first name column, A to Z.



But wait, I have a couple of first names that are the same yet the last name is out of order. Let's go back to that dialog again. We need to add a level. Now, we'll sort by first name then last and now, both are in the right order.



Last	First	Location
Cook	Aaron	Colorado
Fultz	Aaron	Philadelphia
Harang	Aaron	Cincinnati
Miles	Aaron	Colorado
Bernero	Adam	Atlanta
Dunn	Adam	Cincinnati
Eaton	Adam	San Diego
Everett	Adam	Houston
Hyzdu	Adam	San Diego
LaRoche	Adam	Atlanta
Burnett	AJ	Florida
Otsuka	Akinori	San Diego
Leiter	Al	Florida
Reyes	Al	St. Louis
Pujols	Albert	St. Louis
Cintron	Alex	Arizona
Escobar	Alex	Washington
Gonzalez	Alex	Florida
Amezaga	Alfredo	Colorado

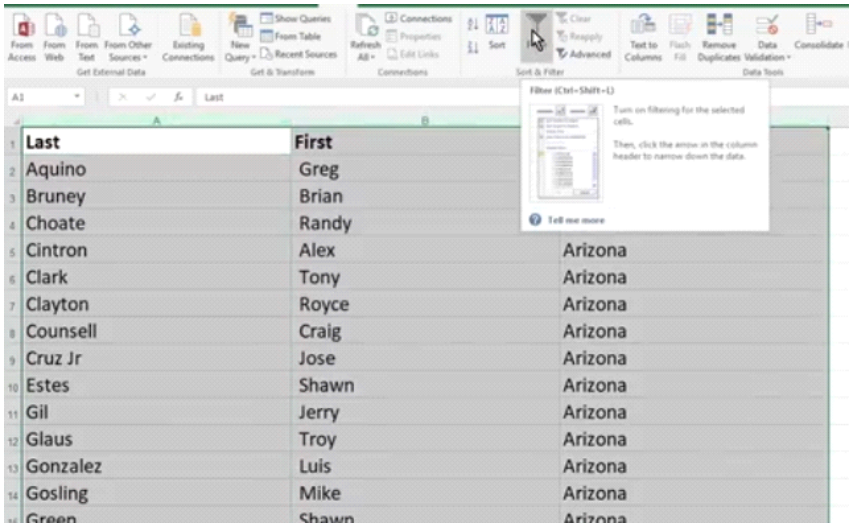
Filter Data

Filtering data is a way to group data by selecting characteristics from our data columns and not looking at the other data.

Looking at a list of names, suppose we want to find every one with the first name Luis.

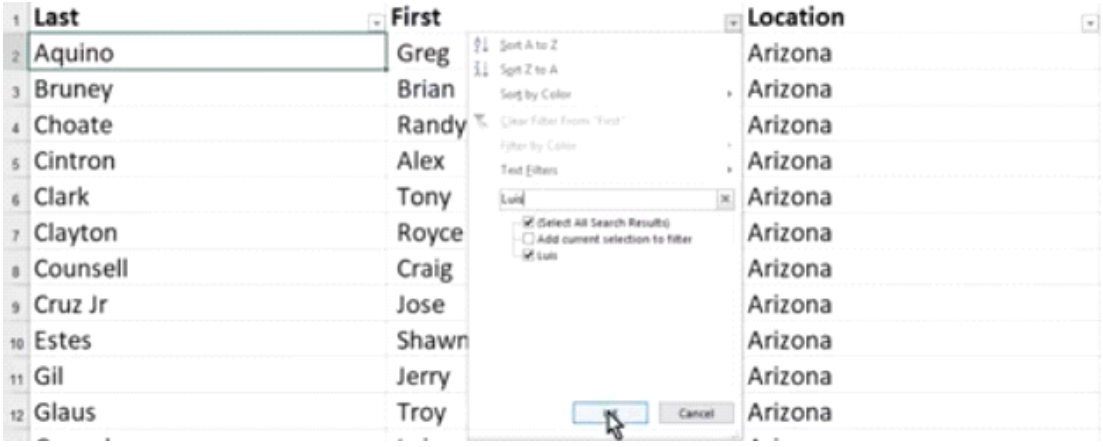
We could sort on the first name and scroll down and find everyone named Luis.

But this changes the data and maybe we don't want to do that. It seems like there's a better way and there is using filters. Select your data. On the Data menu, click the filter icon, which looks like a funnel.



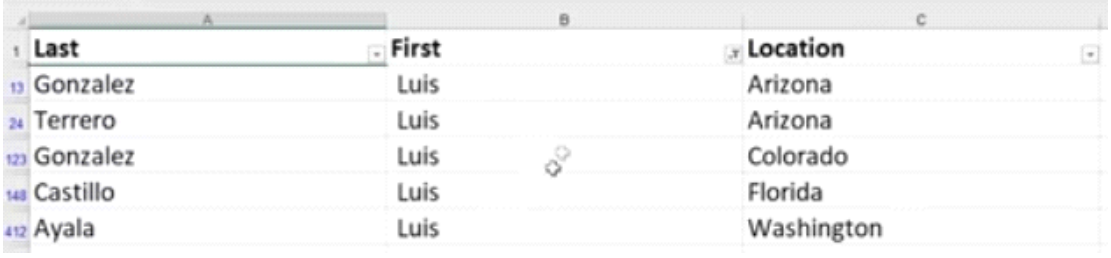
	Last	First	
2	Aquino	Greg	
3	Bruney	Brian	
4	Choate	Randy	
5	Cintron	Alex	Arizona
6	Clark	Tony	Arizona
7	Clayton	Royce	Arizona
8	Counsell	Craig	Arizona
9	Cruz Jr	Jose	Arizona
10	Estes	Shawn	Arizona
11	Gil	Jerry	Arizona
12	Glaus	Troy	Arizona
13	Gonzalez	Luis	Arizona
14	Gosling	Mike	Arizona

When the filter is on, we see little filter boxes with pull down arrows at the top of each column. This is where you choose filters. Click the filter box on the first name. There's a dialog. There are some choices for sorting which we don't want right now. And there's a list of check boxes for every value found in the column.
To look at only at the Luis names, de-select all the boxes and select only the Luis box. Or you can search for it and click okay.



	Last	First	Location
2	Aquino	Greg	Arizona
3	Bruney	Brian	Arizona
4	Choate	Randy	Arizona
5	Cintron	Alex	Arizona
6	Clark	Tony	Arizona
7	Clayton	Royce	Arizona
8	Counsell	Craig	Arizona
9	Cruz Jr	Jose	Arizona
10	Estes	Shawn	Arizona
11	Gil	Jerry	Arizona
12	Glaus	Troy	Arizona

We now see only the matches. At this point, we could add additional filters or copy our subset to somewhere else.



	A	B	C
1	Last	First	Location
13	Gonzalez	Luis	Arizona
24	Terrero	Luis	Arizona
123	Gonzalez	Luis	Colorado
148	Castillo	Luis	Florida
412	Ayala	Luis	Washington

To return your view to the full original data set, click on the clear icon or just turn filtering off altogether by clicking on the filter funnel again.

Intro to Analyze Data

In the previous lesson, you learned techniques for **manipulating data**. In this lesson, you'll explore

additional functions and techniques useful for **analyzing data** in your spreadsheet.

What Does "Analyze the Data" Mean?

Analyzing data involves starting with raw information—our data—and extracting meaningful insights by asking and answering questions.
For example:

- How many apples and oranges were ordered in the fruit stand example?
- What's the average salary of an American League Baseball player in 2003?

This process often uses **aggregation functions** to summarize data. You've already encountered two of these:

- SUM
- AVERAGE

In the previous lesson, you used these functions to begin analyzing data.

Aggregation Functions

In this lesson, you'll learn about several additional aggregation functions to answer more complex questions.

Logical Functions

Aggregation alone may not address all your questions. Sometimes, we need to determine if certain conditions are met, such as:

- "If a student scored above a certain value on a test, assign them an A."

For these conditional scenarios, you'll learn about **logical functions**, including:

1. IF function and its comparison operators
2. Logical functions:
    - AND
    - OR
    - NOT

Conditional Aggregation Functions

You'll also explore logical functions that combine **aggregation** with **conditions**, referred to as **conditional aggregation functions**.

Pivot Tables

Next, you'll learn to create **pivot tables**, a powerful tool for summarizing and aggregating data efficiently in one step.

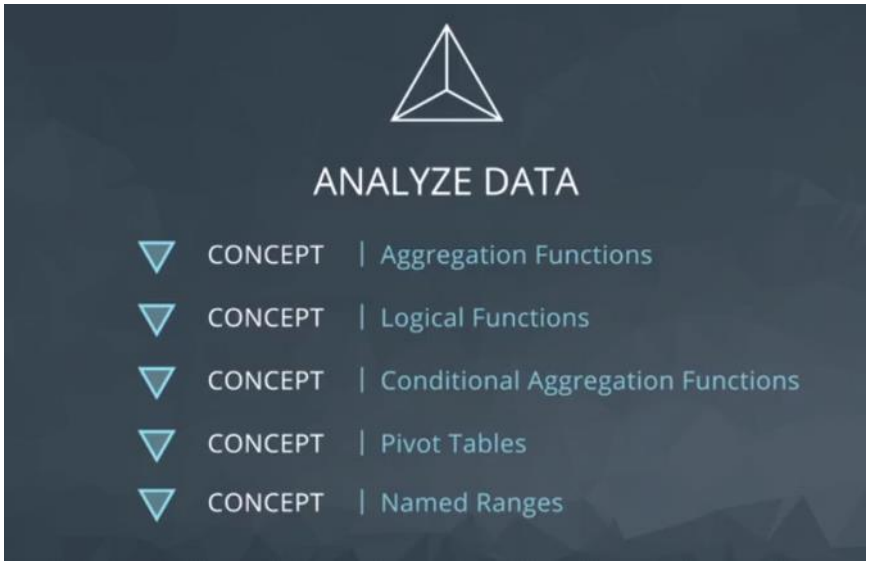
Named Ranges

We'll then shift focus to **named ranges**. These simplify working with large datasets by replacing complex cell addresses (e.g., Sheet3!A1:J943) with descriptive names like **StudentRoster**. Named ranges:

- Reduce errors
- Make formulas easier to understand

- Are especially useful when working with **lookup functions**

Lookup Functions
Finally, you'll learn to use **lookup functions** to find and retrieve specific data from your spreadsheets.
Let's Get Started!



Aggregation Functions

In an earlier lesson, you learned two ways to add numbers:

1. **Using the plus sign (+)**, which requires referencing each cell individually.
2. **Using the SUM function**, which allows you to reference an entire range of cells at once.

The SUM function is an **aggregation function**, meaning it operates over a group of data to produce a single value.

Key Aggregation Functions

Aggregation functions are especially useful for statistical analysis. The primary ones include:

- **SUM**: Adds all values in a range.
- **AVERAGE**: Calculates the mean of values in a range.
- **MAX**: Finds the largest value in a range.
- **MIN**: Finds the smallest value in a range.
- **MEDIAN**: Determines the middle value in a sorted range.
- **STANDARD DEVIATION**: Measures the spread of values around the average.

Example: Fruit Stand Data

Imagine a fruit stand with various products:
- To find the total number of fruits sold, you could:
    1. Add each row of data manually and then total those sums.
    2. Use the **SUM** function to aggregate the entire range at once.
- To answer specific questions, such as:

- **"What was the most fruit sold in a single order?"** Use the **MAX** function.
- **"What was the least fruit sold in a single order?"** Use the **MIN** function.
- **"What is the median value of all orders?"** Use the **MEDIAN** function.

	A	B	C	D	E	F	G	H
1	Order	Apples	Oranges	Pears	Kiwi	Fruit Total		
2	1	3	0	3	1	7		
3	2	2	5	6	9	22		
4	3	2	1	2	4	9		
5	4	6	5	5	9	25		
6	5	2	3	9	12	26		
7	6	2	6	4	3	15		
8	*Grand Total*					**104**		
9	*Another Grand Total*					**104**		
10	*What's the maximum number of fruit in an order?*					**26**		
11	*What's the minimum number of fruit in an order?*					**7**		
12	*What's the median number of fruit in an order?*					**18.5**		
13	*Average*					**17.33333**		
14	*Standard Deviation*					**8.21381**		
15								

The formula
=SUM(F2:F7)
=SUM(B2:E7)
=MAX(F2:F7)
=MIN(F2:F7)
=MEDIAN(F2:F7)
=AVERAGE(F2:F7)
=STDEV(F2:F7)

Statistical Insights

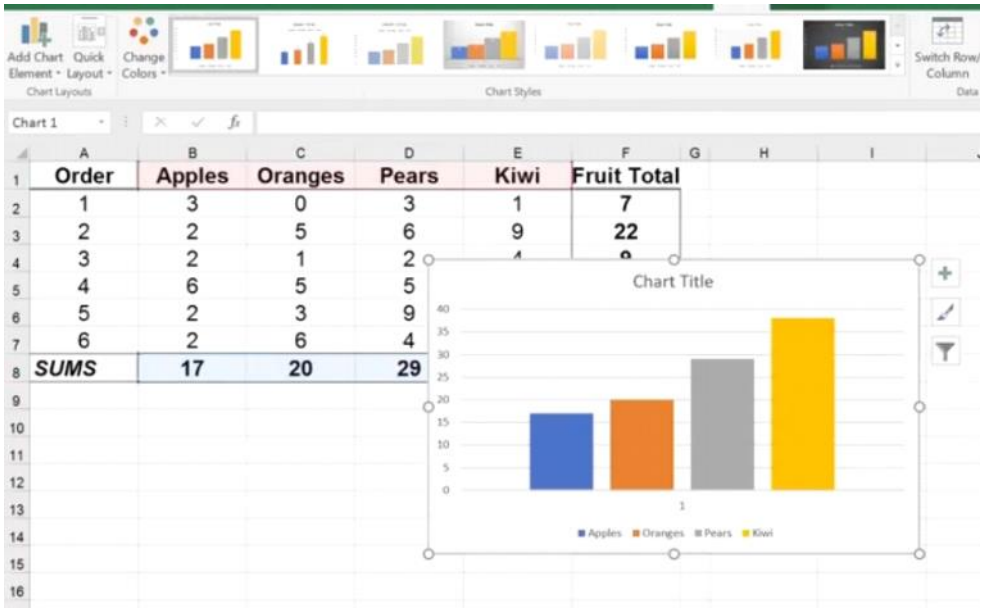
Using **AVERAGE** and **STANDARD DEVIATION**, you can analyze the spread of total order values over time, identifying trends or anomalies.
For more advanced aggregation functions, refer to the **Help files** in your spreadsheet application.

Teaser: Data Visualization

Aggregation functions become even more powerful when paired with data visualization.

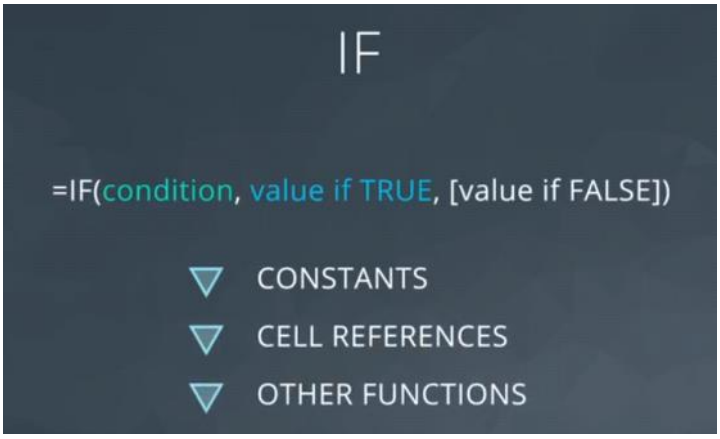
- For example, to identify the best-selling fruit, you could highlight your data, select **Insert Chart**, and choose a bar chart.
- Visualizing aggregated data provides impactful insights at a glance.

In the upcoming **Visualized Data** lesson, you'll learn when and how to use various chart types effectively.



Logical Functions: IF

Conditional functions and spreadsheets are part of the logical functions group.
We'll start with If.
The function If, like other functions, starts out with an equal sign, then if, then some parameters which are a condition, what to do if the condition is true, and optionally what to do if it's false.

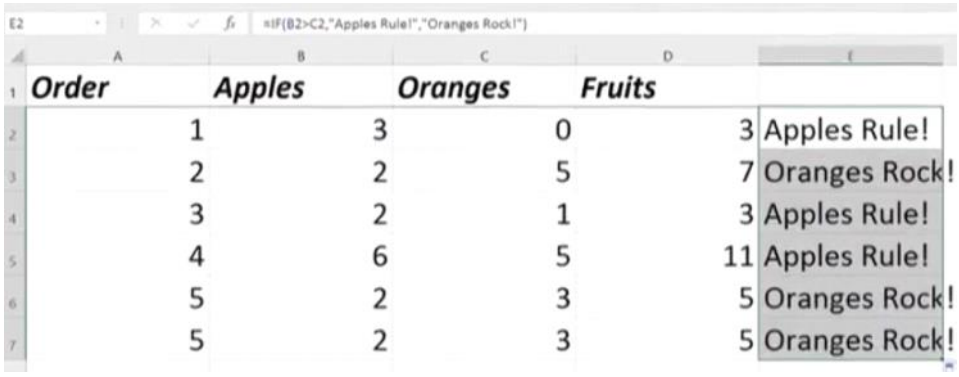


These parameters can be constants or cell references or other functions. Here's an example.
A conditional sentence written in English might be something like, if there were more apples than oranges in the order, print Apples Rule.
The condition is that there are more apples than oranges, and the value, if true, is the statement, Apples Rule. How would we do this in a fruit stand spreadsheet? It's a function, so start with equal, if, open parentheses, the apples sell, now a comparison operator meaning greater than, then the oranges sell, comma, and what to print in double quotes: Apples Rule, close parentheses and fill down.

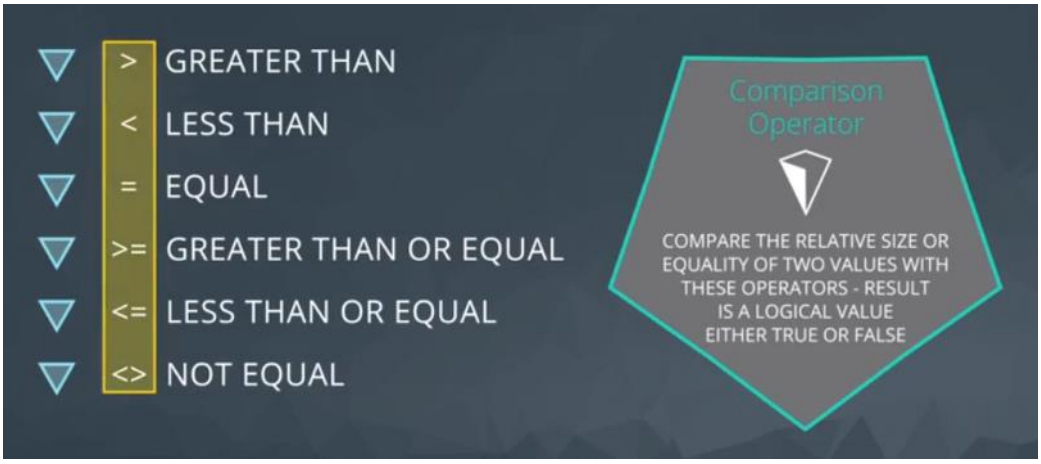
E2 =IF(B2>C2,"Apples Rule!")

	A	B	C	D	E
1	Order	Apples	Oranges	Fruits	
2	1	3	0	3	Apples Rule!
3	2	2	5	7	FALSE
4	3	2	1	3	Apples Rule!
5	4	6	5	11	Apples Rule!
6	5	2	3	5	FALSE
7	5	2	3	5	FALSE
8					

Notice that it just says false, if apples don't rule because I didn't fill in what to do if the value is false.
I can fix that by changing my statement to include an output such as Oranges Rock for the false situation.
So my new statement in English would be, if there are more apples than oranges in the order, print Apples Rule, else, print Oranges Rock.



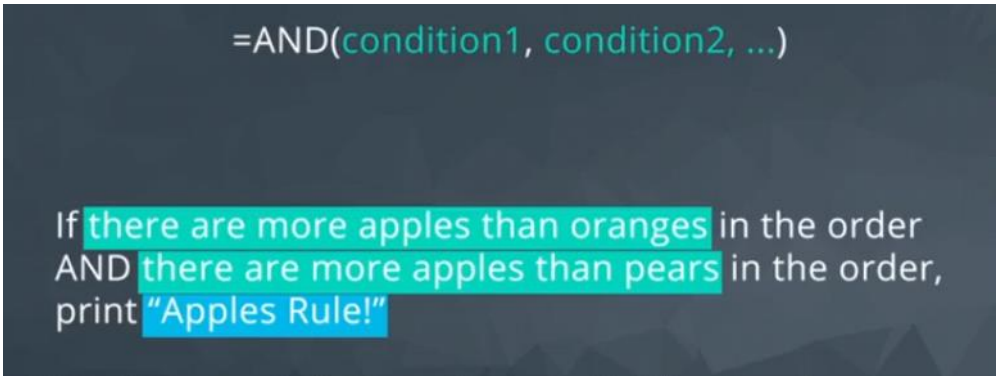
For this little example, I use the greater than sign as a comparison operator. Comparison operators are used in logic statements to compare two values.
In addition to greater than, there are operators for less than, equal, greater than or equal, less than or equal, and not equal.



They're shown here with an operator symbol in Excel. Any of these comparison operators can be used to create a true or false condition, and that's what the if statement will evaluate following your instructions for each.

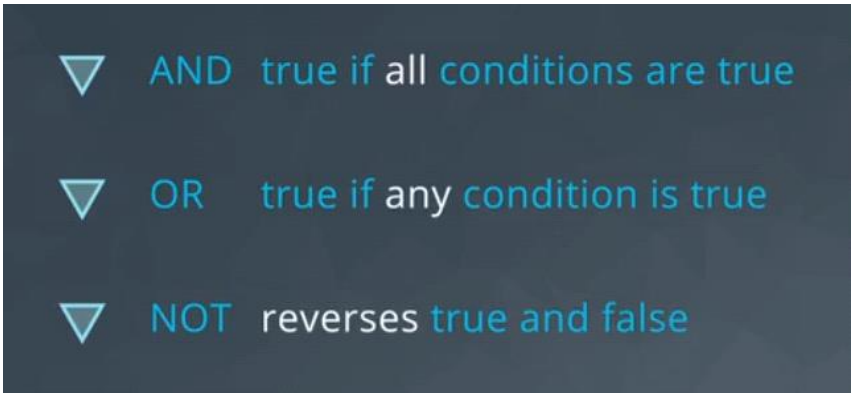
Logical Functions: AND, OR, NOT

What if we want to know whether there are more apples than any other fruits in the list? In that case, our sentence in English would be something like, if there are more apples than oranges in the order, AND there are more apples than pairs in the order, print 'Apples Rule!' AND is another logical function in Excel, which is only true if all of its conditions are true.



Notice that AND like IF is a function not an operator. It has to be entered with its keyword AND followed by parentheses with its parameters which are conditional statements.
To use the AND function in an IF statement, we will have to nest it inside the IF function. The entire AND function is now the conditional part of the IF function.

Recapping. The result from the **AND** function is true if all its conditions are true, and false otherwise.
The **OR** function works much the same way except that it evaluates as true if any of its parameters are true.
And one more for you, the **NOT** function simply reverses the logic. It changes true to false and false to true.
That may seem complicated, but you'll get better at it with practice.



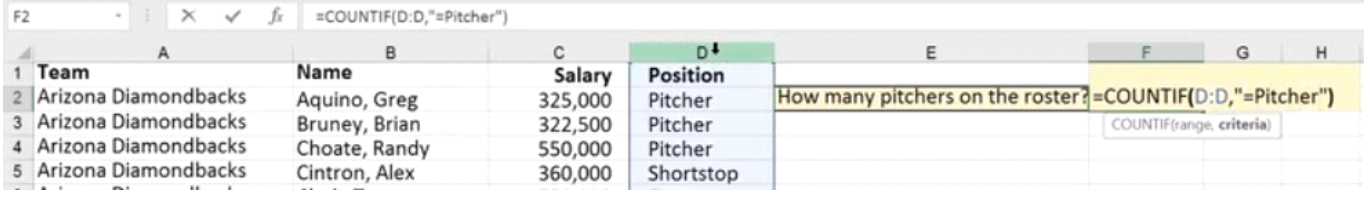
Conditional Aggregation Functions

Conditional aggregation functions are statistical functions that combine aggregation with logical conditions. They operate across a group of data with some logical conditions.
We'll focus on two of these: **COUNTIF** and **SUMIF**.

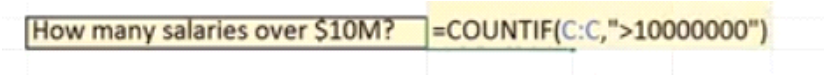
Let's say, you have a baseball data set and that your analysis question is, how many pitchers are there on the roster? We could filter the data by pitcher and count the number of lines showing or we could just use a formula.
With COUNTIF, we can get this information with a single function:

=COUNTIF(range, criteria)

- **Range**: The group of cells to evaluate.
- **Criteria**: The condition to test (e.g., a number, text string, or expression).



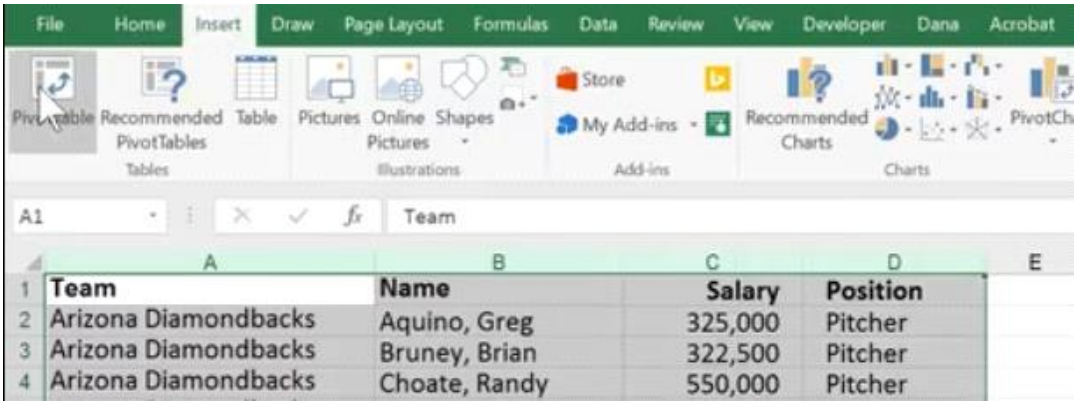
If we wanted to count the number of salaries greater than 10 million, we would need an expression with the comparison operator greater than 10 million. An expression must also be in double quotes.



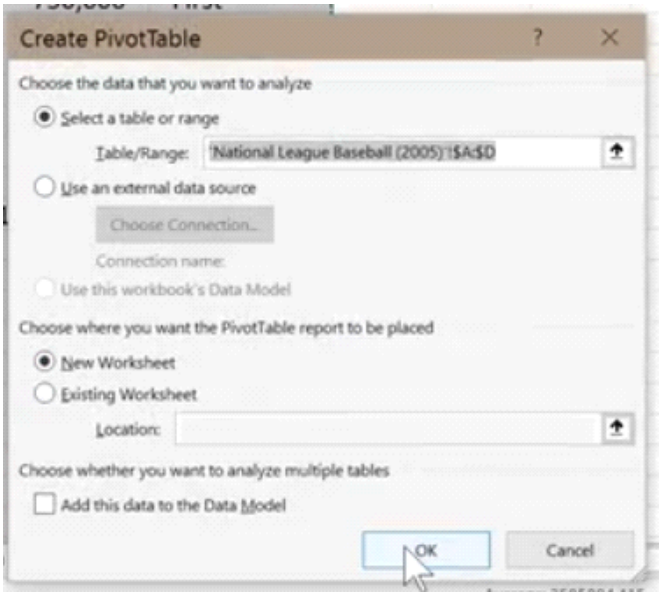
SUMIF works similarly but adds values if the criteria is true rather than counting them.

Pivot Tables

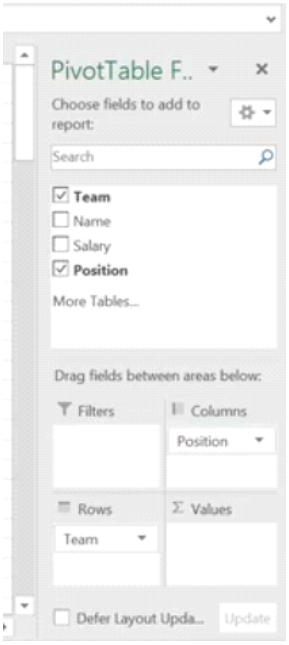
Pivot tables summarize an aggregate all in one step. In a baseball dataset, we see several rows of each team and several rows for each playing position. A pivot will let us group all the teams and all the positions. Select the dataset and go to insert pivot table.



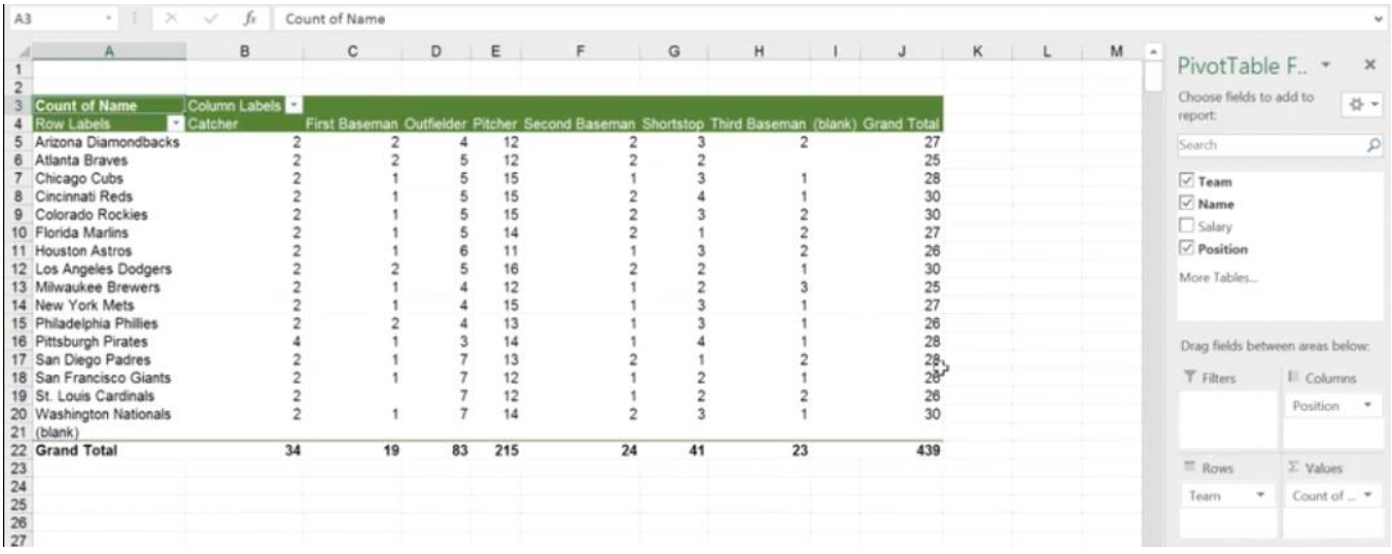
The default is to put it in a new sheet and that's what we want.



Now, we have some drag and drop dialogs. We're looking for the rows and columns to be the grouped categories of teams and positions, and the cells in between to be the aggregations of those groups such as a count of how many pitchers on each team.
So let's slide the team values to rows and here they are now listed, one for each unique team, and the playing positions to columns and here they are, unique positions listed.



Now, we have salaries and names left to choose from for the aggregation. Since we just want to know how many players in each position for each team, it makes sense to just add the names. The default aggregation function is count.



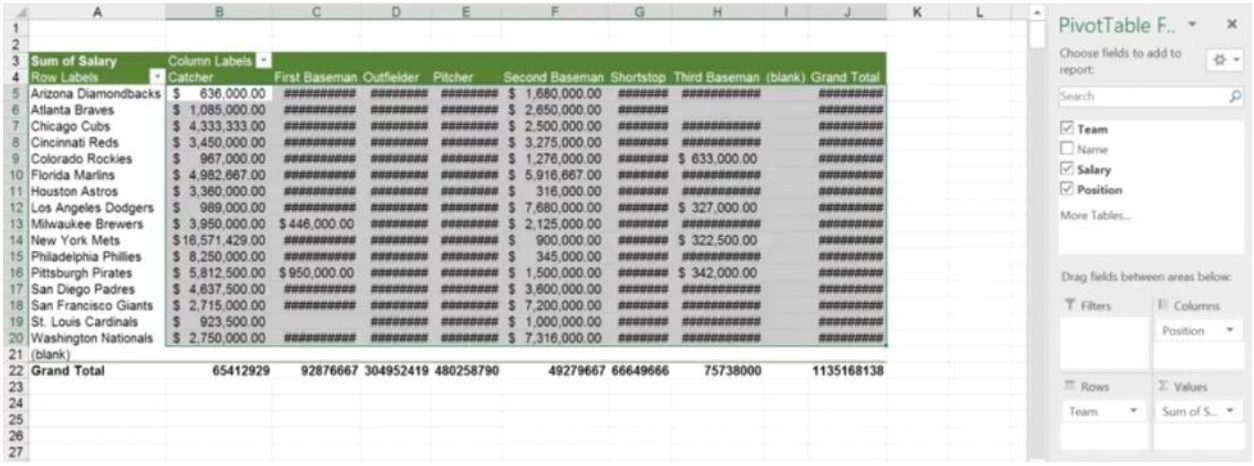
Now, we can look at the table and see how many players in each position each team has on their roster, and how many players are on each team, and how many catchers, first baseman and pitchers there are in the entire league. We just found all of that out in a few clicks.

What if we wanted to know how much money each team spends on which positions? In that case our rows and columns can stay the same, but **the data we need is the salaries**, so replace the names with the salary information.

Nothing really changed because we're still counting the number of salaries not summing them.

To use a different aggregation function, click the pull down menu and select field settings.

There are several that we learned about earlier. Select the Sum aggregation.



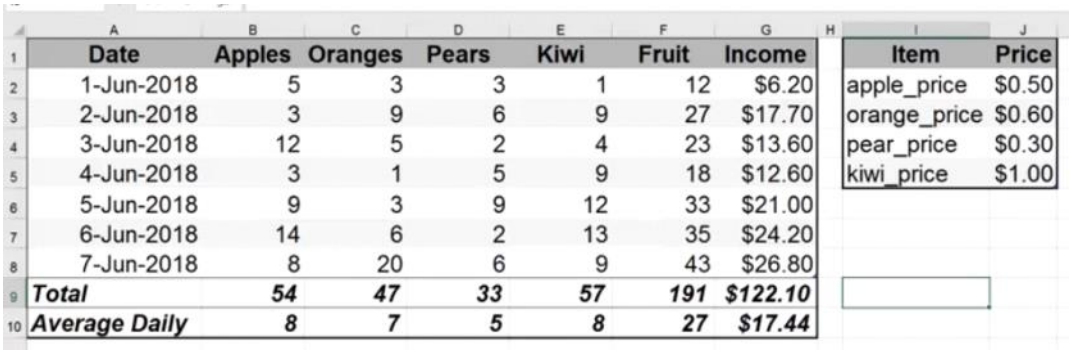
Sum of Salary	Column Labels							
Row Labels	Catcher	First Baseman	Outfielder	Pitcher	Second Baseman	Shortstop	Third Baseman	(blank)	Grand Total
Arizona Diamondbacks	$ 636,000.00	#########	########	########	$ 1,680,000.00	########	#########		#########
Atlanta Braves	$ 1,085,000.00	#########	########	########	$ 2,650,000.00	########			#########
Chicago Cubs	$ 4,333,333.00	#########	########	########	$ 2,500,000.00	########	#########		#########
Cincinnati Reds	$ 3,450,000.00	#########	########	########	$ 3,275,000.00	########	#########		#########
Colorado Rockies	$ 967,000.00	#########	########	########	$ 1,276,000.00	########	$ 633,000.00		#########
Florida Marlins	$ 4,982,667.00	#########	########	########	$ 5,916,667.00	########	#########		#########
Houston Astros	$ 3,360,000.00	#########	########	########	$ 316,000.00	########	#########		#########
Los Angeles Dodgers	$ 989,000.00	#########	########	########	$ 7,680,000.00	########	$ 327,000.00		#########
Milwaukee Brewers	$ 3,950,000.00	$ 446,000.00	########	########	$ 2,125,000.00	########	#########		#########
New York Mets	$ 16,571,429.00	#########	########	########	$ 900,000.00	########	$ 322,500.00		#########
Philadelphia Phillies	$ 8,250,000.00	#########	########	########	$ 345,000.00	########	#########		#########
Pittsburgh Pirates	$ 5,812,500.00	$ 950,000.00	########	########	$ 1,500,000.00	########	$ 342,000.00		#########
San Diego Padres	$ 4,637,500.00	#########	########	########	$ 3,600,000.00	########	#########		#########
San Francisco Giants	$ 2,715,000.00	#########	########	########	$ 7,200,000.00	########	#########		#########
St. Louis Cardinals	$ 923,500.00		########	########	$ 1,000,000.00	########	#########		#########
Washington Nationals	$ 2,750,000.00	#########	########	########	$ 7,316,000.00	########	#########		#########
(blank)									
Grand Total	65412929	92876667	304952419	480258790	49279667	66649666	75738000		1135168138

Now we have the salary sums for each team in position, as well as the subtotals and the grand total. To make this more readable, select the whole area and change the data type to either accounting or currency.

Named Ranges

In the first lesson, when ranges were introduced, I mentioned that it's possible to name ranges instead of referencing them by column and row addresses.

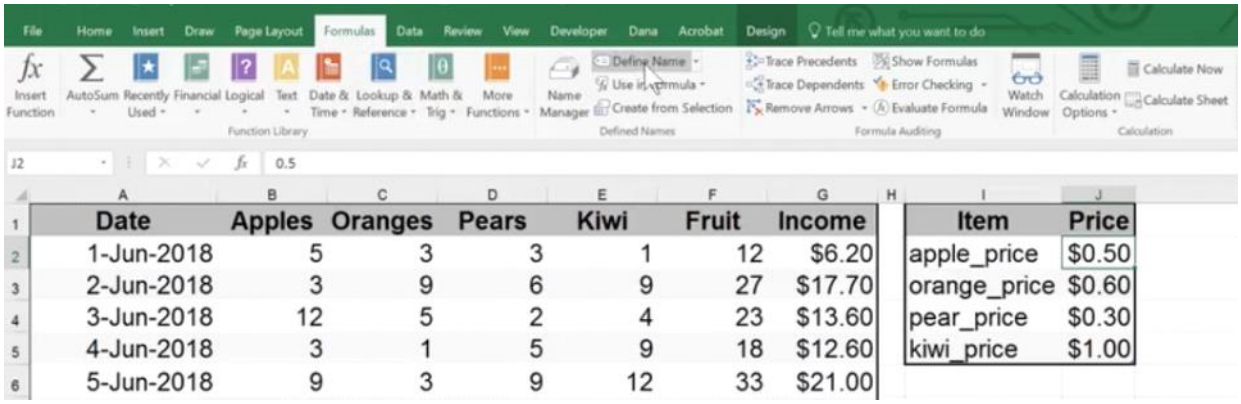
To demonstrate the usefulness of named ranges, we'll go back to the fruit stand spreadsheet. I've changed it a bit with formatting and I've given it dates instead of order numbers. There are four different fruits for sale at the fruit stand, and likewise for item prices.



Date	Apples	Oranges	Pears	Kiwi	Fruit	Income
1-Jun-2018	5	3	3	1	12	$6.20
2-Jun-2018	3	9	6	9	27	$17.70
3-Jun-2018	12	5	2	4	23	$13.60
4-Jun-2018	3	1	5	9	18	$12.60
5-Jun-2018	9	3	9	12	33	$21.00
6-Jun-2018	14	6	2	13	35	$24.20
7-Jun-2018	8	20	6	9	43	$26.80
Total	***54***	***47***	***33***	***57***	***191***	***$122.10***
Average Daily	***8***	***7***	***5***	***8***	***27***	***$17.44***

Item	Price
apple_price	$0.50
orange_price	$0.60
pear_price	$0.30
kiwi_price	$1.00

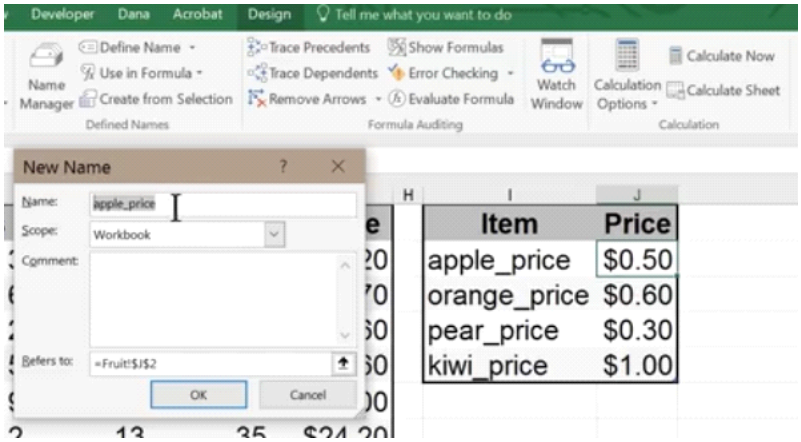
I've also chosen names I want to use for the prices in my formulas. I've chosen Apple price, Orange price, etc. because that's more descriptive than simply apple or orange for the name ranges I am going to create. The basic method to name a cell or range is to select the cell or range. I'm going to select the cell with apple price in it, and then go to the formula menu and click define name.



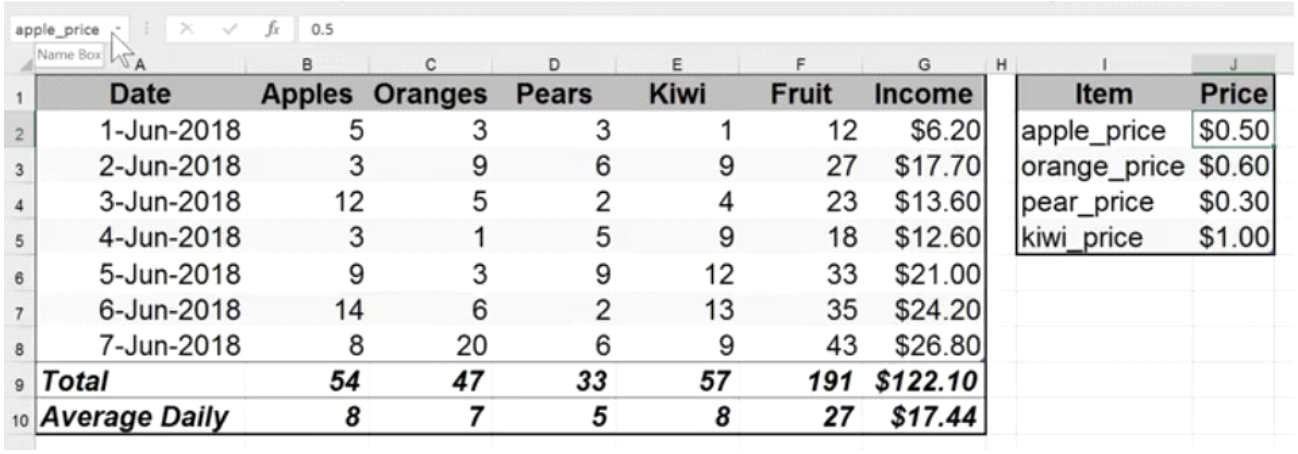
Date	Apples	Oranges	Pears	Kiwi	Fruit	Income
1-Jun-2018	5	3	3	1	12	$6.20
2-Jun-2018	3	9	6	9	27	$17.70
3-Jun-2018	12	5	2	4	23	$13.60
4-Jun-2018	3	1	5	9	18	$12.60
5-Jun-2018	9	3	9	12	33	$21.00

Item	Price
apple_price	$0.50
orange_price	$0.60
pear_price	$0.30
kiwi_price	$1.00

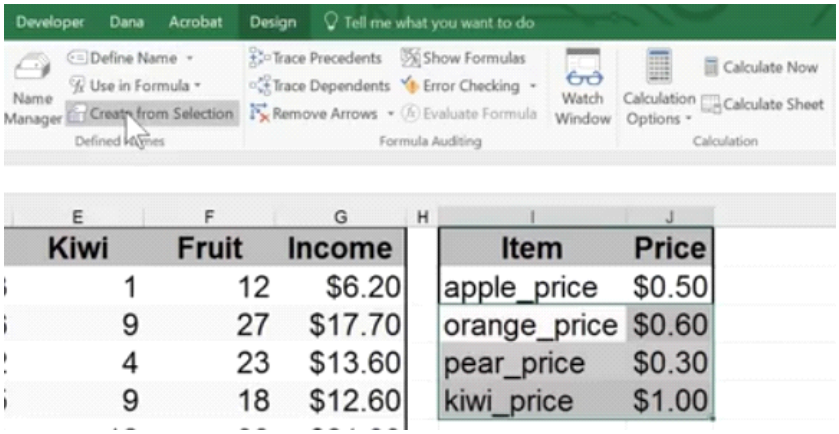
I have a place to type in the name I want to use, but Excel has already assumed I will want to use the text from the adjacent cell to be my label. Excel also shows the range address that will be named.



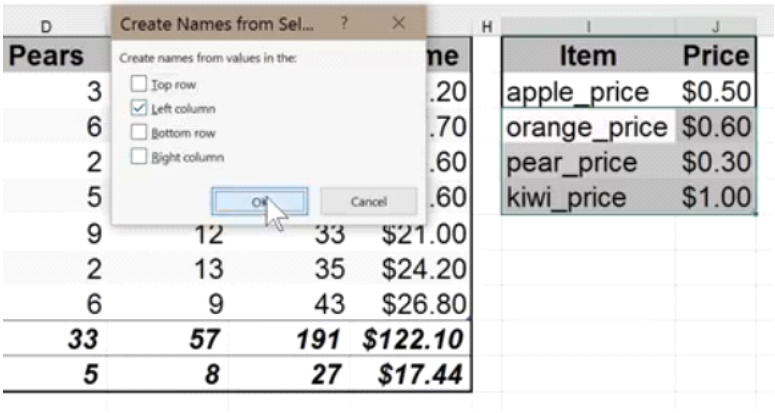
It includes the sheet name followed by an exclamation point, followed by the absolute address. That's what we want here, so I'll just accept it with an OK. Now I have a cell that I can reference not only by its column and row which still works, but by the name apple price.
If I select that cell it even shows up in the name box over here on the left to confirm that this cell is named, apple price.



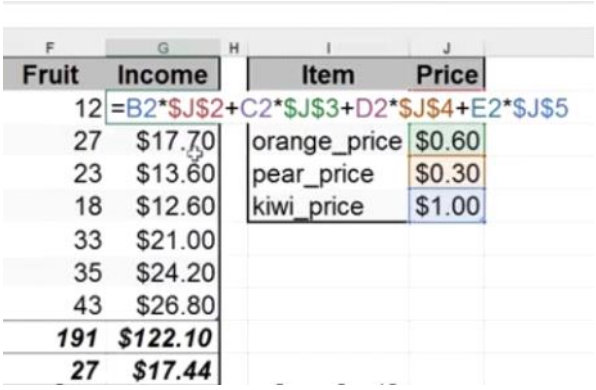
For the other prices, I can repeat the process. But in Excel there's a better way which I'll use for the remaining prices. Select the name and price cells, go to the formula menu and click on Create from selection.



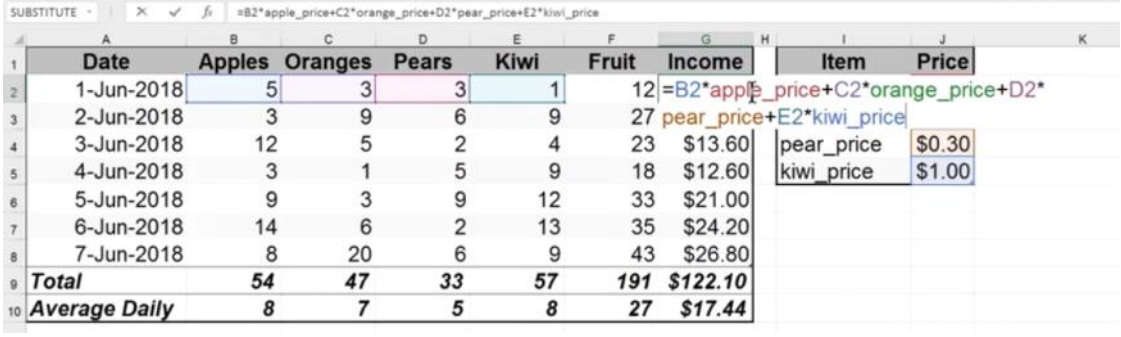
A dialog pops up and excel once again has assumed I want to use the left column for the name, so I'll say OK.



Now, all the price cells have names and we can see them in the name box. When we created our income calculation for the fruit stand before, we went through the task of setting the prices as absolutes when creating the formula.



Here it is and it's long and it's pretty hard to tell exactly what's going on. I'm going to rewrite the formula using named ranges. Again, we have the number of apples times the price which is apple price.



There, that's much more readable. I can see at a glance what this formula is doing and I'll just fill it down. That was one used for named ranges, identifying some single values by name that we went to use in formulas. Named ranges are also useful for naming a larger range to use for lookups and we'll see how to do that next.

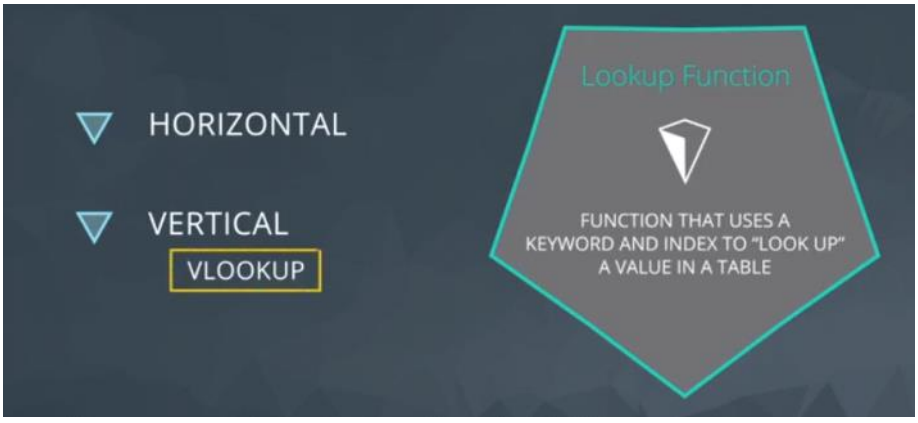
Quiz Question

Which of the following are true about Named Ranges?

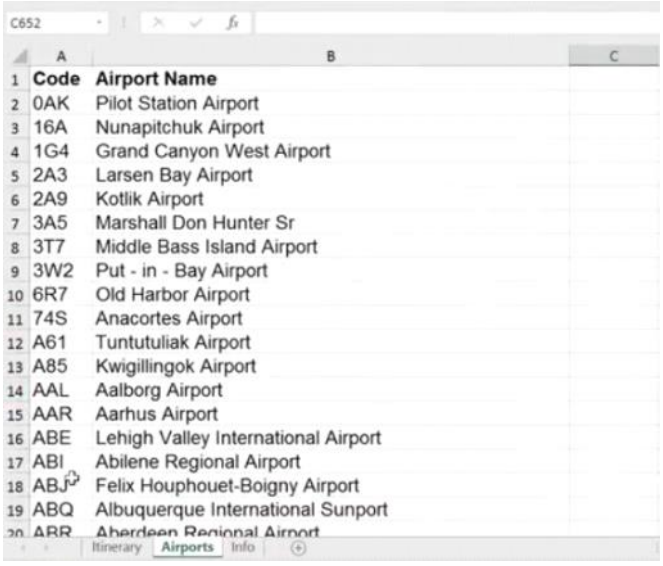
- [x] The addresses are absolute by default ✓
- [ ] The addresses are relative by default
- [x] Formulas using named ranges are easier to create ✓
- [x] Formulas using named ranges are easier to read and maintain ✓

Lookup Functions

Suppose we have a list of part numbers or abbreviations that correspond to prices or descriptions, we might have the short version in one table but need to be able to look up more information elsewhere. That's what Lookup up functions do. They provide a way to use a keyword to look some other information up in a table. There are both horizontal and vertical lookup functions, but we'll focus on just one, VLOOKUP.

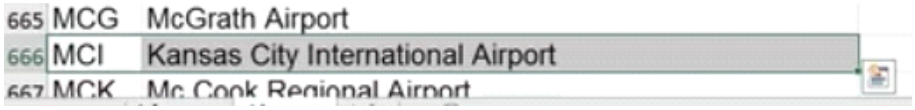


I have a list of airport codes. Suppose I want to know which airport is code MCI. The leftmost column is the key and the second column in this case is the answer.



	A	B	C
1	**Code**	**Airport Name**	
2	0AK	Pilot Station Airport	
3	16A	Nunapitchuk Airport	
4	1G4	Grand Canyon West Airport	
5	2A3	Larsen Bay Airport	
6	2A9	Kotlik Airport	
7	3A5	Marshall Don Hunter Sr	
8	3T7	Middle Bass Island Airport	
9	3W2	Put - in - Bay Airport	
10	6R7	Old Harbor Airport	
11	74S	Anacortes Airport	
12	A61	Tuntutuliak Airport	
13	A85	Kwigillingok Airport	
14	AAL	Aalborg Airport	
15	AAR	Aarhus Airport	
16	ABE	Lehigh Valley International Airport	
17	ABI	Abilene Regional Airport	
18	ABJ	Felix Houphouet-Boigny Airport	
19	ABQ	Albuquerque International Sunport	
20	ABR	Aberdeen Regional Airport	

Scrolling down, I find out MCI is Kansas City.



The function that does this is VLOOKUP.
There are four pieces of information that you will need in order to build the VLOOKUP syntax:

1. The value you want to look up, also called the lookup value.

2. The range where the lookup value is located. Remember that the lookup value should always be in the first column in the range for VLOOKUP to work correctly. For example, if your lookup value is in cell C2 then your range should start with C.

3. The column number in the range that contains the return value. For example, if you specify B2:D11 as the range, you should count B as the first column, C as the second, and so on.

4. Optionally, you can specify TRUE if you want an approximate match or FALSE if you want an exact match of the return value. If you don't specify anything, the default value will always be TRUE or approximate match.

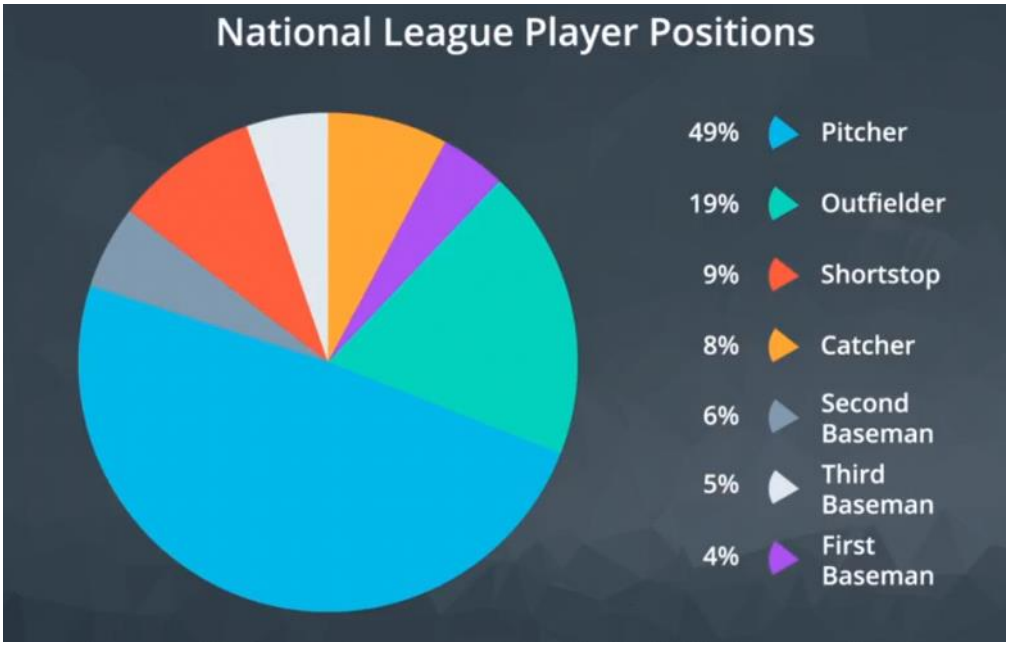
Let's try that. ZZZ is not in the list.

We get an error. We get an error because I put false in the formula to indicate I only want to find an exact match. What if I didn't put false in the formula? Excel would return the closest match which for this problem is just wrong. ***A rule of thumb is, that when looking up information with VLOOKUP, enter false in this optional parameter.*** Finding the wrong airport in this case is worse than getting an error.

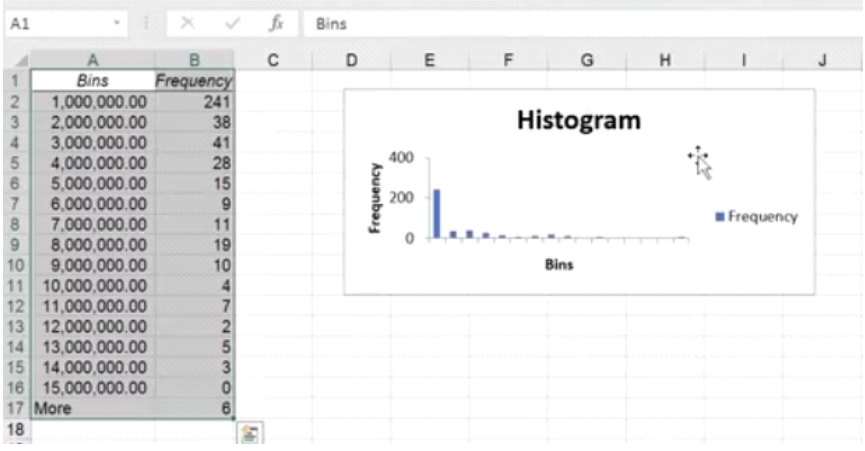
Okay, one more thing. Instead of using columns as my range, which did work just fine, it's really better practice to use a named range. It's more readable and less prone to errors.

Intro to Visualize Data

When we look at spreadsheets and tables full of text and numbers, it's hard to tell much. We might notice trends such as, gee, there seem to be a lot of pitchers in this list.
And we can create some formulas to find out answers to questions like, what percentage of the players are pitchers? Well, now I'm wondering about all the other positions, outfielders, catchers. It would be more helpful to see the data in one big picture.
Here's a pie chart of the percentages of all positions from the baseball set.



This is an example of visualizing the data. When we visualize our data, we understand it more fully. And if you're explaining your data analysis to someone else who's less familiar with it than you are, that is especially important.
When looking at the data as a picture, it's easier to notice patterns, perhaps, spot trends to examine more closely. In this lesson, you're going to learn how to make pie charts as well as bar charts, scatter plots and line graphs from your data.
After learning how to create some of these basic graphs, you'll learn how to create histograms, which display the frequency of values in your data, and box plots, which provide an added glance view of the statistical spread in your data.

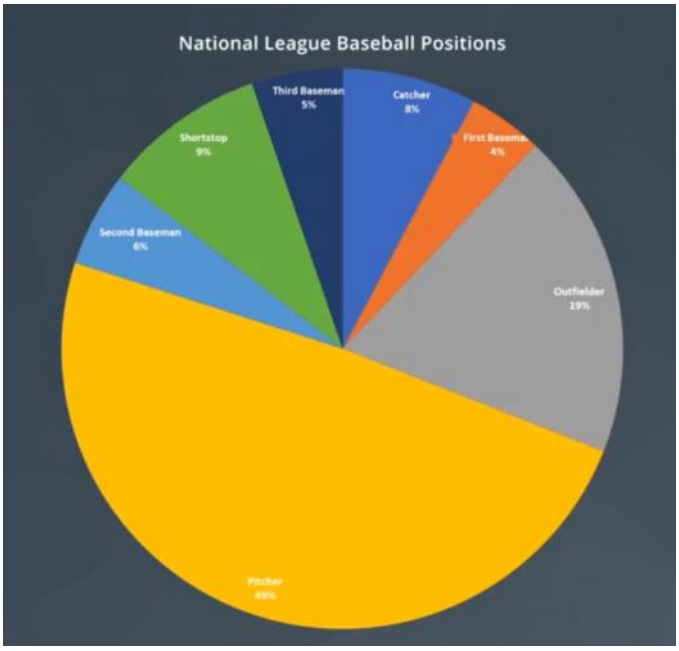


	Bins	Frequency
2	1,000,000.00	241
3	2,000,000.00	38
4	3,000,000.00	41
5	4,000,000.00	28
6	5,000,000.00	15
7	6,000,000.00	9
8	7,000,000.00	11
9	8,000,000.00	19
10	9,000,000.00	10
11	10,000,000.00	4
12	11,000,000.00	7
13	12,000,000.00	2
14	13,000,000.00	5
15	14,000,000.00	3
16	15,000,000.00	0
17	More	6

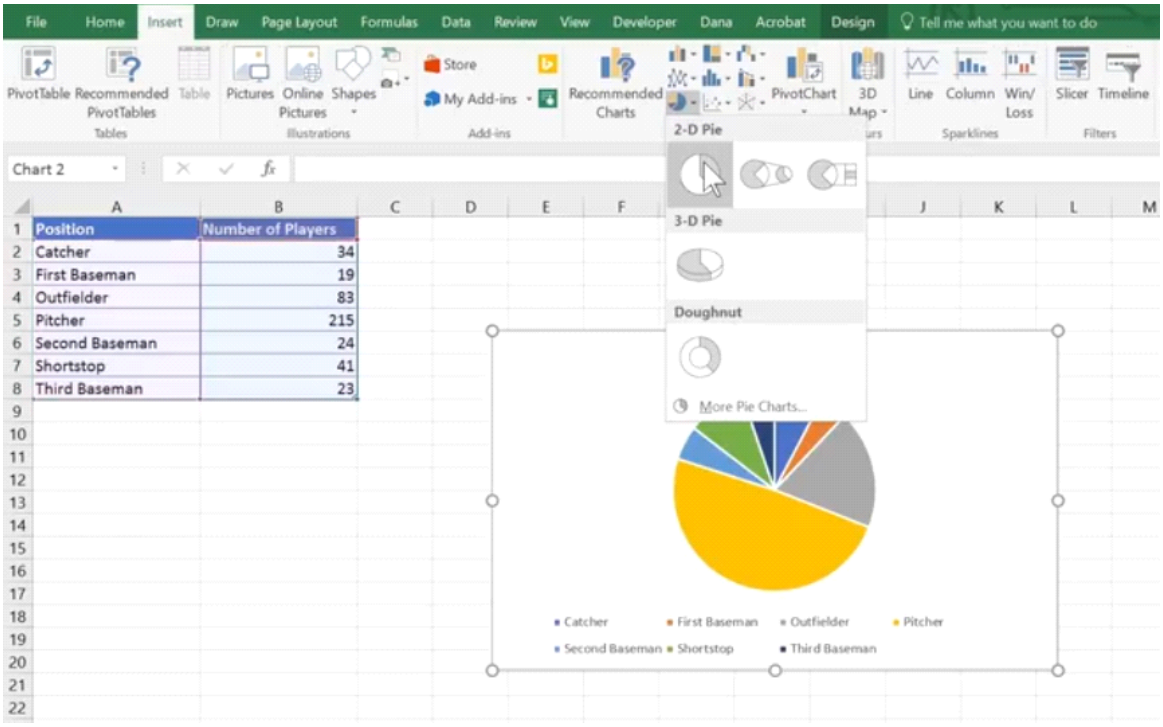
As we work through these various charts, you'll learn some formatting skills that make your charts clear and easy to understand. We'll finish the lesson with a little more formatting to really polish your skills for producing professional looking charts for data analysis reports.

Pie Charts

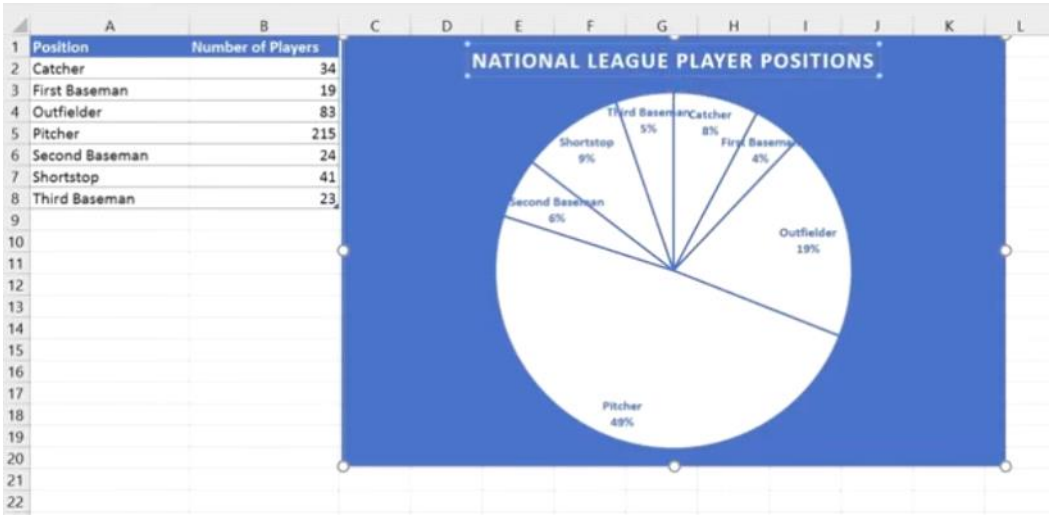
A pie chart is used to illustrate proportionality. If we have a whole group of some sort, such as a list of all players in a baseball league and their positions, a pie chart can tell us what percentages of the whole group are in each position.



Think of it as slicing the pie into pieces, where each piece matches a percentage of the whole list. In spreadsheets this is really easy, because all we need is a list of the categories and matching values such as sums or counts. Select the list of categories and values, go to the Insert menu and choose the pie chart. There it is. It might not look quite the way we want yet, but there are some easy fixes for that too.

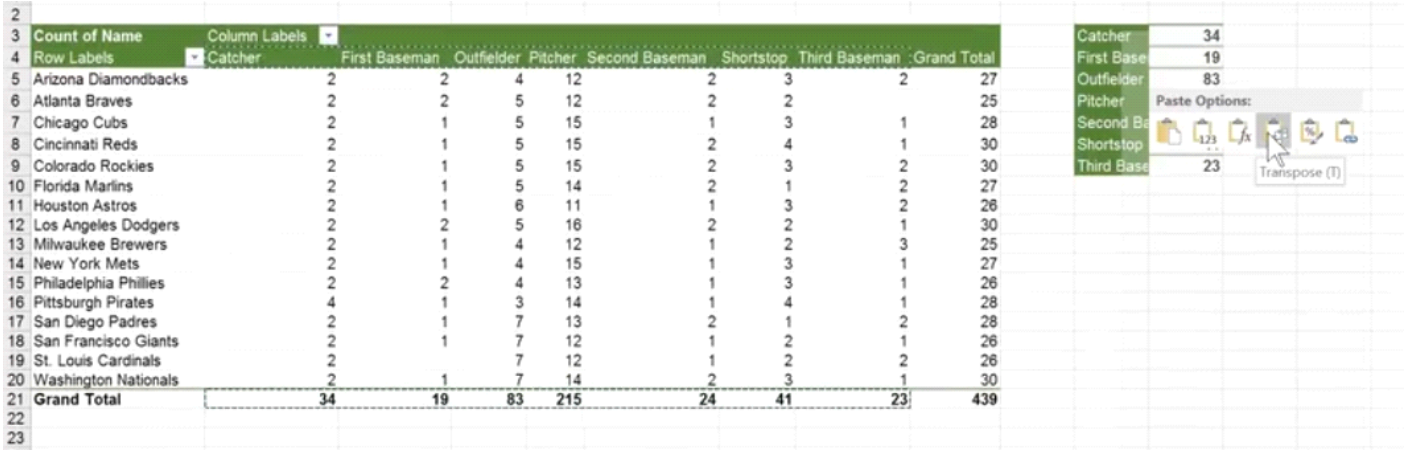


When the chart is selected, a design and format menu are available on the Excel ribbon.
The design menu gives a number of choices: I'll choose this one because I like seeing the percentages on the pie.
I can change the chart title to something more useful by editing it. There's more that you can do to change the appearance of the graph, but this gets it started. We'll go into more detail on how to polish the look later.



Position	Number of Players
Catcher	34
First Baseman	19
Outfielder	83
Pitcher	215
Second Baseman	24
Shortstop	41
Third Baseman	23

For this example I already had a simple list that I highlighted, but I could have used the pivot table we created earlier with some careful selection. Here's the pivot table.
The position categories are in the top row and the totals are in the bottom row. Select the categories, then either hold down the control key on Windows, or the Command key on Apple keyboards to select the bottom row with your mouse.

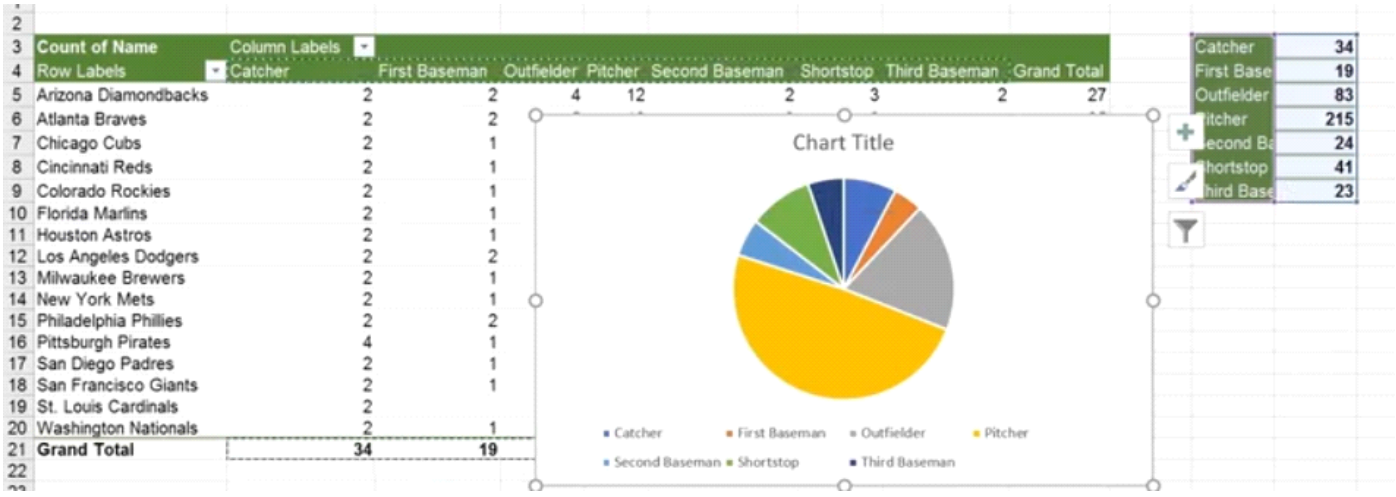


Count of Name	Column Labels							
Row Labels	Catcher	First Baseman	Outfielder	Pitcher	Second Baseman	Shortstop	Third Baseman	Grand Total
Arizona Diamondbacks	2	2	4	12	2	3	2	27
Atlanta Braves	2	2	5	12	2	2		25
Chicago Cubs	2	1	5	15	1	3	1	28
Cincinnati Reds	2	1	5	15	2	4	1	30
Colorado Rockies	2	1	5	15	2	3	2	30
Florida Marlins	2	1	5	14	2	1	2	27
Houston Astros	2	1	6	11	1	3	2	26
Los Angeles Dodgers	2	2	5	16	2	2	1	30
Milwaukee Brewers	2	1	4	12	1	2	3	25
New York Mets	2	1	4	15	1	3	1	27
Philadelphia Phillies	2	2	4	13	1	3	1	26
Pittsburgh Pirates	4	1	3	14	1	4	1	28
San Diego Padres	2	1	7	13	2	1	2	28
San Francisco Giants	2	1	7	12	1	2	1	26
St. Louis Cardinals	2		7	12	1	2	2	26
Washington Nationals	2	1	7	14	2	3	1	30
Grand Total	**34**	**19**	**83**	**215**	**24**	**41**	**23**	**439**

Catcher	34
First Base	19
Outfielder	83
Pitcher	
Second Ba	
Shortstop	
Third Base	23

Paste Options:
Transpose (T)

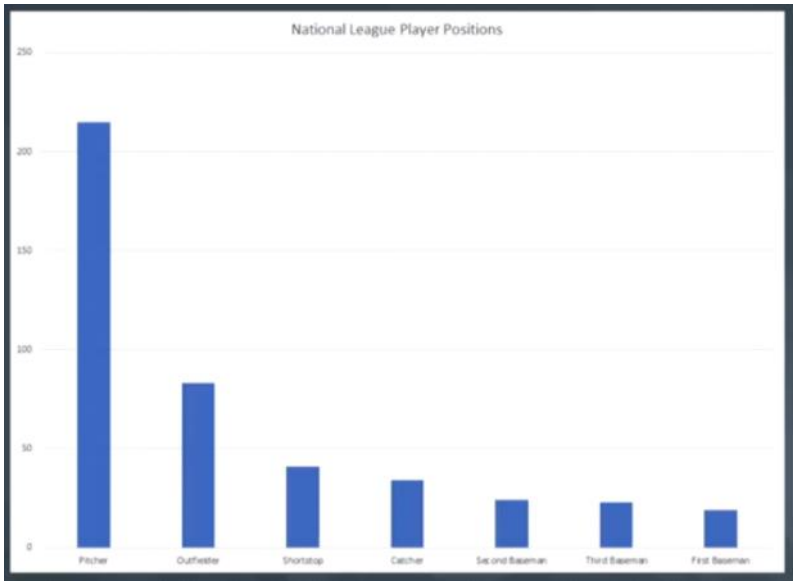
It works most reliably to copy these and then paste this little collection somewhere else. I'm going to paste using the transpose feature. So I have columns, then select and choose the insert pie chart as before. We get the exact same thing.



Count of Name	Column Labels							
Row Labels	Catcher	First Baseman	Outfielder	Pitcher	Second Baseman	Shortstop	Third Baseman	Grand Total
Arizona Diamondbacks	2	2	4	12	2	3	2	27
Atlanta Braves	2	2						
Chicago Cubs	2	1						
Cincinnati Reds	2	1						
Colorado Rockies	2	1						
Florida Marlins	2	1						
Houston Astros	2	1						
Los Angeles Dodgers	2	2						
Milwaukee Brewers	2	1						
New York Mets	2	1						
Philadelphia Phillies	2	2						
Pittsburgh Pirates	4	1						
San Diego Padres	2	1						
San Francisco Giants	2	1						
St. Louis Cardinals	2							
Washington Nationals	2	1						
Grand Total	**34**	**19**						

Catcher	34
First Base	19
Outfielder	83
itcher	215
econd Ba	24
hortstop	41
hird Base	23

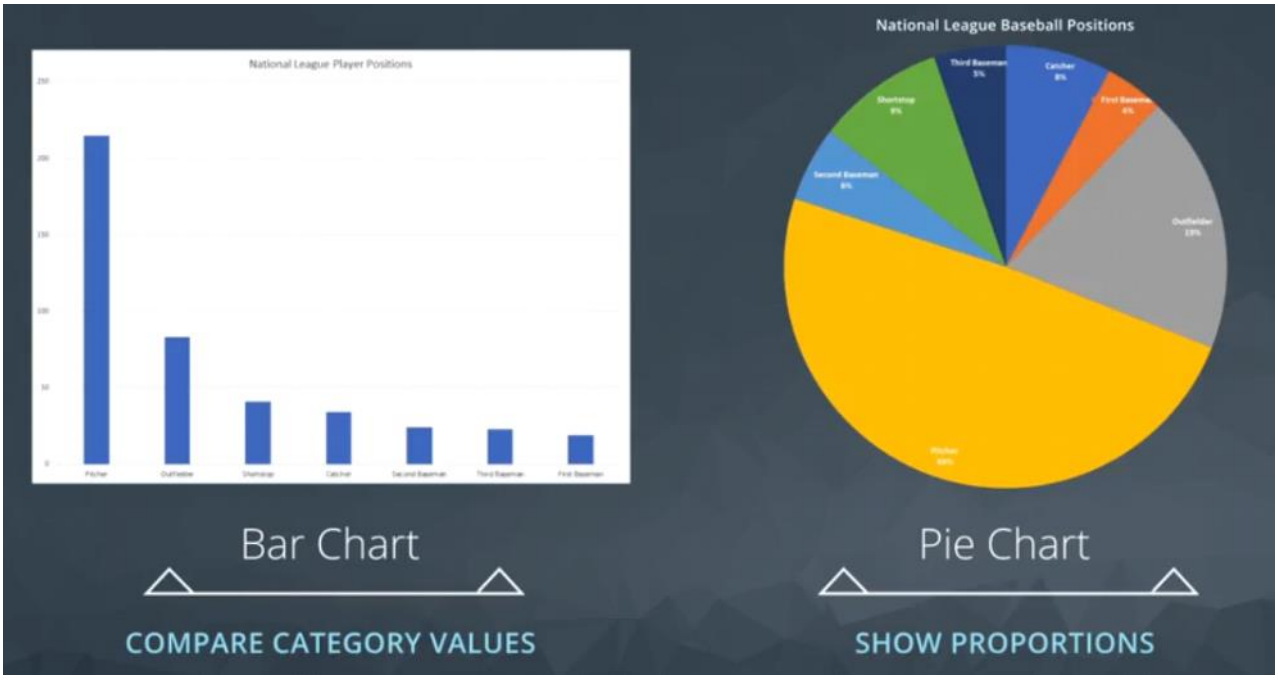
Bar Charts



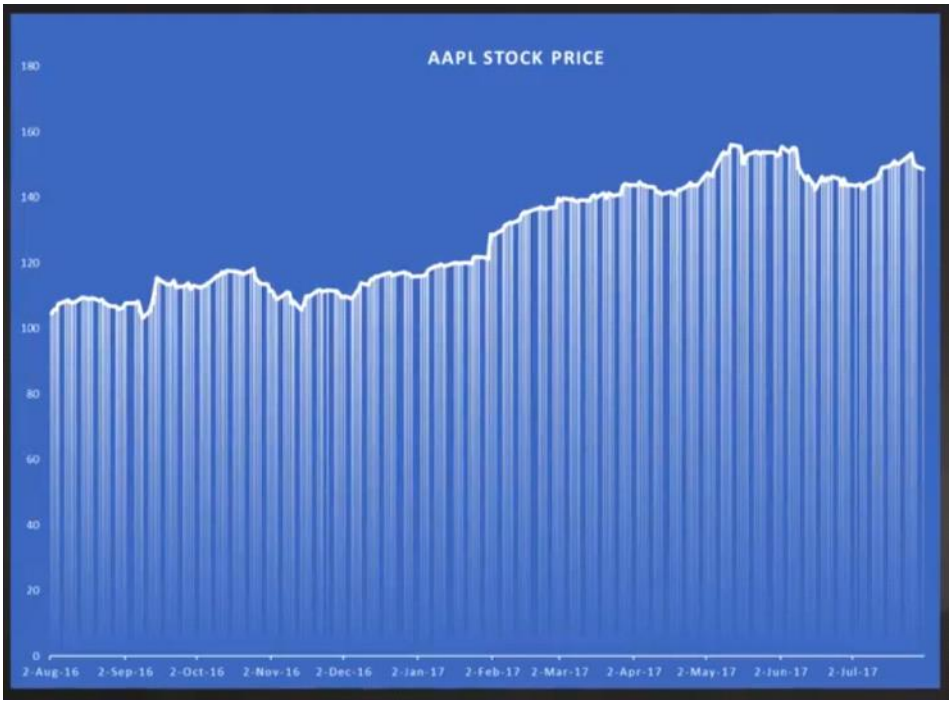
We could use the same information as before and choose a bar or column chart instead of a pie chart. Instead of percentages, it would just show the values with longer bars or columns representing larger values.

This puts a little different spin on things. In this chart, we're comparing the category values against each other and we see their relative sizes. However, we do not have much sense of the whole league or the percentage of each category as we did with the pie charts. **Choosing which kind of chart to use really depends on what patterns you want to highlight and what questions you want to answer.**

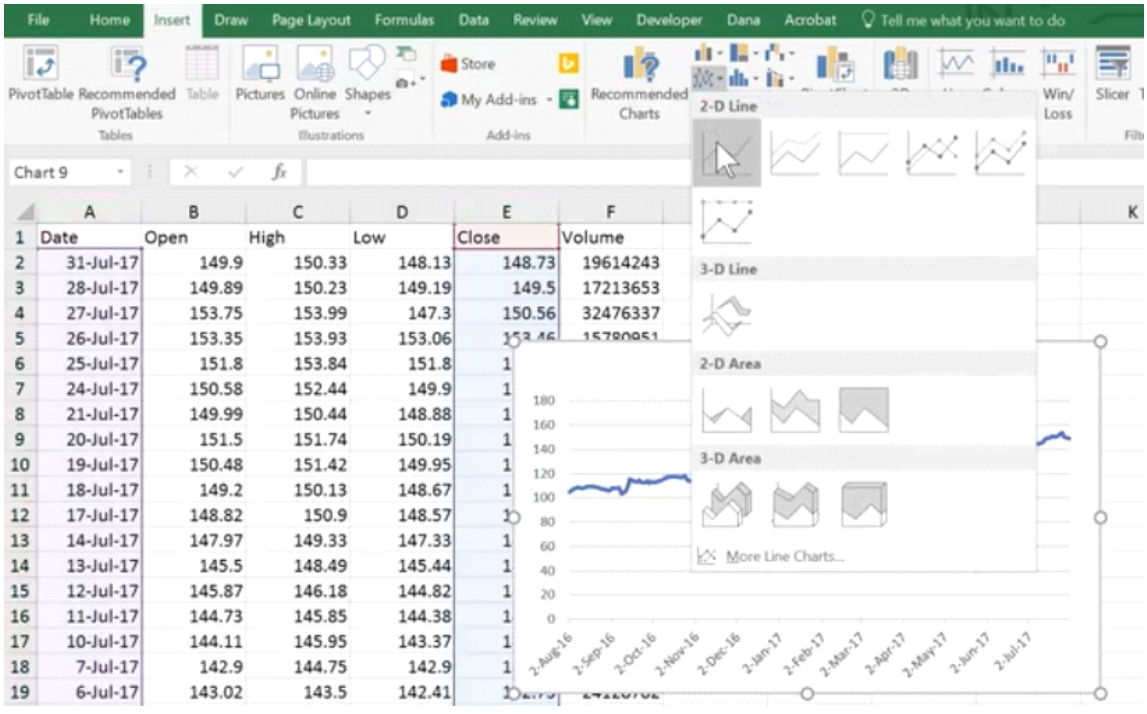
Use bar or column charts to compare category values with each other. Those categories could be player positions as shown here or could be sales years or could be apples and oranges. **Use a pie chart to show proportionality of categories.**



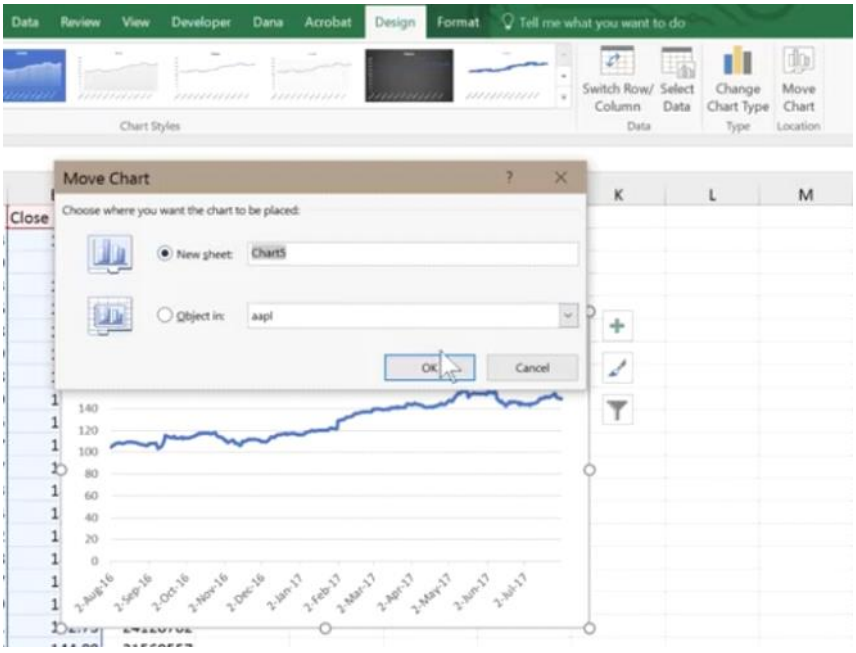
Scatter and Line Plots



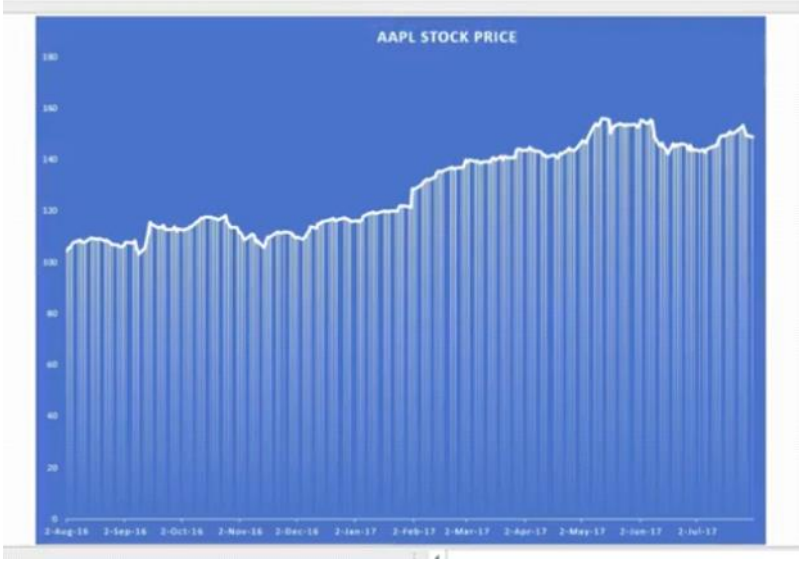
We use pie and bar charts to visualize categorical data. If we have a list of numerical data, such as the list of stock prices over time, a line chart gives us a better picture of the data set. Here's a table of data I downloaded from a financial website, listing prices for AAPL stock.
It has columns for date, open, high, low, close, and volume. I want to select the date column and the close column then go to the insert menu and select a line chart.



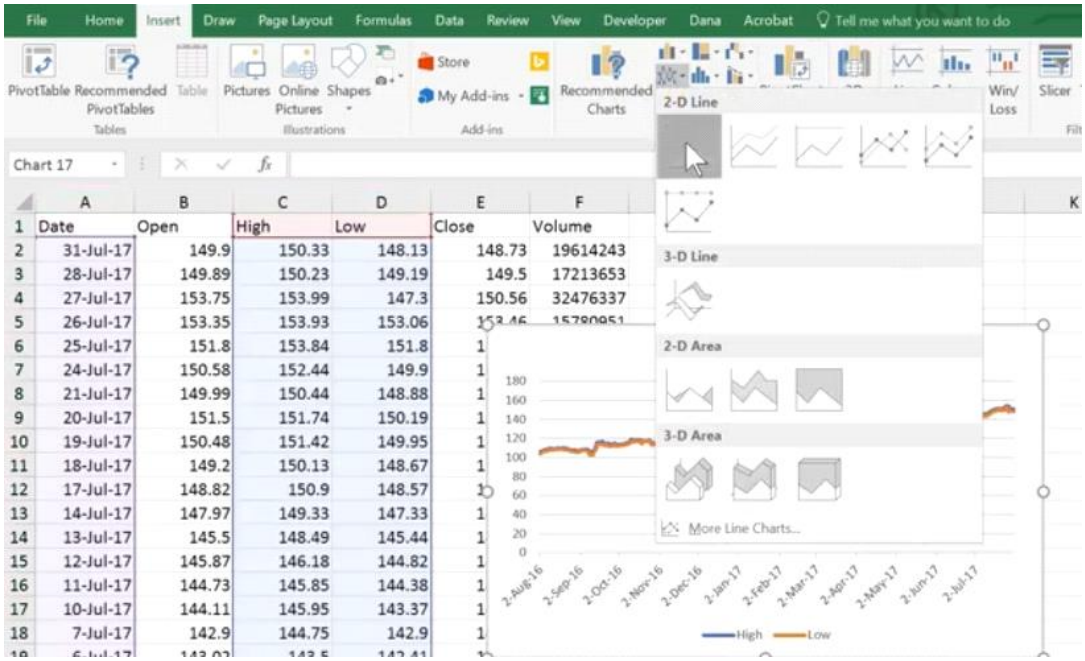
I can move the chart to its own sheet to see the detail better.



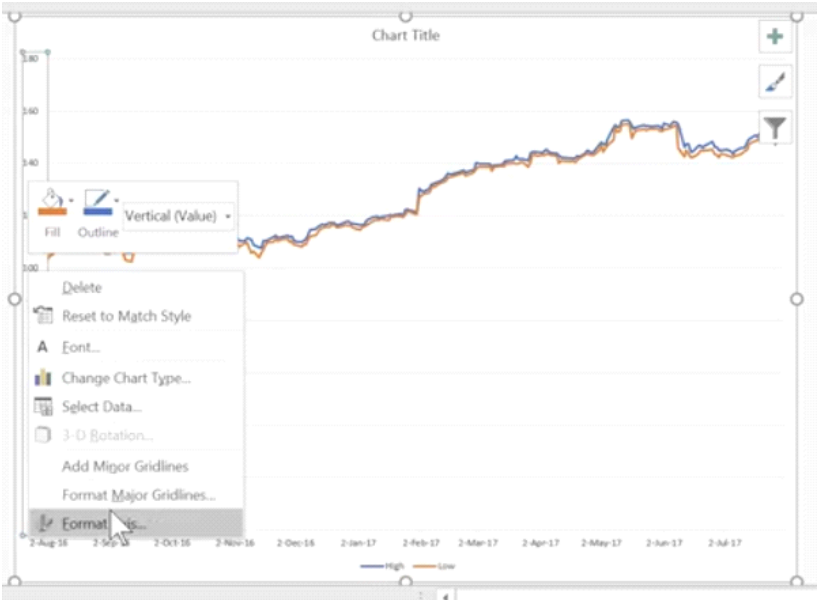
The chart is a little plain, but there are several quick fixes in the chart design menu. I'll choose one and change the title to AAPL Stock Price. That was pretty easy. The horizontal axis shows the dates, and the vertical axis shows dollar values.



Now I can see, at a glance, that over the past year the stock has gone up with a little hiccup about a month ago.
If I have more than one column of data for the same dates, I can show lines for each on the same chart. I'll use the same data set, but I'll select the date plus the high and low values for AAPL stock.

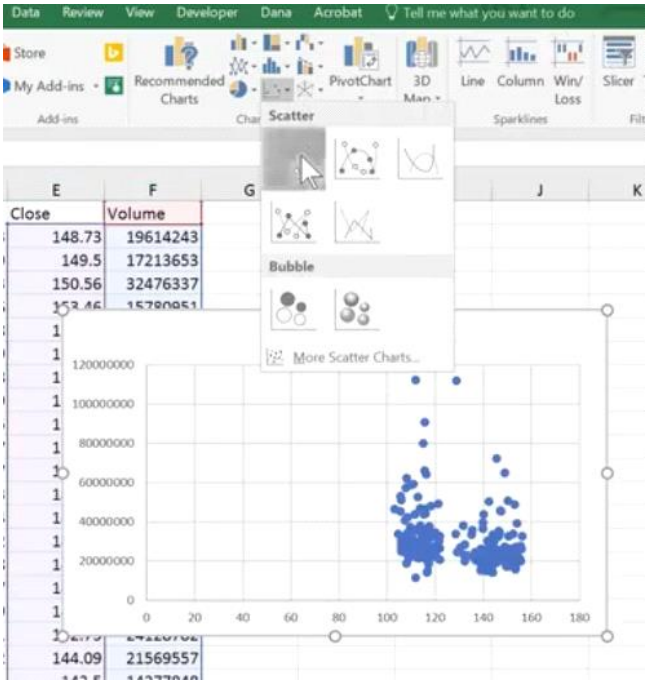


Since the high and low aren't all that far apart, I'll also change the range for the dollar amount on the left to start at 100.

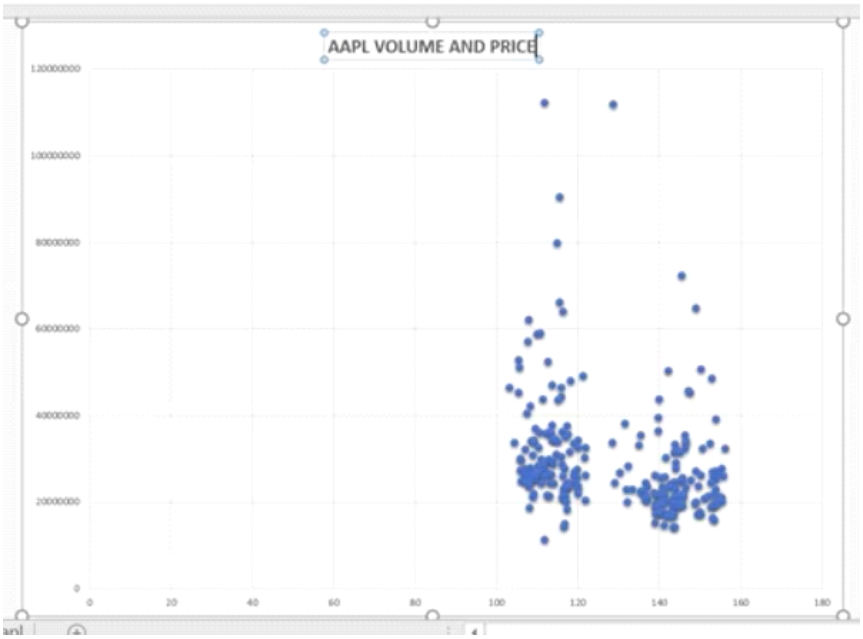


Select the vertical axis, right click, and format axis. You can see both the high and low value lines now and see the spreads between them.

There are a lot of ways to change the look of the same line chart by altering the layout. I could also try to plot two different variables for AAPL stock without dates at all, such as the closing price and volume.



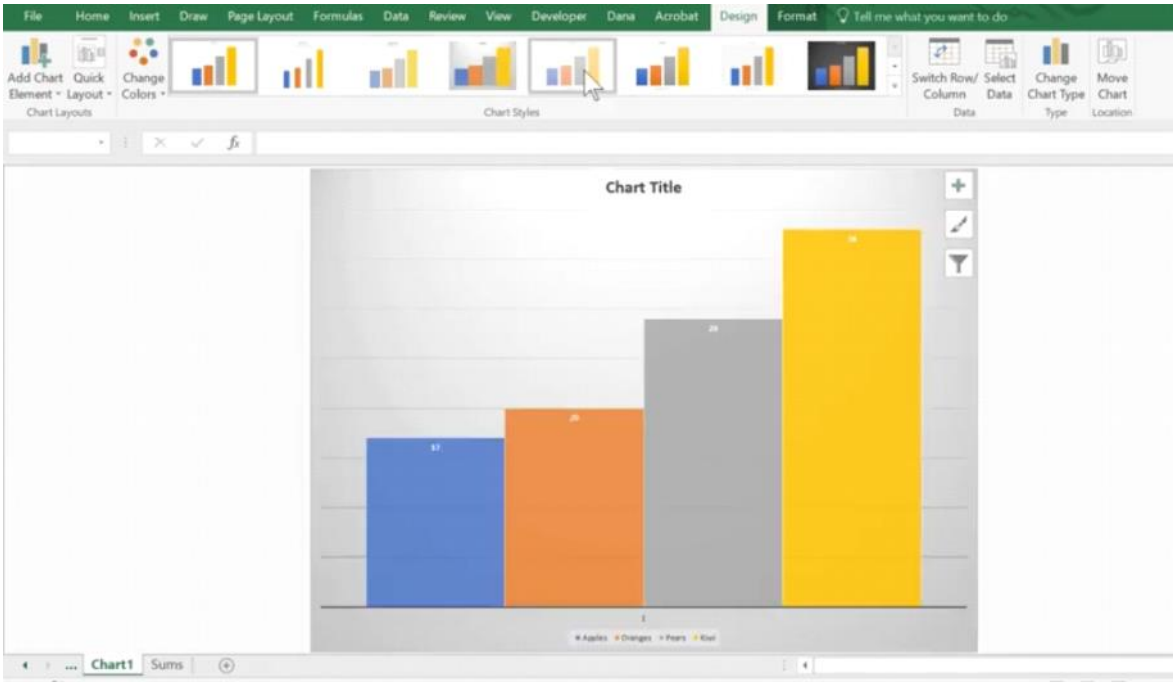
When I choose the scatterplot, I get a graph with the closing price on the horizontal axis and the volume of trade that day on the vertical axis.



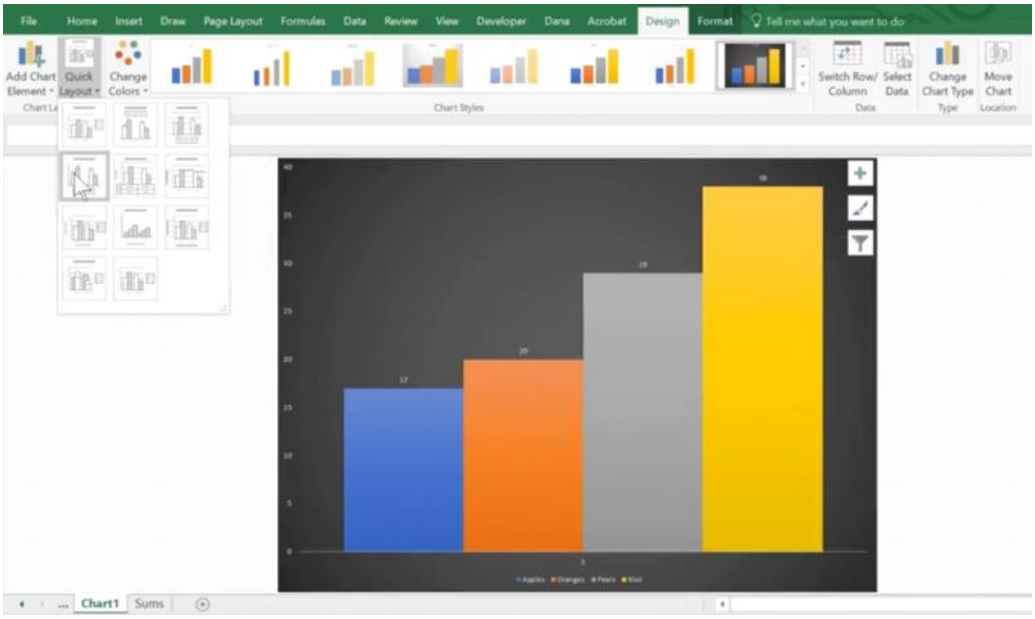
I can see that the prices seem to cluster in a couple of areas and that they have about the same volume generally though there are some high volume days at the lower price.

Chart Layout Tools

Before we go on to the next type of chart, let's take a look at the chart layout options in Excel.
So far, we've used the Chart Design tab that appears when you select a chart to choose some pre-made formatted styles.



Now, I want you to take a look at the Quick Layout menu over here, and the Add Chart Element menu next to it.



The Quick Layout menu is a set of pre-made layouts that affect which chart elements are included and where. Like the axis labels and the legend. You can select various ones and see the effect.
To get really specific about your layout, look at the Add Chart Element menu. With this menu, it's possible to add and remove elements of all types from the chart.

There is a lot here, and rather than try to learn all of it, we'll just use some parts as we need them. It's good to know where to find it though, for those times, when you know what you want to change and all those pre-made formats and layouts just don't have what you need.

Quiz: Chart Layout

Spreadsheet charts have standard elements that can be added or removed by selecting a chart using the **Design** menu, then **Quick Layout** and **Add Chart Element** sub-menus. The chart below has several of these elements. Can you identify them by name?

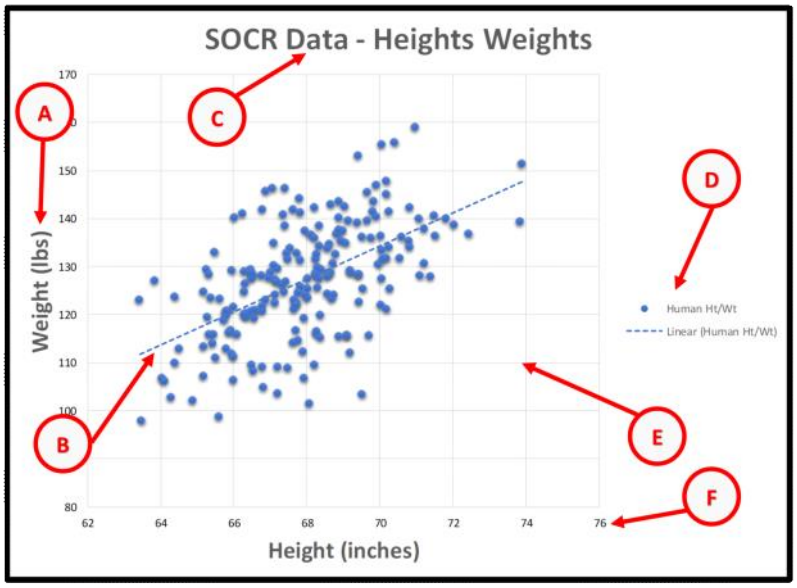


Chart element	Location
Chart Title	C
Horizontal Axis	F
Vertical Axis	None
Trend line	B
Legend	D
Vertical Axis Title	A
Grid Lines	E

Histograms

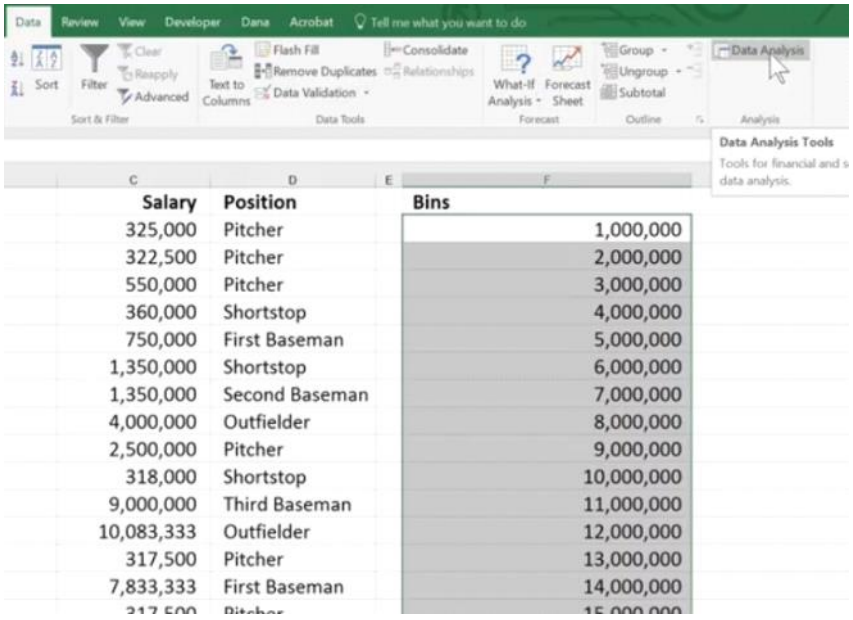
A Histogram is a column chart that measures the frequency of data in a data set and specifically groups numerical values into bins, we define.
Recall that we previously created a column chart to compare counts of categories within a data set. This kind of chart answers a question like, how many players are there in each playing position in the league?

But what if we want to ask the question, how many players made under $1 million in salary, and between $1 and $2 million, and between $2 and $3 million in salary, etc. This kind of a chart is a histogram, and the groupings we choose such as, all salaries between $1 and $2 million and salaries between $2 and $3 million are the **bins**.

There are two ways to do this in Excel.

- We'll start with a method that works on both Windows and Mac using the histogram tool in the analysis tool pack add-in. Instructions for loading the analysis tool pack add-in are given in the Getting Started instructions.
- The second way is a simple insert chart method available for Excel 2016 for PC.

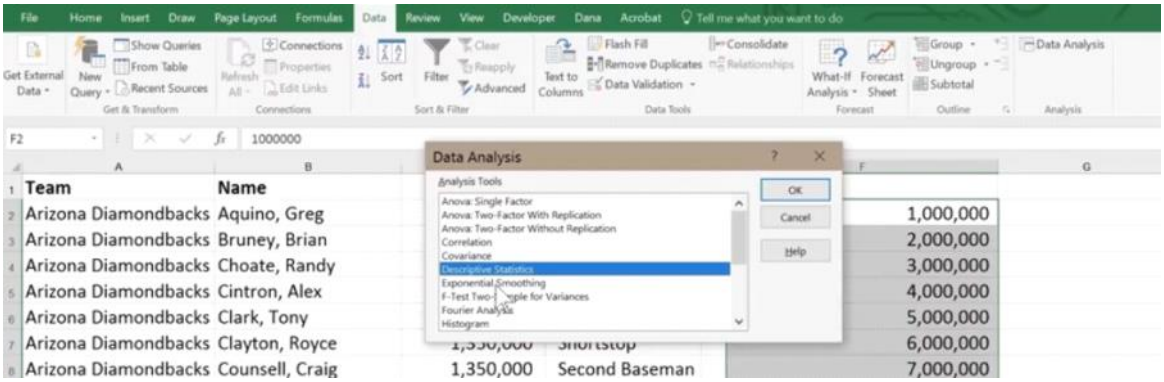
I'll show you that too. The tool pack histogram requires two columns of data. One for data you want to analyze and one for bin levels that represent the intervals for the bins. In this example, I'm starting at $1 million, then $2 million, et cetera, up to $15 million.



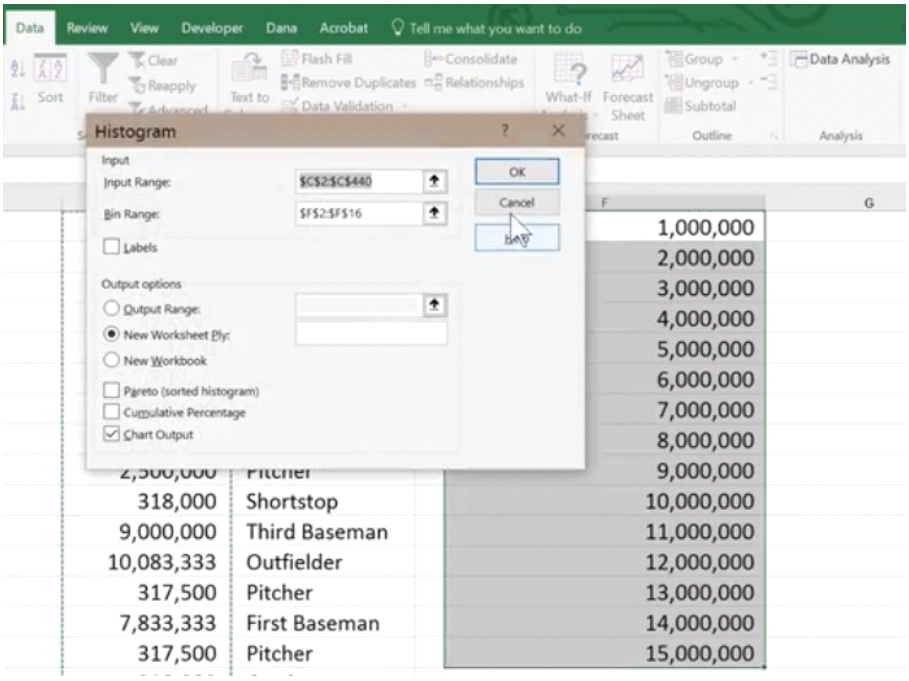
Salary	Position	Bins
325,000	Pitcher	1,000,000
322,500	Pitcher	2,000,000
550,000	Pitcher	3,000,000
360,000	Shortstop	4,000,000
750,000	First Baseman	5,000,000
1,350,000	Shortstop	6,000,000
1,350,000	Second Baseman	7,000,000
4,000,000	Outfielder	8,000,000
2,500,000	Pitcher	9,000,000
318,000	Shortstop	10,000,000
9,000,000	Third Baseman	11,000,000
10,083,333	Outfielder	12,000,000
317,500	Pitcher	13,000,000
7,833,333	First Baseman	14,000,000

When I create the histogram, the number of values in the salaries lists that are below $1 million will be in

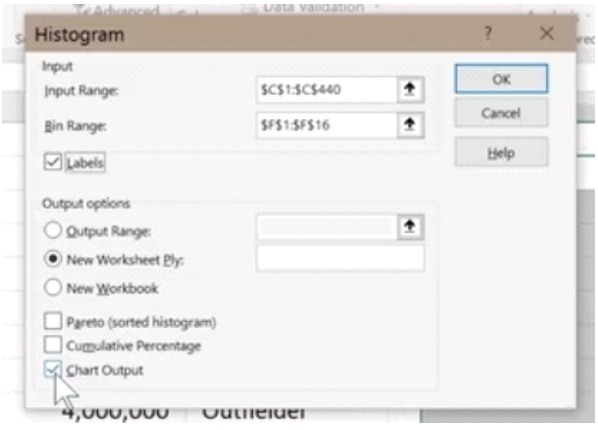
the first bin. The numbers of salaries between $1 and $2 million will be in the second bin, and so on. To create the histogram, choose data analysis from the data menu on Windows or from the tools menu on Mac.



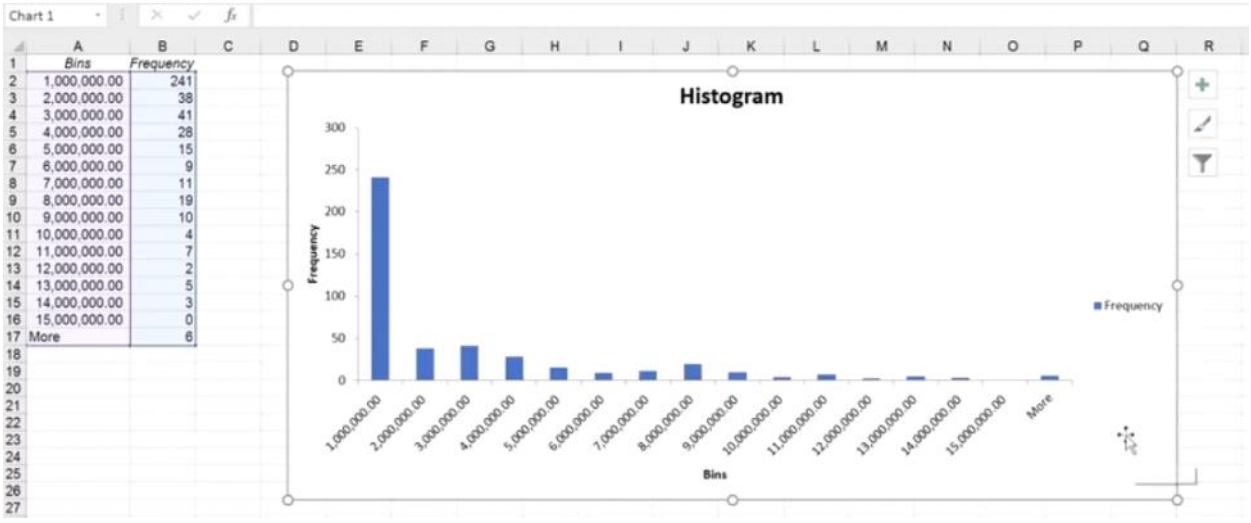
Choose histogram, which opens a dialog. For the input range, select the data from the salaries column. For the bin range, select the bin intervals you've created.



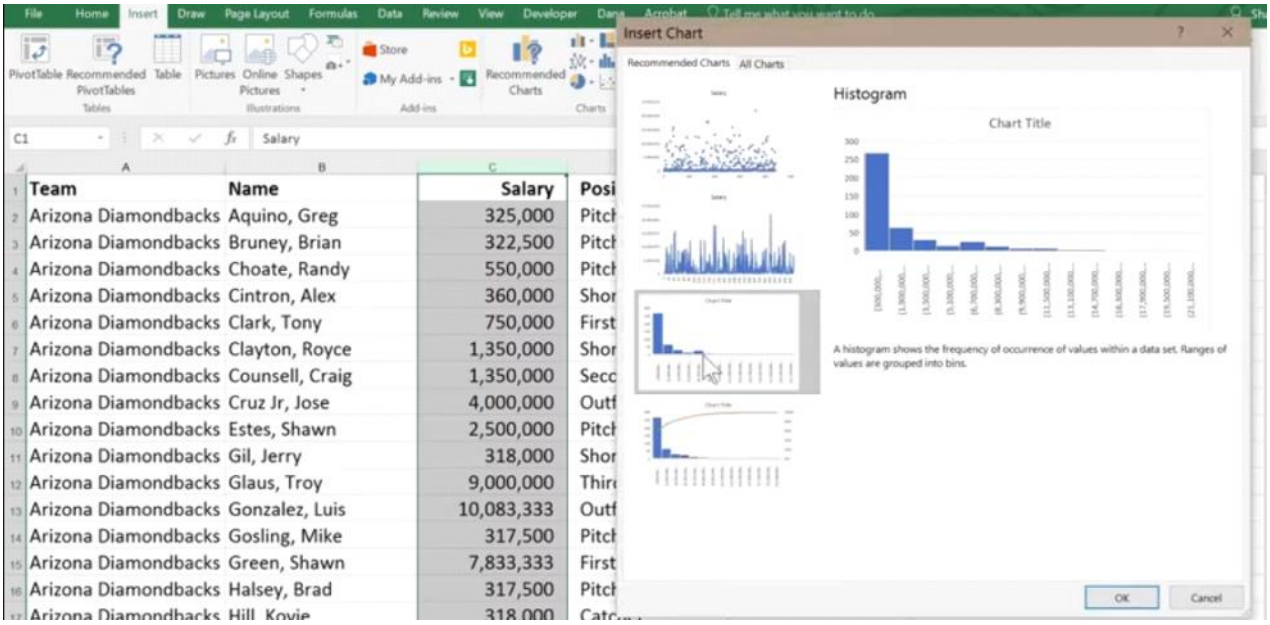
If you have a label at the top of your columns as I do here, click labels. For the output options, select new worksheet and chart output, and press OK.



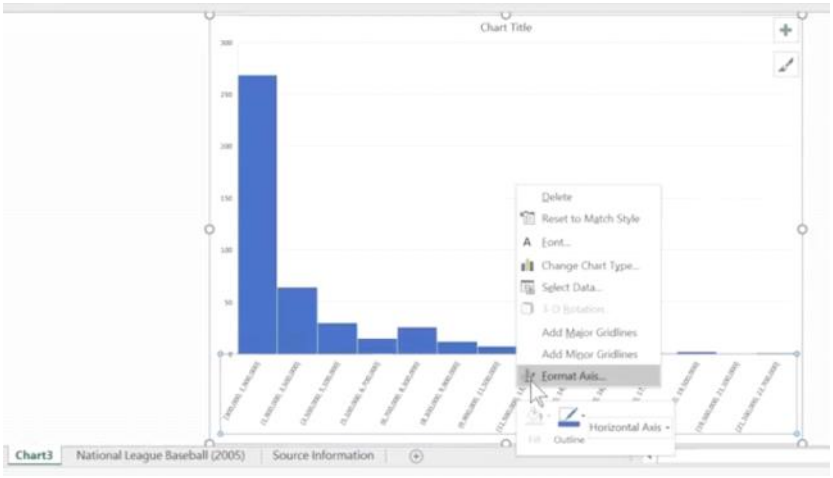
Excel will analyze the list and create a count for the bins on a new worksheet, and then create a column chart to match it.



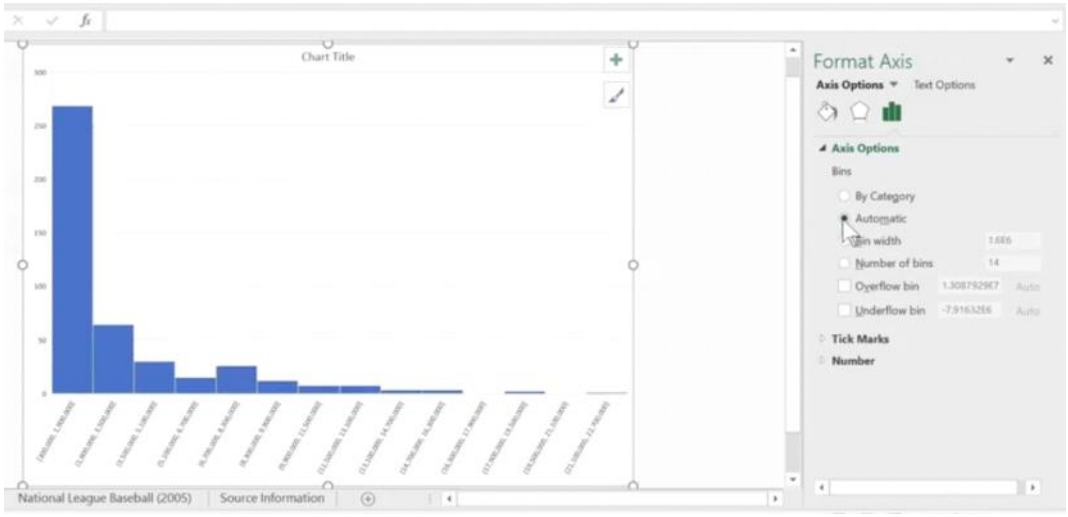
The second method is only available in Excel 2016 on Windows. Just select your data and click insert, recommended charts and choose the histogram chart.



To configure details about the bins, right click the horizontal axis of the chart, click format axis and then click axis options.

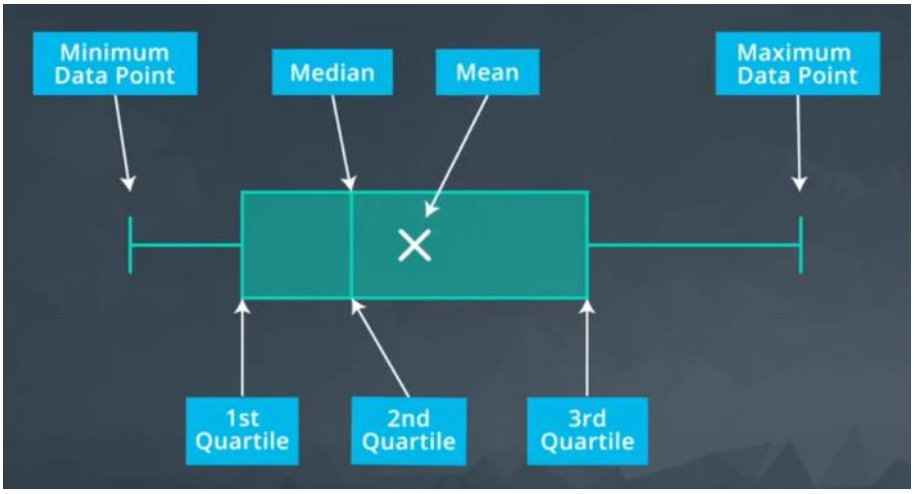


The dialog provides options for choosing categorical data like the player positions or automatic for numerical data. You can specify the number of bins which you might choose to experiment with a bit.
If you choose bins that are too narrow, the result can be noisy.
On the other hand, too few bins will hide details. Here as always, think about what you're trying to learn and convey about the data, and use that as your guide for making these determinations.
As with other charts, design and layout can be further customized from the **design menu** when the chart is selected.

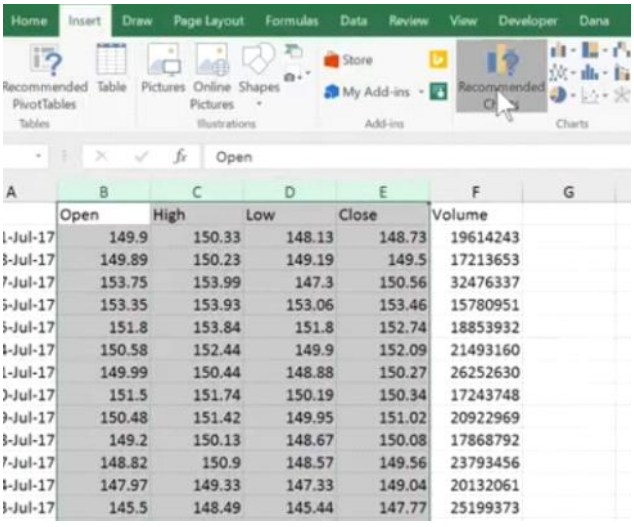


Box Plots

A box plot, which in our case is really a box and whisker plot, is the visualization of statistical spread in a data set of values. A traditional box plot is built using the five numbers summary.
The five numbers summary consists of five values. The maximum, the minimum, the 1st quartile, the 2nd quartile, also called the median, and the 3rd quartile.
If we have a list of numbers sorted and divided as evenly as possible into fourths, we have the quartiles of data. Where we make a box and whisker plot, the maximum becomes the tip of the upper whisker and the minimum becomes the tip of the lower whisker. The box represents the middle half of the data with a line where the median is.
Excel will give us a bonus six number in the summary by placing an X at the mean or average value of the set.

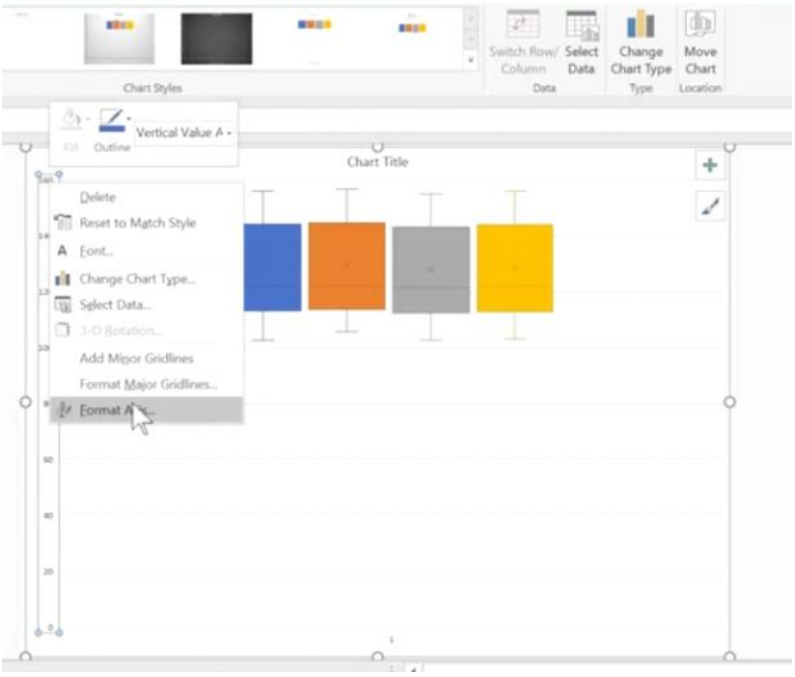


Creating a box plot in Windows Excel 2016 is as easy as any other chart.
Select columns of data.

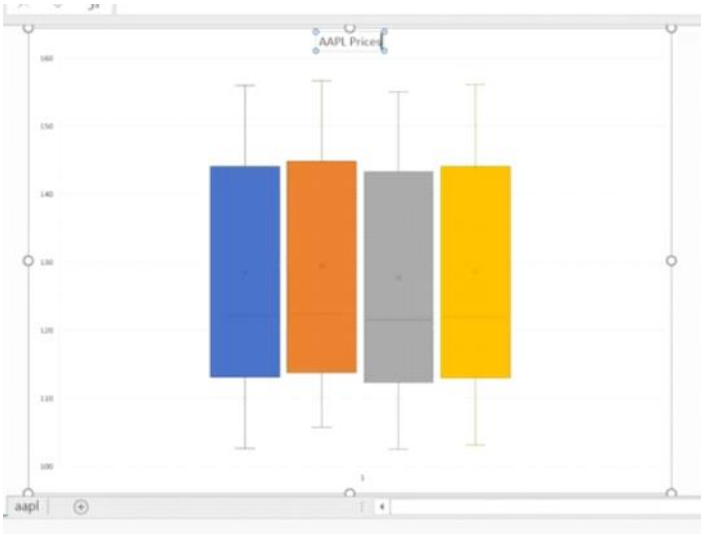


	Open	High	Low	Close	Volume
1-Jul-17	149.9	150.33	148.13	148.73	19614243
3-Jul-17	149.89	150.23	149.19	149.5	17213653
7-Jul-17	153.75	153.99	147.3	150.56	32476337
5-Jul-17	153.35	153.93	153.06	153.46	15780951
6-Jul-17	151.8	153.84	151.8	152.74	18853932
4-Jul-17	150.58	152.44	149.9	152.09	21493160
1-Jul-17	149.99	150.44	148.88	150.27	26252630
0-Jul-17	151.5	151.74	150.19	150.34	17243748
9-Jul-17	150.48	151.42	149.95	151.02	20922969
8-Jul-17	149.2	150.13	148.67	150.08	17868792
7-Jul-17	148.82	150.9	148.57	149.56	23793456
4-Jul-17	147.97	149.33	147.33	149.04	20132061
3-Jul-17	145.5	148.49	145.44	147.77	25199373

Click insert in recommended charts then click the box & whisker chart. Remember that a box plot represents statistics for a single list of numbers.

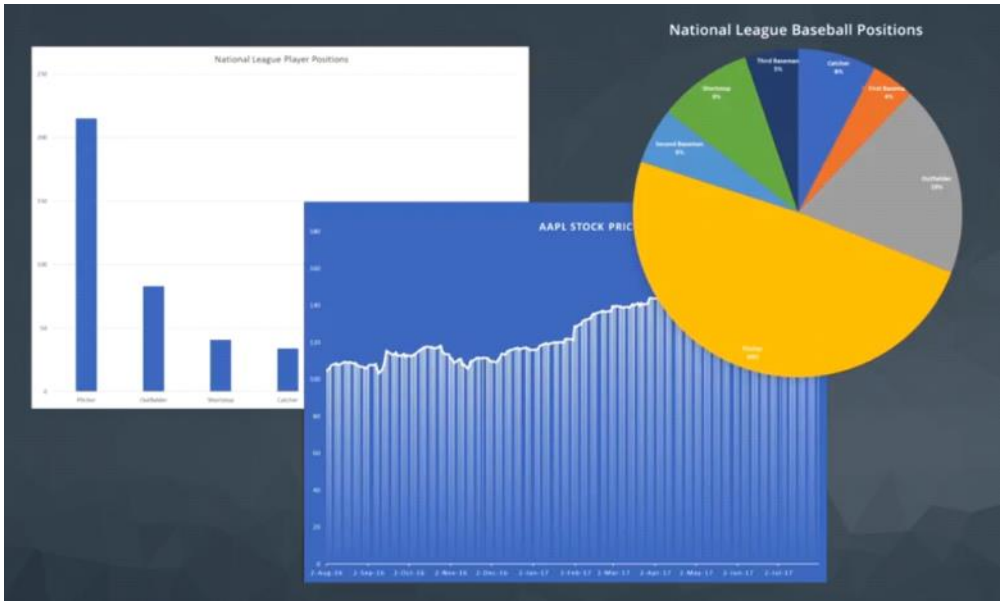


So, each list you select will be represented by its own box plot. The box plot now visually gives a sense of the spread of the value list. I'll just adjust the range so that I can see the plots a little better. And, give the chart a title.

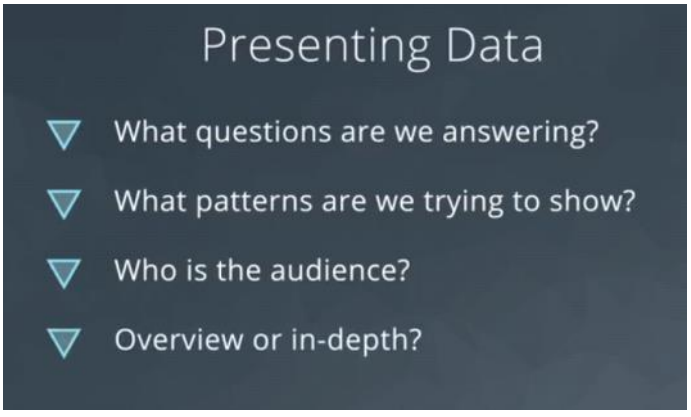


Professional Presentations

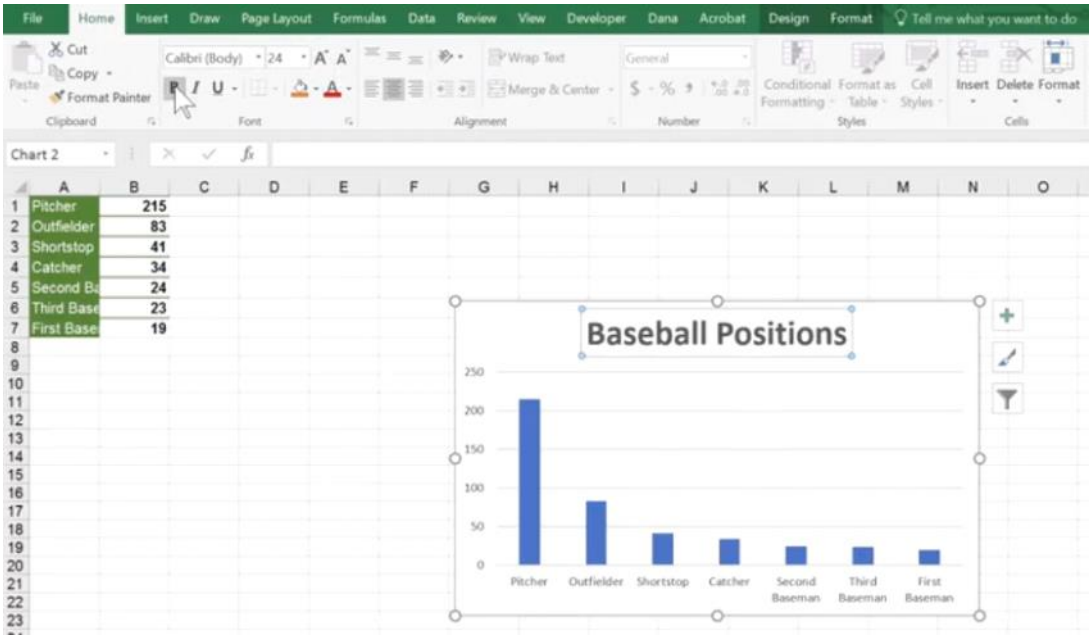
Excel and other spreadsheet applications can do a nice job of creating attractive tables and charts.
It's up to us to make them look their absolute best though. They should be readable, interesting and show the required information, but not show extra information that doesn't matter.



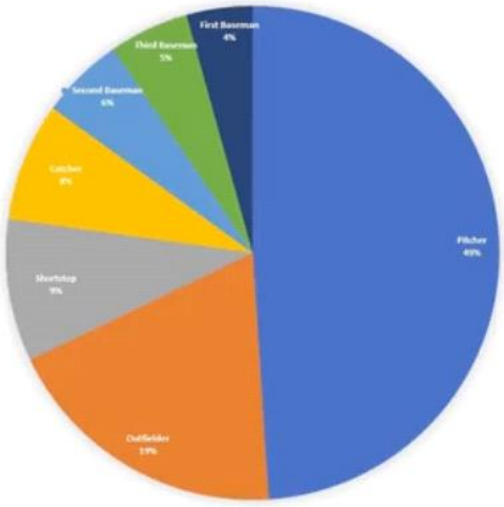
When you present data, think about what questions you're trying to answer and the patterns you're trying to show, then use fonts and layouts appropriately to include and emphasize the elements that matter and exclude elements that don't. Also, ask yourself who is your audience. Is this a quick overview presentation on a slide with the data backup elsewhere or is it a written technical review where more in-depth data should be presented?



Let's look at some of the charts we created earlier and see if we can improve upon them. For categorical column data, like the counts of the baseball positions, the relative sizes are easier to see at a glance if the data is sorted by size. This would not be something we would want to do for sequential data like daily stock values or histogram bins, but it is more readable in this case.
There should be a descriptive title, so I'll put one in that says baseball positions. It should also be big enough to read. If you need to change the font or font size, this can be done by selecting the title and changing it in the Home tab.

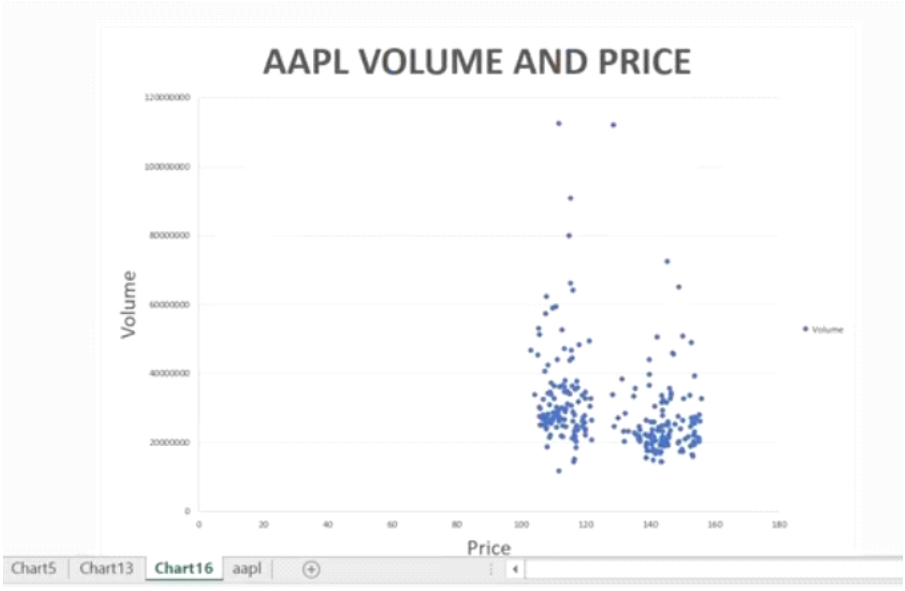


There, does this chart need a legend? If there are more than one data series, it does, but if the labels are already showing up in the data as in this pie chart, then an extra legend is just noise, so remove it.



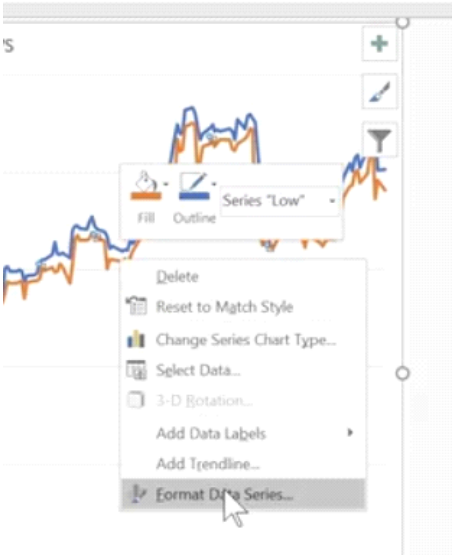
For a graph, do you need or want grid lines? In a detailed technical presentation, it is easier to see the values if there are grid lines, but for quick overviews that are just emphasizing the trend, those extra lines look busy.

What about the axis labels? I hate it when I look at a graph and can't tell what the numbers on the axis represent, so I would generally say put them there, but there are always exceptions.
If the units are redundant, somehow, they can be removed.
The font size is a judgement call, but bigger is easier to read and draws attention.

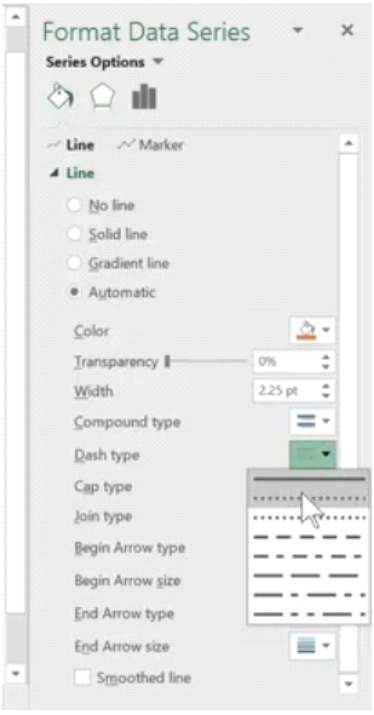


So take a look at your chart and see if it can be improved. Color is a similar issue. It just depends.

Keep in mind that your chart may be printed in grayscale or viewed by someone who's colorblind.



So if it's possible to distinguish groupings in additional ways such as different shapes and scatter plots, or using dashes in lines, it's a best practice to do so.

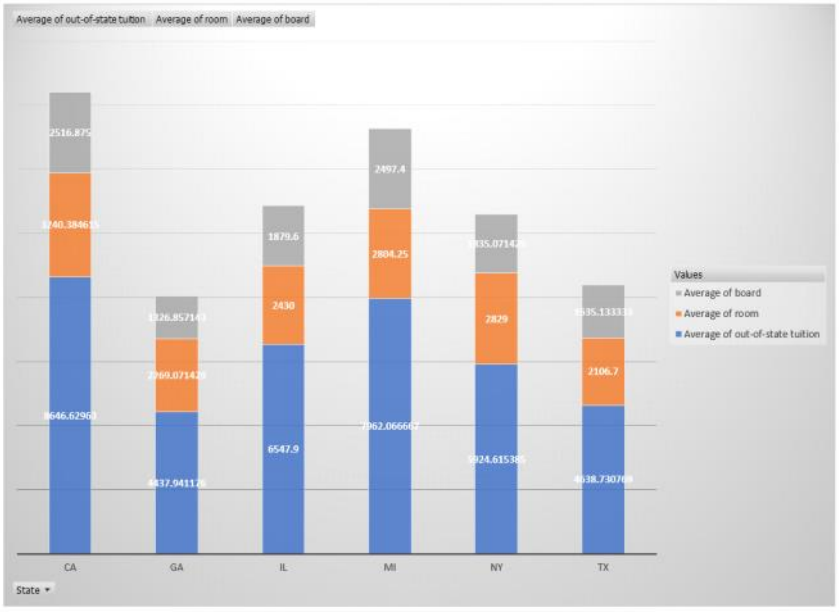


The format data series dialogue has a number of options for changing the look of your graph.

This was just a sampling of some of the modifications you can make to your charts to give them the impact you want in a professional presentation. The best way to learn is to just try things, to explore.

Exercise: Professional Presentations

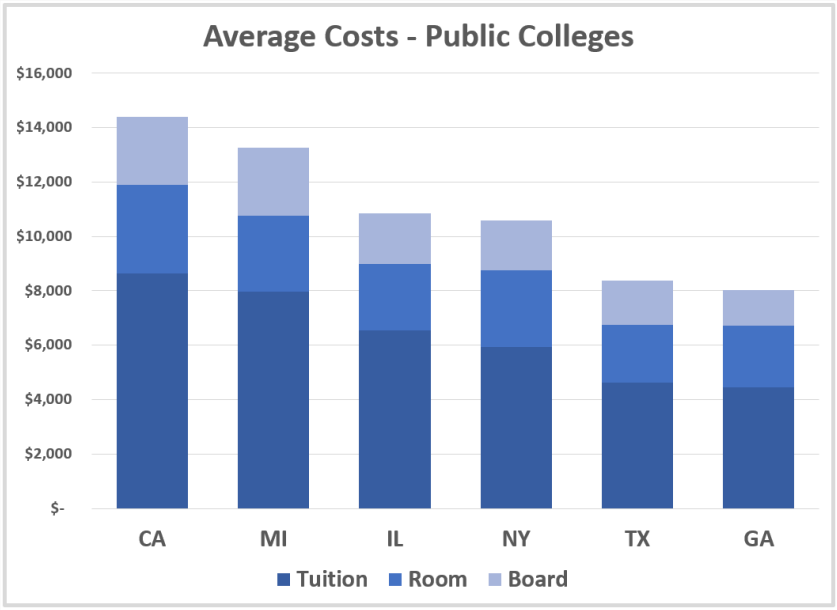
Spreadsheets 4: Visualize Data - Exercise: Professional Presentations



Quiz Question

The chart above is a stacked column chart with all the info requested, but it doesn't look quite right. Which of the following changes could make it a better chart for the counselor's overview presentation?

- [ ] Change the background color to something bright and cheery like hot pink
- [x] Change the background color to something that makes the graph easy to read like white
- [x] Change the numbers to dollars with no decimals
- [x] Reorder the categories with the highest total on the left to the lowest total on the right
- [x] Hide all Field Buttons (**PivotChart Tools->Analyze->Field Buttons**)
- [x] Add a Title
- [x] Remove the individual values in the stacked columns, but add a labeled axis on the left
- [x] Make all the labels and number bigger so the audience can see them



Recap

Congratulations, you've completed the final lesson on spreadsheets. I hope you've enjoyed the spreadsheets portion of the course as much as I've enjoyed teaching it. You've come a long way. You've learned to manipulate data sets with all sorts of functions and tools.

You've learned to answer questions about the data, to analyze the data with aggregation tools and pivot tables. And finally you learned how to create a number of different charts and graphs to visualize and present your analysis professionally. Now you're ready to use spreadsheets in projects and presentations of your own.