

Lab 05 – Microsoft Excel

Cleaning data

Contents

Column and Row Re-arranging	2
Transpose.....	5
Sort.....	6
Empty rows	8
Replacing data at character level.....	9
Removing quotes	12
Trailing minus.....	13
Adjusting Date-based Data	15
Text to Column Dates.....	18
Date Value and Time Value.....	20
Times.....	21
Working with Number to Text Conversion Issues	23
Convert text data to values.....	23
Add leading zeroes: Zip codes, other numerical codes and text codes.....	25
Cleaning up Text Data	27
Two new functions: CONCAT and TEXTJOIN to combine data	28
Combine data via concatenation	30
Split data into columns with the Text to Columns feature	33
Flash fill	36
Fill in Missing Title Information	39
Add phone no's and SS# consistently	41
Remove duplicate rows from a list	44

Column and Row Re-arranging

- Download Ch-01 - Open Move Data tab
- Rearrange the order of the column, e.g. phone number next to department. Rather than inserting a new column, do this in a single action
 - Simply select, in this case, the entire column, F, point to the left edge, arrow changes to a four-way arrow.
 - Hold down the shift key and with left mouse button, drag that left edge leftward to the left of the SS# column

	A	B	C	D	E	F	G	H	I	J	K	L
1	Employee Name	Status	Department	Phone	SS#	Building	Hire Date	Years	Month	Benefits	Compensation	Job Rating
2	Barnes, Grant	Hourly	ADC	(919) 581-8082	411-52-6157	Taft	18/01/1998	20	January		42,816	2
3	Edwards, Phillip	Contract	ADC	(252) 353-9786	948-19-5711	West	24/07/2012	5	July		51,048	5
4	Garrett, Chris	Full Time	ADC	(402) 280-4104	100-43-2924	West	31/05/2004	13	May	DMR	29,460	1
5	Saunders, Corey	Full Time	ADC	(252) 813-8394	991-65-6720	Watson	20/05/2013	4	May	M	87,396	2
6	Simmons, Robert	Half-Time	ADC	(252) 177-4590	914-42-8485	West	11/07/1996	21	July	DM	32,154	4
7	Alexander, Charles	Full Time	Admin Training	(919) 844-9868	840-31-3216	Watson	07/05/2012	5	May	M	45,204	3
8	Anderson, Teason	Hourly	Admin Training	(252) 287-2439	771-27-7493	West	30/03/2015	2	March		12,763	4
9	Andrews, Diane	Full Time	Admin Training	(252) 164-1031	638-27-1383	Main	23/10/2009	8	October	DMR	59,220	4
10	Ballard, Martin	Contract	Admin Training	(252) 364-6601	767-96-1463	West	27/01/2006	11	January		92,028	3
11	Bass, Justin	Full Time	Admin Training	(252) 731-7354	542-05-1793	North	28/08/2004	13	August	DMR	90,180	1
12	Bush, Rena	Full Time	Admin Training	(919) 269-3355	975-60-3308	Taft	16/10/2007	10	October	DMR	36,936	4
13	Clarke, Dennis	Half-Time	Admin Training	(252) 793-6742	356-11-0882	Taft	16/07/2004	13	July	DMR	18,288	1
14	Davenport, Troy	Full Time	Admin Training	(252) 785-2326	475-25-6935	South	29/03/2003	14	March	R	102,360	2
15	Fowler, John	Full Time	Admin Training	(919) 407-5460	608-79-6012	North	03/01/1998	20	January	DMR	95,712	5
16	Kent, Angus	Full Time	Admin Training	(802) 167-3267	768-68-1542	North	24/06/2002	15	June	DMR	72,996	2
17	Kramer, Faye	Half-Time	Admin Training	(919) 788-9149	781-91-3936	Main	28/10/2008	9	October	D	21,282	3
18	Love, Danny	Hourly	Admin Training	(919) 523-0846	415-07-6748	South	19/01/1997	21	January	DMR	34,884	3
19	McKinney, Chris	Half-Time	Admin Training	(252) 349-2633	535-53-9723	Main	11/01/1997	21	January	DM	36,534	1
20	Rhodes, Brenda	Contract	Admin Training	(252) 846-7597	202-81-5919	Main	05/09/1998	19	September		79,896	5
21	Salazar, Ruben	Full Time	Admin Training	(919) 468-0033	456-94-6966	Watson	14/01/2006	12	January	D	89,808	4
22	Stephenson, Matt	Full Time	Admin Training	(919) 647-9087	481-33-6564	Taft	03/05/2004	13	May	R	86,508	5
23	Wong, Dennis	Full Time	Admin Training	(252) 583-2994	297-85-2686	West	01/02/2014	3	February	D	69,948	5
24	Adkins, Michael	Full Time	Audit Services	(252) 438-3168	963-02-8490	West	29/03/1997	20	March	DM	49,620	2
25	Gray, Mark	Full Time	Audit Services	(919) 588-2405	237-35-9447	North	18/04/1999	18	April	DMR	88,128	1

- The same can be done with copying multiple columns, highlight column “Phone” and “SS#”, hold down shift key, using the four-way arrow, drag the 2 columns to the column after “Building”, when you see the I-beam indicator, let the mouse go in that position.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Employee Name	Status	Department	Building	Phone	SS#	Hire Date	Years	Month	Benefits	Compensation	Job Rating
2	Barnes, Grant	Hourly	ADC	Taft	(919) 581-8082	411-52-6157	18/01/1998	20	January		42,816	2
3	Edwards, Phillip	Contract	ADC	West	(252) 353-9786	948-19-5711	24/07/2012	5	July		51,048	5
4	Garrett, Chris	Full Time	ADC	West	(402) 280-4104	100-43-2924	31/05/2004	13	May	DMR	29,460	1
5	Saunders, Corey	Full Time	ADC	Watson	(252) 813-8394	991-65-6720	20/05/2013	4	May	M	87,396	2
6	Simmons, Robert	Half-Time	ADC	West	(252) 177-4590	914-42-8485	11/07/1996	21	July	DM	32,154	4
7	Alexander, Charles	Full Time	Admin Training	Watson	(919) 844-9868	840-31-3216	07/05/2012	5	May	M	45,204	3
8	Anderson, Teason	Hourly	Admin Training	West	(252) 287-2439	771-27-7493	30/03/2015	2	March		12,763	4
9	Andrews, Diane	Full Time	Admin Training	Main	(252) 164-1031	638-27-1383	23/10/2009	8	October	DMR	59,220	4
10	Ballard, Martin	Contract	Admin Training	West	(252) 364-6601	767-96-1463	27/01/2006	11	January		92,028	3
11	Bass, Justin	Full Time	Admin Training	North	(252) 731-7354	542-05-1793	28/08/2004	13	August	DMR	90,180	1
12	Bush, Rena	Full Time	Admin Training	Taft	(919) 269-3355	975-60-3308	16/10/2007	10	October	DMR	36,936	4
13	Clarke, Dennis	Half-Time	Admin Training	Taft	(252) 793-6742	356-11-0882	16/07/2004	13	July	DMR	18,288	1
14	Davenport, Troy	Full Time	Admin Training	South	(252) 785-2326	475-25-6935	29/03/2003	14	March	R	102,360	2

- Staying in the same workbook - open worksheet, 'Regional'. We can move rows in this example:
- After looking at the data for awhile you will recognise the need to move rows to suit the data. Row three is about sales, then expenses, then profits. Down in row 9, we're measuring the percent of sales change, but for some reason we've got profits here and expenses here, and this is not quite in-sync with what we see up above. So, in this case, we might want to move the entire row. In other words, but the expenses between sales and profits. Click row 11, and drag upward.

Regional Sales, Expenses, and Profits								
	Jan	Feb	Mar	Apr	May	Jun	Total	Average
Sales	\$ 120.00	\$ 180.00	\$ 250.00	\$ 240.00	\$ 300.00	\$ 450.00	\$ 1,540.00	\$ 256.67
Expenses	100.00	130.00	120.00	220.00	260.00	350.00	1,180.00	196.67
Profits	20.00	50.00	130.00	20.00	40.00	100.00	360.00	60.00
YTD Profits	20.00	70.00	200.00	220.00	260.00	360.00		
YTD Average								
% Sales Change		50.0%	38.9%	-4.0%	25.0%	50.0%	275.0%	30.3%
% Expenses Change		30.0%	-7.7%	83.3%	18.2%	34.6%	250.0%	28.5%
% Profits Change		150.0%	160.0%	-84.6%	100.0%	150.0%	400.0%	38.0%
Sales:Expenses	1.2	1.4	2.1	1.1	1.2	1.3	1.3	
Sales:Profits	6.0	3.6	1.9	12.0	7.5	4.5	4.3	
Expenses:Profits	5.0	2.6	0.9	11.0	6.5	3.5	3.3	

- Moving this row may affect the data across – 'Sales person'

Global Sales, Expenses, and Profits							Salesperson
Feb	Mar	Apr	May	Jun	Total	Average	
\$ 180.00	\$ 250.00	\$ 240.00	\$ 300.00	\$ 450.00	\$ 1,540.00	\$ 256.67	Loman, Willy
130.00	120.00	220.00	260.00	350.00	1,180.00	196.67	Babowsky, Bill
50.00	130.00	20.00	40.00	100.00	360.00	60.00	Furness, Betty
70.00	200.00	220.00	260.00	360.00			Reimers, Ed
							Tilley, Ernest
							Hutton, Lauren
							Popiel, Ron
50.0%	38.9%	-4.0%	25.0%	50.0%	275.0%	30.3%	Pardo, Don
30.0%	-7.7%	83.3%	18.2%	34.6%	250.0%	28.5%	Moss, Ed
150.0%	160.0%	-84.6%	100.0%	150.0%	400.0%	38.0%	Levene, Shelley
1.4	2.1	1.1	1.2	1.3	1.3		

- Just undo that last action.
- This time move only cells. Reduce width of cells in between so we can see the column for 'Salesperson'. Try again, by just selecting the cells and dragging up the cells in row 11 A-I, hold down shift key and drag to the position row 10. This time it should not affect the 'Salesperson' column.
- Back to 'MoveData' worksheet for another technique. Following the same process of selecting a column F, point to edges for four-way arrow. This time when you drag it to the position of column D, click right button of mouse for a menu to appear. Choose 'shift right and move', SSN to column D.

A	B	C	D:D	E	F	G	H	I	J
1	Employee Name	Status	Department	Build	SS#	Hire Date	Years	Month	Benefits
2 Barnes, Grant	Hourly	ADC	Taft	52-6157	1/18/1998	18	January		
3 Edwards, Phillip	Contract	ADC	Wes	19-5711	7/24/2012	4	July		
4 Garrett, Chris	Full Time	ADC	Wes	43-2924	5/31/2004	12	May	DMR	
5 Saunders, Corey	Full Time	ADC	Watso	55-6720	5/20/2013	3	May	M	
6 Simmons, Robert	Half-Time	ADC	Wes	42-8485	7/11/1996	20	July	DM	
7 Alexander, Charles	Full Time	Admin Training	Watso	31-3216	5/7/2012	4	May	M	
8 Anderson, Teason	Hourly	Admin Training	Wes	27-7493	3/30/2015	1	March		
9 Andrews, Diane	Full Time	Admin Training	Mair	Shift Down and Move	27-1383	10/23/2009	7	October	DMR
10 Ballard, Martin	Contract	Admin Training	Wes	Shift Right and Move	96-1463	1/27/2006	10	January	
11 Bass, Justin	Full Time	Admin Training	Nort	Cancel	05-1793	8/28/2004	12	August	DMR
12 Bush, Rena	Full Time	Admin Training	Taft	(919) 269-3355	975-60-3308	10/16/2007	9	October	DMR
13 Clarke, Dennis	Half-Time	Admin Training	Taft	(252) 793-6742	356-11-0882	7/16/2004	12	July	DMR
14 Davenport, Troy	Full Time	Admin Training	South	(252) 785-2326	475-25-6935	3/29/2003	13	March	R
15 Fowler, John	Full Time	Admin Training	North	(919) 407-5460	608-79-6012	1/3/1998	18	January	DMR
16 Kent, Angus	Full Time	Admin Training	North	(802) 167-3267	768-68-1542	6/24/2002	14	June	DMR
17 Kramer, Faye	Half-Time	Admin Training	Main	(919) 788-9149	781-91-3936	10/28/2008	8	October	D
Lewis, Danny	Hourly	Admin Training	South	(919) 572-0846	415-07-6740	1/10/2007	10	January	DMR

MoveData Regional Transpose ExtraRows EmptyRows

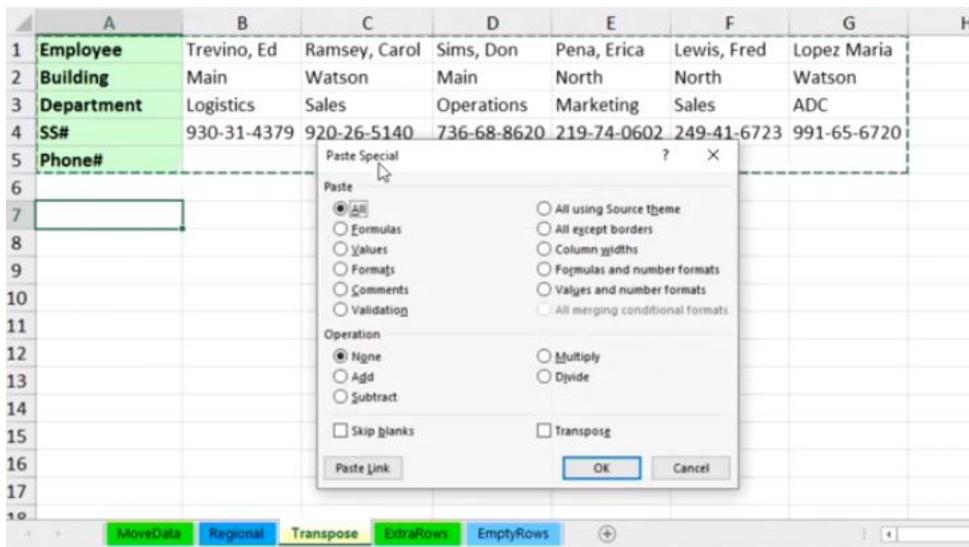
Transpose

The transpose feature is easy to use, and highly recommended, as you work with data, helps to visualize how data might look differently – changing from column to row layout and vice versa.

- In this worksheet called Transpose, there's a list of important contacts within this organization, if these people need to be reached often, we need access to some information about them. Also the list will grow a little bit from time to time, so this layout is not suitable for extending. Categories will be added too so the list might also grow this way as well, but at some point, maybe the thought occurs, could this data have been entered better if we had put the employee name with the heading, and then employee names underneath it.
- To allow the list to grow substantially more, than we're currently seeing, and to improve it so its more readable, we can transpose it. Highlight this data here and copy:

A	B	C	D	E	F	G	H
1 Employee	Trevino, Ed	Ramsey, Carol	Sims, Don	Pena, Erica	Lewis, Fred	Lopez Maria	
2 Building	Main	Watson	Main	North	North	Watson	
3 Department	Logistics	Sales	Operations	Marketing	Sales	ADC	
4 SS#	930-31-4379	920-26-5140	736-68-8620	219-74-0602	249-41-6723	991-65-6720	
5 Phone#							
6							
7							
8							

- Click in cell A7. Then transpose the version, CTRL ALT V, choose transpose from paste special menu. We're going to be flipping this, imagine a diagonal from the upper left to the lower right, rotating the data along the imaginary diagonal.



A	B	C	D	E	F	G	H	I	J	K	Salesperson
1 Employee	Trevino, Ed	Ramsey, Carol	Sims, Don	Pena, Erica	Lewis, Fred	Lopez Maria					
2 Building	Main	Watson	Main	North	North	Watson					Loman
3 Department	Logistics	Sales	Operations	Marketing	Sales	ADC					Babov
4 SS#	930-31-4379	920-26-5140	736-68-8620	219-74-0602	249-41-6723	991-65-6720					Furne
5 Phone#											Reime
6											Tilley,
7 Employee	Building	Department	SS#	Phone#							Hutton
8 Trevino, Ed	Main	Logistics	930-31-4379								Popiel
9 Ramsey, Carol	Watson	Sales	920-26-5140								Pardo
10 Sims, Don	Main	Operations	736-68-8620								Leven
11 Pena, Erica	North	Marketing	219-74-0602								Moss,
12 Lewis, Fred	North	Sales	249-41-6723								
13 Lopez Maria	Watson	ADC	991-65-6720								
14											
15											
16											
17											
18											

- We don't need to keep the old information, although sometimes you might want to compare it, make sure everything got copied. If you have formulas involved, we don't in this case, but if you do, they too get transposed as well.
- Task: Try out the column L for Salesperson – try paste special for different transpose of the data.

Sort

In the worksheet ExtraRows, we've got some valuable information here. Within each department, for example, ADC we see a list of information about the people who work in that department. And scrolling up and down you can see over in column A how the data is indented and every time we have a new department, we've got a new list with a new heading. That heading does repeat the information that we see to the right.

- If we wanted to print this and distribute it to people, we might hide column C, right click Hide.
- If you wanted to work with the data to create a pivot table and do some filtering the worksheet is not usable in its current form. Unhide column C.
- There are many rows with subtotals ADC, Admin training with empty cells. It would not be efficient to delete each row, there's hundreds of them.
- To deal with this issue we could sort the data. Pick one of the columns where there's empty information next to that. Sort based on any column except column A. Choose Department:
- Click in column C, choose the Sort button on Data ribbon:

CH-01 - Excel

Employee Name Building Department Status Hire Date Month Years Benefits Salary Job Rating

Barnes, Grant Taft ADC Hourly 1/11/98 January 18 46,384 2

Simmons, Robert West ADC Half-Time 7/4/96 July 20 DM 34,834 4

Saunders, Corey Watson ADC Full Time 5/13/13 May 3 M 94,679 2

Edwards, Phillip West ADC Full Time 7/17/12 July 4 55,302 5

Garrett, Chris West ADC Full Time 5/24/04 May 12 DMR 31,915 1

Davenport, Troy South Admin Training Full Time 3/22/03 March 13 R 110,890 2

McKinney, Chris Main Admin Training Half-Time 1/4/97 January 19 DM 39,579 1

Fowler, John North Admin Training Full Time 12/27/97 December 18 DMR 103,688 5

Anderson, Teason West Admin Training Hourly 3/23/15 March 1 13,827 4

Ballard, Martin West Admin Training Contract 1/20/06 January 10 99,697 3

Bass, Justin North Admin Training Full Time 8/21/04 August 12 DMR 97,695 1

Salazar, Ruben Watson Admin Training Full Time 1/7/06 January 10 D 97,292 4

Alexander, Charles Watson Admin Training Full Time 4/30/12 April 4 M 48,971 3

Stephenson, Matt Taft Admin Training Full Time 4/26/04 April 12 R 93,717 5

Kramer, Faye Main Admin Training Half-Time 10/21/08 October 8 D 23,056 3

Rhodes, Brenda Main Admin Training Contract 8/29/98 August 18 86,554 5

Louis, Dennis South Admin Training Unknown 1/1/07 January 10 DMR 27,701 2

Choose Sort by: Department in A to Z.

Excel recognises 'My data has headers' so row 1 doesn't get involved...

Ok

CH-01 - Excel

Employee Name Building Department Status Hire Date Month Years Benefits Salary Job Rating

Barnes, Grant Taft ADC Hourly 1/11/98 January 18 46,384 2

Simmons, Robert West ADC Half-Time 7/4/96 July 20 DM 34,834 4

Saunders, Corey Watson ADC Full Time 5/13/13 May 3 M 94,679 2

Edwards, Phillip West ADC Full Time 7/17/12 July 4 55,302 5

Garrett, Chris West ADC Full Time 5/24/04 May 12 DMR 31,915 1

Davenport, Troy South Admin Training Full Time 3/22/03 March 13 R 110,890 2

McKinney, Chris Main Admin Training Half-Time 1/4/97 January 19 DM 39,579 1

Fowler, John North Admin Training Full Time 12/27/97 December 18 DMR 103,688 5

Anderson, Teason West Admin Training Hourly 3/23/15 March 1 13,827 4

Ballard, Martin West Admin Training Contract 1/20/06 January 10 99,697 3

Bass, Justin North Admin Training Full Time 8/21/04 August 12 DMR 97,695 1

Salazar, Ruben Watson Admin Training Full Time 1/7/06 January 10 D 97,292 4

Alexander, Charles Watson Admin Training Full Time 4/30/12 April 4 M 48,971 3

Stephenson, Matt Taft Admin Training Full Time 4/26/04 April 12 R 93,717 5

Kramer, Faye Main Admin Training Half-Time 10/21/08 October 8 D 23,056 3

Rhodes, Brenda Main Admin Training Contract 8/29/98 August 18 86,554 5

Louis, Dennis South Admin Training Unknown 1/1/07 January 10 DMR 27,701 2

- What happened to those titles, they're down at the bottom. Ctrl End. The titles can be deleted or moved to an empty sheet.

- Now we can work with the data without concern about those headings. We can sort the data, we can filter the data, we can create a Pivot Table with this, we can perform other features as well without worrying about those embedded headings that we had.

Empty rows

- Open the **EmptyRows** tab next.
- In this data we have empty rows, which we need to be removed. We want to highlight all the data. Go to the end, Ctrl End, click in the very last cell, L764. To highlight to the top, hold down shift key and choose Ctrl Home on keyboard.
- Choose sort on data ribbon, sort

The screenshot shows an Excel spreadsheet titled "CH-01 - Excel". The Data ribbon tab is selected. A "Sort" dialog box is open, showing the settings: "Column" is "Employee Name", "Sort On" is "Values", and "Order" is "A to Z". The "My data has headers" checkbox is checked. The data range selected in the dialog is E756:L764. The main spreadsheet area shows several rows of employee data, and the bottom status bar indicates "Average: 1068011337 Count: 8699 Sum: 4.74838E+12".

- Sort by Employee Name, A to Z. Excel picks up headers in our list, so make sure that's checked ok.
- The data's been sorted. The empty rows are at the bottom and there's our contiguous list.

C1	Employee Name	Building	Department	SS#	Phone	Status	Hire Date	Month	Years	Benefits	Salary	Job R
1	Abbott, James	Watson	Professional Training Group	834-06-1135	(919) 847-2270	Full Time	2/12/97	February	19	M	57,928	2
2	Acosta, Robert	Watson	Manufacturing	484-21-7278	(919) 562-7374	Hourly	3/9/14	March	2		13,744	4
3	Adams, David	Main	Professional Training Group	444-15-9297	(252) 245-6406	Full Time	3/7/97	March	19	DMR	105,989	5
4	Adkins, Michael	West	Audit Services	963-02-8490	(252) 438-3168	Full Time	3/22/97	March	19	DM	53,755	2
5	Aguilar, Kevin	Taft	Quality Control	688-76-9770	(919) 241-6398	Full Time	2/12/11	February	5	DMR	57,889	2
6	Alexander, Charles	Watson	Admin Training	840-31-3216	(919) 844-9868	Full Time	4/30/12	April	4	M	48,971	3
7	Allen, Thomas	West	Pharmacokinetics	425-94-3144	(252) 291-1046	Contract	7/27/99	July	17		93,210	2
8	Allison, Timothy	North	Quality Assurance	323-70-1315	(919) 447-9196	Full Time	6/20/16	June	0	M	104,338	3
9	Alvarado, Sonia	Main	Quality Assurance	412-15-9105	(919) 825-2392	Hourly	4/16/02	April	14		43,560	4
10	Alvarez, Steven	South	Major Mfg Projects	477-11-0649	(919) 135-1512	Full Time	10/13/06	October	10	DM	58,695	1
11	Anderson, Teason	West	Admin Training	771-27-7493	(252) 287-2439	Hourly	3/23/15	March	1		13,827	4
12	Andrews, Diane	Main	Admin Training	638-27-1383	(252) 164-1031	Full Time	10/16/09	October	7	DMR	64,155	4
13	Anthony, Robert	Main	Process Development	931-97-7751	(919) 447-1952	Full Time	12/1/00	December	15	DMR	33,579	5
14	Armstrong, David	Watson	Manufacturing	378-18-9642	(252) 622-8199	Contract	2/15/09	February	7		83,486	5
15	Arnold, Cole	North	Logistics	659-92-9807	(919) 308-9561	Full Time	6/19/11	June	5	DMR	29,133	4
16	Ashley, Michael	Main	Quality Control	876-08-2195	(252) 604-9607	Full Time	10/30/04	October	12	D	80,405	2
17	Atkins, Kevin	West	Operations	361-03-0327	(252) 333-3772	Full Time	7/14/02	July	12	D	112,070	3

- We cleaned up our data, by getting rid of rows.

Replacing data at character level

This feature Find and Replace is very useful for replacing a digit or a letter or a word once a column is selected. You can replace a single character with zero, one or many characters. You can imagine all kinds or variations on how you could use this feature. The following will show you the capabilities of using the Replace command:

- Download file for this exercise CH-02, open sheet **Substitute**
- We need to make a change in Column B. Company's restructuring it's HR categories, so hourly is no longer an option. Also, we want to change Half-Time to Part-Time.
- Select Column B, on the Home tab, the far right editing group, go to Find & Select, Replace.
- Replace that phrase Half-Time, it's got a dash in it, tab, Part-Time, and we went to Replace all of them

A	B	C	D	E	F	G	H
Employee Name	Status	Department	Phone	Hire Date			Part #
Barnes, Grant	Hourly	ADC					\$D8
Edwards, Phillip	Contract	ADC					\$D8
Garrett, Chris	Full Time	ADC					\$D4
Saunders, Corey	Full Time	ADC					\$D4
Simmons, Robert	Half-Time	ADC					\$D9
Alexander, Charles	Full Time	Admin Tra					\$D1
Anderson, Teason	Hourly	Admin Tra					\$D6
Andrews, Diane	Full Time	Admin Tra					\$D9
Ballard, Martin	Contract	Admin Training	(252) 364-6601	27/01/2006			258754D8
Bass, Justin	Full Time	Admin Training	(252) 731-7354	28/08/2004			919255D1
Bush, Rena	Full Time	Admin Training	(919) 269-3355	16/10/2007			919749D2
Clarke, Dennis	Half-Time	Admin Training	(252) 793-6742	16/07/2004			252131D9
Davenport, Troy	Full Time	Admin Training	(252) 785-2326	29/03/2003			252333D7
Fowler, John	Full Time	Admin Training	(919) 407-5460	03/01/1998			919460D9
Kent, Angus	Full Time	Admin Training	(802) 167-3267	24/06/2002			919433D1

SUBSTITUTE REPLACE Quotes Trailing Minus +

- It will tell us how many it has replaced, in this case 96.
- TASK: In Column H, try replacing D8 with 'x'.
- There's also a function called Replace, but it's not like the Replace command, and strangely enough even though those words are the same, there is another function that's very similar to Replace, but it's called Substitute. The Substitute function is like the Replace command, its no faster than the above. Here's an example:

	E	F	G	H	I	J	K	L
1	Hire Date			Part #			Project #	
2	1/18/1998			402204D8	=substitute(H2,8,"q")		3848	
3	7/24/2012			258190D8			91*5	
4	5/31/2004			252386D4			1742	
5	5/20/2013			988582D4			8*65	
6	7/11/1996			252894D9			2804	

- The only time you might use it is if you wanted to state the replacing of a digit that's appearing at the beginning or at a particular position, e.g. if we just wanted to replace the first '8', see below:

1	Hire Date	Part #	Project #
2	1/18/1998	=SUBSTITUTE(H2,8,"q",1)	3848
3	7/24/2012	258190D8	91*5
4	5/31/2004	252386D4	1742
5	5/20/2013	988582D4	8*65
6	7/11/1996	252894D9	2804

- Putting in the '1' for instance_num means you will replace all first 8's with a 'q'.
- In column K there are '*' asterisks. These are special characters, like the '?' in column M. The Find and replace does not work with special characters, unless you include a tilde '~' see below:

E	F	G	H	I	J	K
Hire Date						Project #
18/01/1985						3848
24/07/1985						91*5
31/05/1985						1742
20/05/1985						8*65
11/07/1985						2804
07/05/1985						83*5
30/03/1985						3143
23/10/2009						6*12
27/01/2006						6275

- TASK: Try replacing the '?' in column M with lowercase 'x'.
- In column 'O' every time we see G and Y, we want to replace it and everything including the characters between them, so cells have 2 characters between them, some have 5.
- We are going to use the wildcard symbol to represent the characters between G and Y. Replace G...Y with 'x'. See below:

Code #	SUBSTITUTION
103634g4849359y111	Replace
103638g5422596y534	Replace
104010gy158	Replace
1545g2y74	Replace
17g118y59	Replace
18182g3891591y263	Replace
2200g6508258y786	Replace
2265g3456y130	Replace
29334g851432y696	Replace
30063g7449060v484	Replace

- The Replace function, works with data differently. The Replace function is unlike the Replace command. The Replace function is based on location or position.
- Open worksheet Replace.
- We've got some code numbers in column B. Some codes are positional, meaning that maybe the third character has to do with a color of an item, the fourth character might have to do with the year that it was made. These could have something to do with size, location, any number of different things. We need to make some changes, see the example below how replace function works:

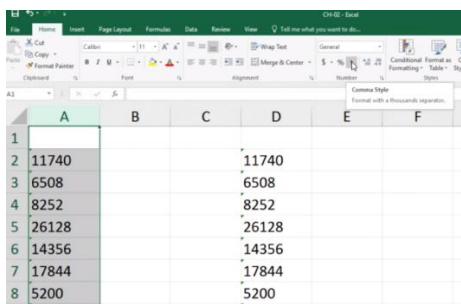
Department	Code	
ADC	40224D8	=replace(B2,3,2,"AA")
ADC	25210D7	REPLACE(old_text, start_num, num_chars, new_text)
ADC	25236D4	
ADC	91952D4	
ADC	25284D9	
Admin Training	80217D1	Replace multiple character
Admin Training	25272D6	Replace multiple character

Department	Code
ADC	40224D8=REPLACE(B2,3,2,"AA")
ADC	25210D725AA0D7
ADC	25236D425AA6D4
ADC	91952D491AA2D4
ADC	25284D925AA4D9
Admin Training	80217D180AA7D1
Admin Training	25272D625AA2D6
Admin Training	25287D925AA7D9
Admin Training	25274D825AA4D8

- We can use this replace to replace 3 characters instead of 2. There's lots of variations on this.
- TASK: Try replacing the second position characters with nothing?

Removing quotes

- Open the 'quotes' worksheet.
- Sometimes when you get data from another source, you might see numbers that have a single quote in front of them. Any number of obscure reasons for this, but when we deal with situations like this, sometimes it's confusing. You cannot format the cells with quotes unless you turn them into numbers.

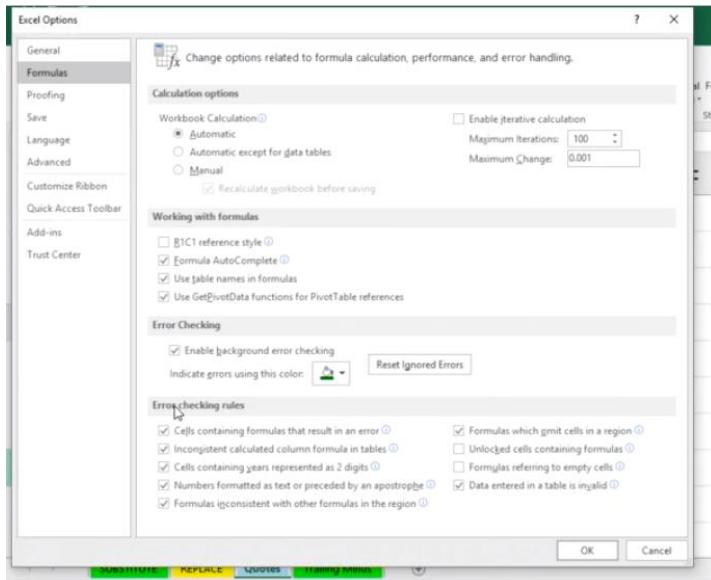


The screenshot shows a Microsoft Excel spreadsheet titled "CH02 - Basic". The data is located in column A, starting from row 1. The values are: 11740, 6508, 8252, 26128, 14356, 17844, and 5200. Each value has a single quote character (') at the beginning. The ribbon at the top shows the Home tab selected. The formula bar above the first cell contains the formula '=A1'. The status bar at the bottom indicates "CH02 - Basic".

A	B	C	D	E	F
1					
2 11740		11740			
3 6508		6508			
4 8252		8252			
5 26128		26128			
6 14356		14356			
7 17844		17844			
8 5200		5200			

- You might see green triangles. This is a particular Excel setting that you can turn on or off. File tab in the ribbon, down the left-hand side, Options. And under the Excel Options, go to Formulas. And at the bottom here, there are error checking rules. This is a bit of a misnomer.

Not all of these are errors. There's a box next to a choice here that says "Numbers formatted as text or preceded by an apostrophe."



- This can remove the green triangle warning.
- Otherwise, leave the green warning and check...

	A	B	C
1			
2	11740		
3	6508		
4	8252		
5	26128		
6	14356		

- In this example, choose convert it to a number, highlight all column A and choose.

Trailing minus

- Open worksheet, 'Trailing minus'. To deal with minus
- Select column A, go to data ribbon, 'Text to Columns'...

CH-02 - Excel

A	B	C
1	Amount	
2	56.56-	
3	317.42	
4	23.00-	
5	26.41-	
6	23.45-	

- Next, until step 3. Then choose 'Advanced'.

A	B
1	Amount
2	56.56-
3	317.42
4	23.00-
5	26.41-
6	23.45-
7	321.42
8	3.45-

- Trailing minus for negative numbers. Click ok and finish and they're gone. And the negative appears here. Can we use this now?

A	B	C
1	Amount	
2	-56.56 =A2+1	
3	317.42	
4	-23	
5	-26.41	
6	-23.45	
7	321.42	
8	-3.45	
9	46.78	

- Yes we can. Try = this value, plus one! The main idea here was of course, we got rid of those trailing minuses and turn them into minuses that Excel can work with. And the negative data appears the way we want it based on our own formatting.

Adjusting Date-based Data

Sometimes when you get information from other sources, the dates are put in in an unusual way and sometimes incomplete.

- Open CH-03 Workbook, tab/sheet: **DateFormats**
- There are some entries in column A, also in column C, that are similar, they're in reverse order. They've got month and year but not a day associated with them.
- In column E there's all three of the date parts, but in a different layout. The key to converting these many times is a combination of using a function called DATE as displayed in this generic form here in cell B1 and also some of the text functions, like left and right, that allow us to extract data from the left side or right side of a cell. See the format below:

A	B	C	D	E
Month Year	=DATE(Year, Month, Day)	Year Month	=DATE(Year, Month, Day)	Year Month Day
10 2016		2016 10		2015 Sep 21
4 2016		2016 4		2016 Jan 1
5 2016		2016 5		2015 MAR 13
9 2015		2015 9		2016 Jan 1
10 2016		2016 10		2015 Jun 1
6 2016		2016 6		2016 AUG 12
4 2015		2015 4		2015 Jan 1
5 2016		2016 5		2016 Feb 14
1 2016		2016 1		2015 Jan 12
12 2015		2015 12		
12 2015		2015 10		

- The key to converting these many times is a combination of using a function called DATE as displayed in this generic form here in cell B1 and also some of the text functions, like left and right, that allow us to extract data from the left side or right side of a cell.

A	B	C	D	E
Month Year	=DATE(Year, Month, Day)	Year Month	=DATE(Year, Month, Day)	Year Month Day
10 2016	=date(right(A2,4),left(A2,2),1)	2016 10		2015 Sep 21
4 2016		2016 4		2016 Jan 1
5 2016		2016 5		2015 MAR 13
9 2015		2015 9		2016 Jan 1
10 2016		2016 10		2015 Jun 1
6 2016		2016 6		2016 AUG 12
4 2015		2015 4		2015 Jan 1
5 2016		2016 5		2016 Feb 14
1 2016		2016 1		2015 Jan 12
12 2015		2015 12		
12 2015		2015 10		

- In cell B2: =date ... link to the year out of cell A2. To access the four right-most characters, we use the function right. Cell A2 comma 4, meaning pull out those four right-most characters, that's the year.

- Then we need the month, that's the two left-most characters. We're in cell A2 again comma two. Now there's no day over there, so make it day one, first day of the month, comma one. Right parenthesis. We've got a date.
- Copy/autofill this down the column.
- We might also need something almost the same in column D. Since this formula we just created is based on the cell to its left, we can copy this formula we just created in column B. We've a different layout. So you can edit it as follows:

	A	B	C	D	E
1	Month Year	=DATE(Year, Month, Day)	Year Month	=DATE(Year, Month, Day)	Year Month Day
2	10 2016	10/1/2016	2016 10	=DATE(left(C2,4),right(C2,2),1)	2015 Sep 21
3	4 2016	4/1/2016	2016 4		2016 Jan 1
4	5 2016	5/1/2016	2016 5		2015 MAR 13
5	9 2015	9/1/2015	2015 9		2016 Jan 1
6	10 2016	10/1/2016	2016 10		2015 Jun 1
7	6 2016	6/1/2016	2016 6		2016 AUG 12
8	4 2015	4/1/2015	2015 4		2015 Jan 1
9	5 2016	5/1/2016	2016 5		2016 Feb 14
10	1 2016	1/1/2016	2016 1		2015 Jan 12
11	12 2015	12/1/2015	2015 12		
12	12 2015	12/1/2015	2015 10		

- In column E the date is formatted differently, it includes the 3 parts to the date. Notice J and K showing each month and sequence – this could be helpful here.
- To input a formula in F1 to format column E we could use a lookup table to somehow come up with the actual numbers. E.g., September is nine, the ninth month and so on. This is done inside out here. So in cell E2, how to get Sep to recognise as '9'.
- Use a function called Mid. '6' represents the space where 'S' for Sep starts in cell E2. '3' represents the 3 characters – 'Sep'. Hit enter, that will give us 'Sep' in cell F2.

	CH-BB - Excel											
	CH-BB - Excel											
	CH-BB - Excel											
	E	F	G	H	I	J	K	L	M	N	O	P
1	Year Month Day	=DATE(Year, Month, Day)				Jan	1					
2	2015 Sep 21	=mid(E2,6,3)				Feb	2					
3	2016 Ján 1					Mar	3					
4	2015 MAR 13					Apr	4					
5	2016 Jan 1					May	5					
6	2015 Jun 1					Jun	6					
7	2016 AUG 12					Jul	7					
8	2015 Jan 1					Aug	8					
9	2016 Feb 14					Sep	9					
10	2015 Jan 12					Oct	10					
11						Nov	11					
12						Dec	12					

- To replace the 'Sep' with '9', we use the Vlookup function. The purpose of the Vlookup is: I've got a value somewhere, in other words for Sep...I want to compare it with a table of data somewhere. That table in this example is over in columns J and K. See function format below:

	E	F	G	H	I	J	K	L
1	Year Month Day	=DATE(Year, Month, Day)				Jan	1	
2	2015 Sep 21	=VLOOKUP(MID(E2,6,3),J:K,2,FALSE)				Feb	2	
3	2016 Jan 1		1			Mar	3	
4	2015 MAR 13		3			Apr	4	
5	2016 Jan 1		1			May	5	
6	2015 Jun 1		6			Jun	6	
7	2016 AUG 12		8			Jul	7	
8	2015 Jan 1		1			Aug	8	
9	2016 Feb 14		2			Sep	9	
10	2015 Jan 12		1			Oct	10	
11						Nov	11	
12						Dec	12	

- Start with: =VLOOKUP(MID(E2,6,3),J:K,2,FALSE) The '2' means the second column which is 'K' and 'FALSE' means – choosing exact match. If all goes well, it should have '9' as the answer. Convert this to Number – general format from home ribbon. Just to test this in column 'F', autofill down this column 'F' to see if it picked up the month as the numerical field:

	E	F	G	H	I	J	K	L
1	Year Month Day	=DATE(Year, Month, Day)				Jan	1	
2	2015 Sep 21		9			Feb	2	
3	2016 Jan 1		1			Mar	3	
4	2015 MAR 13		3			Apr	4	
5	2016 Jan 1		1			May	5	
6	2015 Jun 1		6			Jun	6	
7	2016 AUG 12		8			Jul	7	
8	2015 Jan 1		1			Aug	8	
9	2016 Feb 14		2			Sep	9	
10	2015 Jan 12		1			Oct	10	
11						Nov	11	
12						Dec	12	

- All that, to come up with a month!! To format the Date in column E – edit the formula as follows:

	E	F	G	H	I	J	K
1	Year Month Day	=DATE(Year, Month, Day)				Jan	1
2	2015 Sep 21	=date(left(E2,4),VLOOKUP(MID(E2,6,3),J:K,2,FALSE),right(E2,2))				Feb	2
3	2016 Jan 1		1			Mar	3
4	2015 MAR 13		3			Apr	4
5	2016 Jan 1		1			May	5
6	2015 Jun 1		6			Jun	6
7	2016 AUG 12		8			Jul	7
8	2015 Jan 1		1			Aug	8
9	2016 Feb 14		2			Sep	9
10	2015 Jan 12		1			Oct	10
11						Nov	11
12						Dec	12

- At different times you're going to have to go to these unusual lengths to achieve Date formatting:

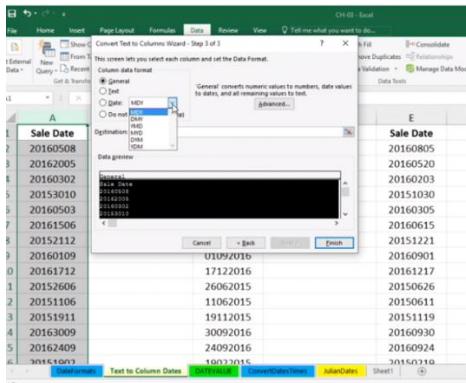
	E	F	G	H	I	J	K
1	Year Month Day	=DATE(Year, Month, Day)					
2	2015 Sep 21	9/21/2015				Jan	1
3	2016 Jan 1	1/1/2016				Feb	2
4	2015 MAR 13	3/13/2015				Mar	3
5	2016 Jan 1	1/1/2016				Apr	4
6	2015 Jun 1	6/1/2015				May	5
7	2016 AUG 12	8/12/2016				Jun	6
8	2015 Jan 1	1/1/2015				Jul	7
9	2016 Feb 14	2/14/2016				Aug	8
10	2015 Jan 12	1/12/2015				Sep	9
11						Oct	10
12						Nov	11
						Dec	12

- To clean up the columns, we won't want the original column now that we have reformatted. To quickly copy, bring your mouse to any edge of column F, hold it, begin to move it to next column, click right button of mouse to choose option with copying.

E	F	G	H
1	Year Month Day	=DATE(Year, Month, Day)	
2	2015 Sep 21	9/21/2015	
3	2016 Jan 1	1/1/2016	
4	2015 MAR 13	3/13/2015	
5	2016 Jan 1	1/1/2016	
6	2015 Jun 1	6/1/2015	
7	2016 AUG 12	8/12/2016	
8	2015 Jan 1	1/1/2015	
9	2016 Feb 14	2/14/2016	
10	2015 Jan 12	1/12/2015	
11			

Text to Column Dates

- Open sheet, **Text to Column Dates**
- View column A, C and E. Each have dates, but they're in unusual formats. It can be difficult to identify which is year and month, especially in row 8 of column A. You have to figure out which is day, month and year. Brief look can help you identify this overall.
- Try this:
 - Highlight column A, click on data ribbon, choose Text to Columns.
 - Use wizard- Next and Next. Step 3 to column data format –



- There are 6 different variations of formatting of how date is formatted.
- Choose YMD as you notice in column A has year first and then finish.

A
Sale Date
20160508
20160505
20160302
20153010
20160503
20161506
20152122
20160109
20161712
20152606
20151106
20151911
20163009
20162409
20151003

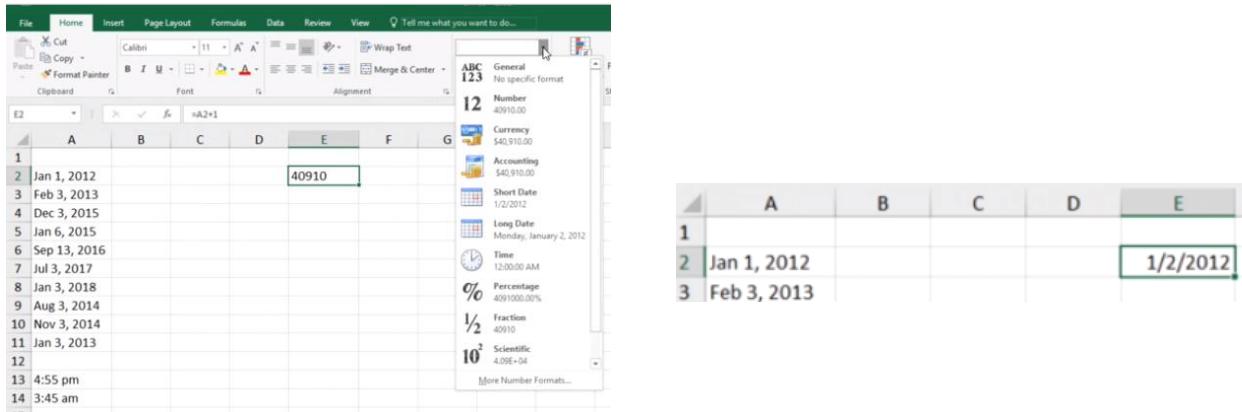
- We've got dates, you could align it...
- Try column C and F.
- Column C. The day on left and year on right:

A	B	C
Sale Date		Sale Date
8/5/2016		05082016
5/20/2016		20032016
2/3/2016		03022016
10/30/2015		30102015
3/5/2016		05032016
6/15/2016		15062016
12/21/2015		21122015
9/1/2016		01092016
12/17/2016		17122016
6/26/2015		26062015
6/11/2015		11062015
11/19/2015		19112015
9/30/2016		30092016
9/24/2016		24092016
2/10/2015		10022015

- Col E – year on left side. YMD finish.

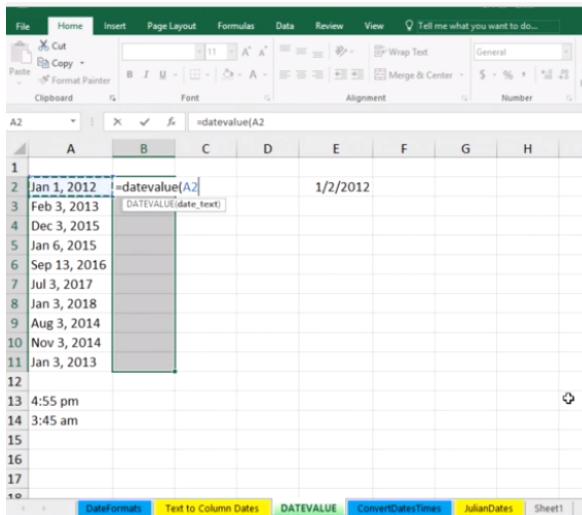
Date Value and Time Value

- Open sheet, **DateValue**.
- Notice on the Home ribbon, you have different ways of dealing with date in the number section. You can format date – short date, long date. If column A was a date format, we could test it by typing in cell E2: =A2+1 . We would get that number in E2. Could it be formatted? Click on cell E2 and choose short date.



The screenshot shows the Excel interface with the Home tab selected. The formula bar shows '=A2+1'. Cell E2 contains the value '40910'. The 'Number' dropdown menu is open, displaying options like General, Number, Currency, Accounting, Short Date, Long Date, and Time. The 'Short Date' option is highlighted, showing the result '1/2/2012'. The rest of the sheet shows dates from Jan 1, 2012, to Feb 3, 2013.

- Another way which is faster will clean up dates. The formula is called DATEVALUE. If you've had a situation where you got dates that have somehow been treated as text or converted into text. That happens sometimes when you're copying data through various interim software packages or sometimes moving between workbooks.
- Highlight column B2:b11, start typing =DATEVALUE (it automatically types in cell B2)



The screenshot shows the Excel interface with the Home tab selected. The formula bar shows '=datevalue(A2)'. The selected range is B2:B11. Cell B2 contains the formula '=datevalue(A2)' and displays the value '1/2/2012'. The formula bar also shows '=DATEVALUE(date_text)'. The status bar at the bottom shows tabs for DateFormats, Text to Column Dates, DATEVALUE, ConvertDateTimes, JulianDates, and Sheet1.

- You get this, it doesn't look valid...

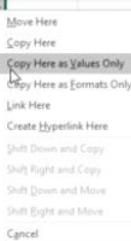
Jan 1, 2012	40909
Feb 3, 2013	41308
Dec 3, 2015	42341
Jan 6, 2015	42010
Sep 13, 2016	42626
Jul 3, 2017	42919
Jan 3, 2018	43103
Aug 3, 2014	41854
Nov 3, 2014	41946
Jan 3, 2013	41277

- From the home tab, number section, convert into a short date display. We now have dates that are workable...

A	B
Jan 1, 2012	1/1/2012
Feb 3, 2013	2/3/2013
Dec 3, 2015	12/3/2015
Jan 6, 2015	1/6/2015
Sep 13, 2016	9/13/2016
Jul 3, 2017	7/3/2017
Jan 3, 2018	1/3/2018
Aug 3, 2014	8/3/2014
Nov 3, 2014	11/3/2014
Jan 3, 2013	1/3/2013

- We don't want the formulas here in column B to work with it. Highlight the cells in B, drag it into another column or move it back to column B on top of itself, right click and choose, "copy here as values only"

Jan 1, 2012	1/1/2012
Feb 3, 2013	2/3/2013
Dec 3, 2015	12/3/2015
Jan 6, 2015	1/6/2015
Sep 13, 2016	9/13/2016
Jul 3, 2017	7/3/2017
Jan 3, 2018	1/3/2018
Aug 3, 2014	8/3/2014
Nov 3, 2014	11/3/2014
Jan 3, 2013	1/3/2013
4:55 pm	
3:45 am	



- Now, these are converted, we've got dates we can work with. So you could remove the original dates in column A.

Times

- Same thing can happen with times, in cell B13, could we add an hour to this? Add 1/24 of a day?

4:55 pm	=A13+(1/24)
3:45 am	

- It's a number – go to the home ribbon again, number section, change to the time format...

A	B	C	D	E	F
1					
2	1/1/2012			1/1/1900	
3	2/3/2013				
4	12/3/2015				
5	1/6/2015				
6	9/13/2016				
7	7/3/2017				
8	1/3/2018				
9	8/3/2014				
10	11/3/2014				
11	1/3/2013				
12					
13	4:55 pm	0.74652778			
14	3:45 am				
15					

4:55 pm 5:55:00 PM
3:45 am

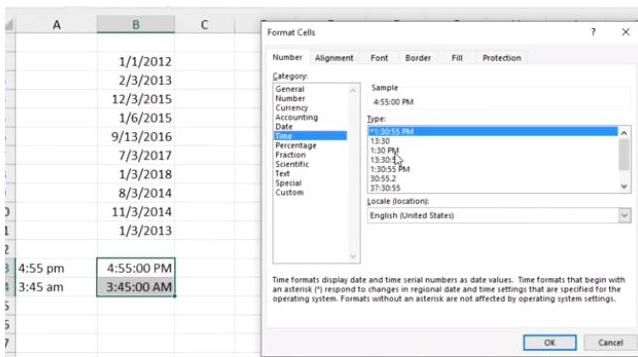
- That's workable.
- It could be faster to use timevalue function:

4:55 pm =timevalue(A13)
3:45 am TIMEVALUE(time_text)

- Press control +enter, changes selected together at one time.
- Sometimes again in the copying of cells back and forth formats get copied. Highlight both and choose time from home ribbon.

4:55 pm 4:55:00 PM
3:45 am 3:45:00 AM

- If those are not to your liking – the formats can be changed from format cells. AMPM without seconds or 24 hour style.



- Converting dates that have been entered as text or converted to text need to be cleaned so they're suitable for working on, same situation with time we use either the function Date Value or Time Value to clean up that information.

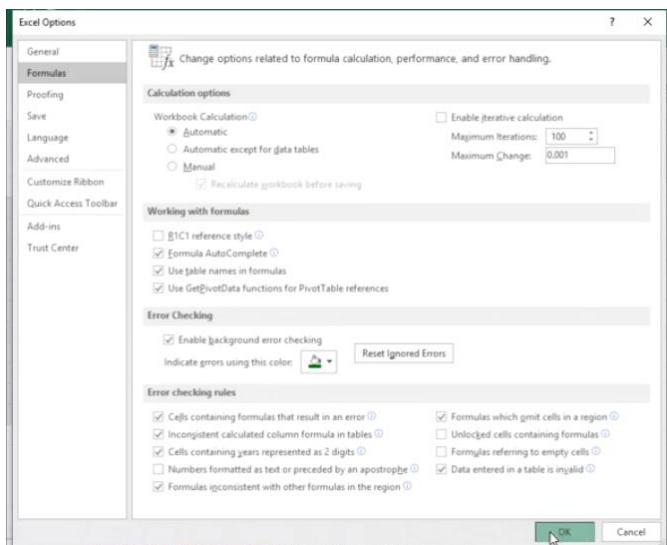
Working with Number to Text Conversion Issues

Convert text data to values

Sometimes you get data from other sources, you don't know how they were displayed as text, or why they were converted. Half the time it's a mistake, or it wasn't done on purpose.

There are two techniques to deal with this, if Excel prompts you by indicating a green triangle you can simply select the conversion by highlighting the cells. The other method is pick an empty cell, and copy that and add it to the numbers that are text by way of paste special.

- Open Ch-04 Workbook, Sheet: **Text to values**
- You will notice green indicators on each cell.
- As mentioned before, in case you don't see the green triangle, it could be that its disabled. Open Excel Options, Formulas, Error Checking Rules. See below "Numbers formatted as text or preceded by an apostrophe". Enable this feature as the green triangle prompt is very useful when you want to convert the values.



- When you accept the options from the green triangle, highlight column B and convert from text to numbers.

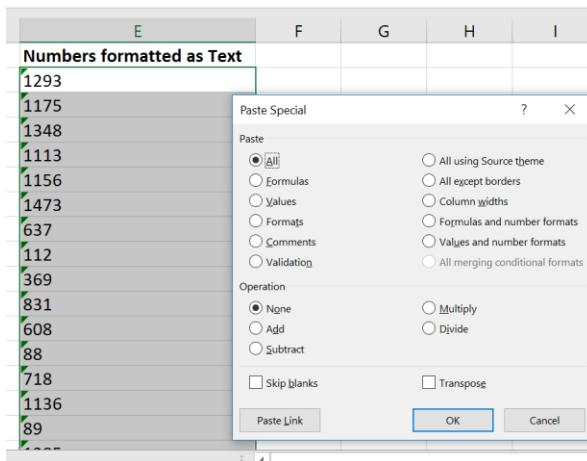
	A	B	C	D	E	F
1					Numbers formatted as Text	
2		1293			1293	
3		1175			1175	
4		1348			1348	
5		1113			1113	
6		1156			1156	
7		1473			1473	
8		637			637	
9		112			112	
10		369			369	
11		831			831	
12		608			608	
13		88			88	
14		718			718	
15		1136			1136	
16		89			89	
17		1385			1385	
18						
					Text to Values	Leading Zeros
					Sheet1	Sheet2
					Sheet3	

- When you click in column E, notice the Home ribbon, number section. The numbers are formatted as text. If you click in cell E2 and try to add a dollar sign or a comma, it does not change.
- If you try to work with the values and click in cell F2 and type a formula =E2*2. Notice the cell F2 format is accounting format, so you can work with these numbers. Add currency, comma etc.

The oddity about text formatting is that when you apply text formatting, when you apply later formats, it doesn't really remove the formatting. That sort of sounds like a contradiction. It can be confusing.

How do we clean these up?

- Go to a blank cell, F3, choose copy.
- Highlight the column E2:E20 and choose paste special, Ctrl-Alt-V. We are copying a blank cell



- Choose Add (we're adding what we're copying – blank cell). The numbers will shift but nothing different to note.
- What you will notice is, if you click in cell E column and go to Home ribbon to change formatting, you can add commas and currency etc.
- We have converted these into standard numbers and removed the text formatting.

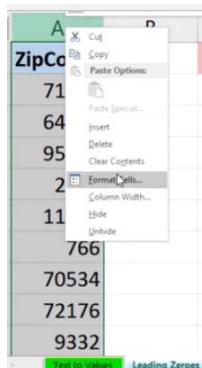
Add leading zeroes: Zip codes, other numerical codes and text codes

When you get data that should've had leading zeroes but doesn't, it poses some problems and some different issues occur. We need to clean up the data for multiple reasons on the following worksheet "Leading Zero's":

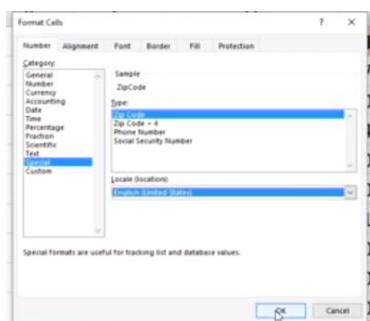
- Open tab/sheet – **Leading Zero's**
- Column A - Someone entered zip codes here, maybe it was just essentially looking at a sheet of paper, typing an entry, pressing Enter, not looking up, and here that person typed 02299 in row 5, and 00766 in row 7. What happens to the leading zeroes? Well they're not there. These are in so-called general format, no special formatting was applied.

1	ZipCode	ID #	6 Characters	ID #	Code #
2	71396	75757		075757	A348U8
3	64134	2026		002026	302U302
4	95192	44613		044613	13891E891
5	2299	814		000814	47006F006
6	11348	1411		001411	45615R152
7	766	16259		016259	4503N503
8	70534	4784		004784	93458D847
9	72176	3247		003247	3485B485
10	9332	4098		004098	51E51

- A Special format could have been applied. Right click column A. format cells.



- Number tab, choose special under category. There's 2 variations on zip codes. Try the first one, with 5 digits.



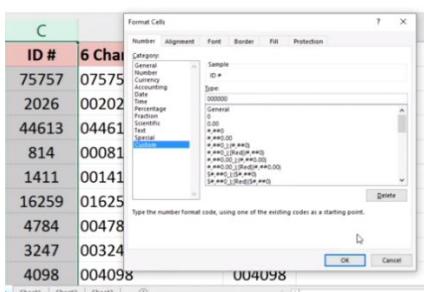
- Watch the entries in column A, they now have leading zero's. That means when you're typing these you don't have to type the leading zero's.
- Column C – these are ID numbers – somebody was typing these and typed leading zeroes but we don't see them there. We want them to appear the way they appear in column E. How can this be fixed – there's 2 ways.

ZipCode	ID #	6 Characters	ID #	Code #
71396	75757	075757	075757	A348U8
64134	2026	002026	002026	302U302
95192	44613	044613	044613	13891E891
00123	814	000814	000814	47006F006
11348	1411	001411	001411	45615R152
00766	16259	016259	016259	4503N503
70534	4784	004784	004784	93458D847
72176	3247	003247	003247	3485B485
09332	4098	004098	004098	51E51

- The first way to try is longer, it's done by using the function TEXT.
- TEXT allows us to pick up numerical information, but we have to provide the form, we want the format to be 6 zero's. Each zero stands for a numerical representation of characters.
- =text(C2, "000000")
- Autofill down the column

C	D
ID #	6 Characters
75757	075757
2026	002026
44613	044613
814	000814
1411	001411
16259	016259
4784	004784
3247	003247
4098	004098

- The other method: highlight column C – right click, format cells and go to custom



- This will only work if all our entries are numerical, as with the previous technique too.
- Off to the right, in column G. We want column G to look like column I. There is 9 characters.
- We need a function to put in as many zero's as we need.

- Try this formula in cell H2: =Rept(9,4) We get value 9, repeated 4 times. If it's a letter or symbol we must put in "" double quotes. What if we put in an unusual character, try: =Rept("%", 4)
- Remove the formula you tried out in cell H2. We want a 9 character representation here. You will need to begin by saying you want to put in a zero, but we're not sure how many ties, so we need another function. How wide is the entry? Or how long is the entry?
- Ultimately we want nine characters, if cell G2 has 6, we need to repeat zero 3 times. We use a function call LEN (think of length)

Code #	9 Characters
A348U8	=REPT("0",9-LEN(G2))&G2
302U302	
13891E891	

- The REPT function says, let's repeat this character, how many times. The number nine minus the length of the entry already. What if it has nine? Look at the entry in row 5, in G5 we see an entry there, it has nine characters. What is nine minus this length of the cell? We put in the number zero, zero times. And will it work here without the double quote? When you work with numbers often you don't need to use the double quote. The & symbol allows us to join data from different locations, in this case G2.
- Copy down the column. They should be all 9 entries wide. In some cases with letters it doesn't work, row 23. Letters are different. The main idea we're able to reconstruct the numbers and leading zero's.
- After creating the information the way we want it to appear, we can paste it back on top of the old data. We don't want formulas, use right mouse button, drag up or down, let go of the mouse button and choose 'Copy Here as Values Only'. Keep the results and you have the data looking the way you want it.

Code #	9 Characters	
A348U8	000A348U8	Move Here
302U302	00302U302	Copy Here
13891E891	13891E891	Copy Here as Values Only
47006F006	47006F006	Copy Here with Formats Only
45615R152	45615R152	Link Here
4503N503	04503N503	Create Hyperlink Here
93458D847	93458D847	Shift Down and Copy
3485B485	03485B485	Shift Right and Copy
51E51	000051E51	Shift Down and Move
		Shift Right and Move
		Shift Eight and Move
		Cancel

Cleaning up Text Data

Cleaning up text data does have the advantage of many different Excel functions called text functions

- Open Ch-05 Workbook, Sheet: **Textfunctions**

- Different techniques are used here for using Text functions. The main issue with this sheet is uppercase.
- Start with column A. The function is proper. Type the following in column B:

Contact	
O'BRIEN, DONNA	=PROPER(A2)
CARSON, BOB	
MCDONALD, MARK	
BAKER, MARK	
HANSEN, SHEILA	

- Copy this down the column. All looks ok, except one exception row 4. "Mcd". You might have to manually override this. There's a flash fill feature that we will try later that will do this more efficiently.
- Column D – is different because it has space issues. E.g. row, 3, 5 and 8.
- To clean these up the function is TRIM. We will include PROPER as we need to sort the capitals.
- See the following:

Name	
BAKER, MARK	=trim(proper(D2))

- Ctrl key + enter (the active cell does not move downward)
- We dealt with the extra spaces in many of those entries, and we uncapitalized the letters
- Column G - we might want to simply capitalize the letters there. Use a function called upper. At the same time, we need to deal with spaces, if we use trim we get extra space:

G	H
Part #	
2w7 145	=upper(trim(G2))

- Instead try the UPPER function and 2 other functions to pull the data together:

G	H	I	Sele
Part #			
2w7 145	=UPPER(left(G2,3)&right(G2,3))		

- Copy down.
- You can see a list of text functions listed at the far right that allow us to deal with data at the character level.

Two new functions: CONCAT and TEXTJOIN to combine data

If you're using Office 365 and you've installed upgrades that were introduced in 2016, you've got access to two new functions in Excel. One is called concat, the other one is called textjoin.

- Open sheet: **CONCAT-TEXTJOIN**

- To see if they are available to you, go to the formulas tab, and you should see them in the category called text. If you're using standard Excel 2016, you will see instead concatenate.

The screenshot shows a Microsoft Excel spreadsheet with data in columns A, B, and C. Column A contains names, column B contains old IDs, and column C contains years. The formula ribbon is open, and the 'Text' function category is selected. Functions like BAHTTEXT, CLEAN, CODE, CONCAT, DOLLAR, EXACT, FIND, FIXED, LEFT, LEN, LOWER, MID, NUMBERVALUE, PROPER, REPLACE, REPT, RIGHT, SEARCH, and TRIM are listed.

Name	Old ID	Y
Alvarado, Sonia	321	2
Anthony, Robert	537	2
Ayala, Polly	146	2
Beck, Craig	437	2
Booker, Judith	282	2
Campos, Richard	973	2
Chang, Thomas	978	2
Cline, Rebecca	119	2
Franklin, Alicia	707	2
Garcia, Brenda	930	2
Gardner, Anthony	571	2
Guzman, Don	723	2
Leach, Jim	793	2
Manning, John	401	2
Mason, Suzanne	473	2014 AL

- In column E – we would like to create a new ID comprised of these three cells, with the existing concatenate function, and you can use this even if you have installed the new functions.
- Using either concat or concatenate formulas, start typing concat and it appears in the list...

A	B	C	D	E
Name	Old ID	Year	State	New ID
Alvarado, Sonia	321	2009	OH	=CONCATENATE(B2,C2,D2)

- Just to compare the 2 formulas. We will try the following:
- In F2, use a formula called formula text so that we have reference. We're simply going to refer to this cell so we can display that actual formula over in F2.

A	B	C	D	E	F
Name	Old ID	Year	State	New ID	
Alvarado, Sonia	321	2009	OH	3212009OH	=FORMULATEXT(E2)

- Back in E column we will try concat – in cell E3, try =concat(B3:D3) Instead of listing individually the range can be highlighted. If you need to remove the middle cell in the range you would remove ‘,’ type the following =concat(B3:D3)
- Across the sheet we will use the new function, one that you can only try if you have the upgrade to Excel 2016 or Office 365 version. We begin in cell L we will concat column I, J, K. We want a space between each one:

I	J	K	L
First	MI	Last	CONCAT
Jeffrey	K	Abrams	=CONCAT(I2, " ", J2, " ", K2)
Karen		Chambers	CONCAT(text1, [text2], [text3], [text4], [text5], ...)

- Copy these down the column
- In some cases where there are no initials in column J, we have spaces appearing between first and last name in our result in column L. This is where TEXTJOIN comes in.

- This new function, textjoin in column M, begins with delimiter, we want to pull this data together with spaces, double-quote, space, double-quote, comma, ignore empty text, so if there's an empty cell, as there is in column J from time to time here, we want to ignore it, and that option is true. Now you can either click true, and tab it into place, that's one option, then pick up the text we want to concat- the 3 cells I2:K2. This will pull together the data from those 3 cells, along with the space between them, but ignore cells that have empty text in them.

I	J	K	L	M
First	MI	Last	CONCAT	TEXTJOIN
Jeffrey	K	Abrams	Jeffrey K Abrams	=TEXTJOIN(" ", t
Karen		Chambers		Ignore empty cell TEXTJOIN(delimiter, ignore_empty, text1, [text2], ...)
Marilyn	E	Eier		(...) FALSE - Include empty cells

- Type TRUE to ignore empty cells

I	J	K	L	M
First	MI	Last	CONCAT	TEXTJOIN
Jeffrey	K	Abrams	Jeffrey K Abrams	=TEXTJOIN(" ", true,I2:K2)

- Copy down column M
- Now see row 3 Karen Chambers, only 1 space between.
- Just display this in column N, using the formula, =FORMULATEXT(M2) so you can view the formula format in column M.
- To make the TEXTJOIN formula more efficient you can take the 'TRUE' out.
- Try choosing FALSE, where you did want to include empty cells, or type '0'. Immediately you can see the space back when you copy formula down. Just undo and go back to the True or space to deal with no spaces.
- In column E if we were to change the format in column E to include a “-”, see the difference between concatenate and textjoin format. TEXTJOIN is simply easier to create:

A	B	C	D	E	F
1	Name	Old ID	Year	State	New ID
2	Alvarado, Sonia	321	2009	OH	=CONCATENATE(B2,"-",C2,"-",D2)
3	Anthony, Robert	537	2014	OR	=TEXTJOIN("-",,B3:D3)
4	Ayala, Polly	146	2014	CA	

Combine data via concatenation

- Open sheet **CombineNames**
- In addition to using functions like CONCAT or CONCATENATE or Text Join to bring together information from different columns, you can also use a feature referred to as concatenation.
- In column D, we need to pull together the data from columns A, B, and C. And we can do this by using the ampersand symbol, that's the and symbol typically found above the number seven key

on most keyboards. We want to bring together this data, and in a particular order. Notice there's a middle initial here. We might have some issues with that.

- Widen column D.

Last Name	First Name	Middle Initial	
Baker	Mark	R	=a2&"&B2&"&C2
Hansen	Sheila	H	
O'Brien	Ted		

- Complete entry, Ctrl+enter
- Copy down the column
- There are trailing spaces in column D now

Last Name	First Name	Middle Initial	
Baker	Mark	R	Baker, Mark R
Hansen	Sheila	H	Hansen, Sheila H
O'Brien	Ted		O'Brien, Ted
Fier	Marilyn		Fier, Marilyn
Morris	Mark	K	Morris, Mark K
Björling	Jussi		Björling, Jussi
Long	Ryan	S	Long, Ryan S
Fitzgerald	Jackie	D	Fitzgerald, Jackie D
Muti	Riccardo	F	Muti, Riccardo F
Eaton	Jeffrey	W	Eaton, Jeffrey W
Chambers	Karen		Chambers, Karen

- So following the first name, that's the B2, instead of putting in a space, because sometimes we will have a middle initial and sometimes not, following this ampersand right here we're going to put in an If function to see if that cell C2 is blank. We're also going to use a function called blank.

Last Name	First Name	Middle Initial	
Baker	Mark	R	=A2&"&B2&if(isblank(C2), "", "&C2&.")
Hansen	Sheila	H	Hansen, Sheila H

- Copy down the column, you should see the period, “.” after the initial every time.
- Like before, we want to throw away the formula but keep the results. Highlight column, drag any edge, left, right, until you bring it right back on to itself in column D, release the right mouse button, “copy here as Values only”.
- Off to the right, not exactly related to cleaning up data, but something you might want to work with numbers like this, e.g. in an invoice where you have a list of Unit Prices and you want to put in a calculation. The display you would like could be one in cell G5. Now, this is simply typed in, we don't want to be doing that all the time, but let's create this same look by using concatenation.
- A simple calculation to multiply the unit price by number is as follows: in cell G2, =E2*F2 To see that along with text, do the following:

- You can put “ “ quotes in for text. To add the currency you need the text function. The Text function says I've got some data, here it is here, it's a numerical bit of information, so let's display this according to a format. If you're familiar with formatting in Excel, the pound sign symbol, when used in other symbols, means suppress the display of leading zeroes. So let's say that sometimes these numbers could go into the tens of thousands, maybe even hundreds of thousands.
- We're putting in an entry here, three pound signs, comma, then two more, and let's say that all these numbers are going to be over a dollar, so we'll put in a zero right here. Possibly we could have some very small numbers here, that might be \$8.75. This format is not exactly obvious at first, it will display the information such that if it were unusually large, we wouldn't have leading zeroes in front of it. So it takes this entry(E2*F2), converts it into text, but displays it in this kind of a format in the process.

	E	F	G	H
1	Unit Price	Number		
2	24.95	45	= "Total Price is: "&text(E2*F2,"###,##0.00")	
3	9.95	20		
4				
5			Total Price is: \$1,122.75	

- Copy it down to the next row:

	E	F	G
1	Unit Price	Number	
2	24.95	45	Total Price is: 1,122.75
3	9.95	20	Total Price is: 199.00
4			
5			Total Price is: \$1,122.75
6			

- One more thing you might want to consider in this format is to put in a dollar sign, select both G2 and G3, press F2 to edit:

	E	F	G
1	Unit Price	Number	
2	24.95	45	= "Total Price is: "&TEXT(E2*F2,"\$###,##0.00")
3	9.95	20	Total Price is: 199.00
4			
5			Total Price is: \$1,122.75

Ctrl+enter

- What happens if the numbers are lower, if the total price is under \$100. Change the number to 5 in cell f3. The entry over to the right, changes to \$49.75 and if you changed the unit in F2 to 545

	E	F	G
1	Unit Price	Number	
2	24.95	545	Total Price is: \$43,597.75
3	9.95	5	Total Price is: \$49.75
4			
5			Total Price is: \$1,122.75
6			

- Like before we could display this formula in the column beside in H.
- =FORMULATEXT(G2)

	E	F	G	H
1	Unit Price	Number		
2	24.95	545	Total Price is: \$13,597.75	= "Total Price is: "&TEXT(E2*F2,"\$###,##0.00")
3	9.95	5	Total Price is: \$49.75	
4				
5			Total Price is: \$1,122.75	

- We will edit the function in cell G3 to make it more efficient, is to replace the function TEXT with DOLLAR. We can get rid of the format that we had ### etc. See below:

	E	F	G	H
1	Unit Price	Number		
2	24.95	545	Total Price is: \$13,597.75	= "Total Price is: "&TEXT(E2*F2,"\$###,##0.00")
3	9.95	5	Total Price is: \$49.75	= "Total Price is: "&DOLLAR(E3*F3)
4				
5			Total Price is: \$1,122.75	

- This has nothing to do, really, with cleaning up data, but it does show us how we can use that ampersand symbol, sometimes referred to as the concatenation symbol, to pull data together.

Split data into columns with the Text to Columns feature

The common problem with data that you get from other sources is too much information in the same column. Now, that's not always bad, and we have to take this on a case by case basis.

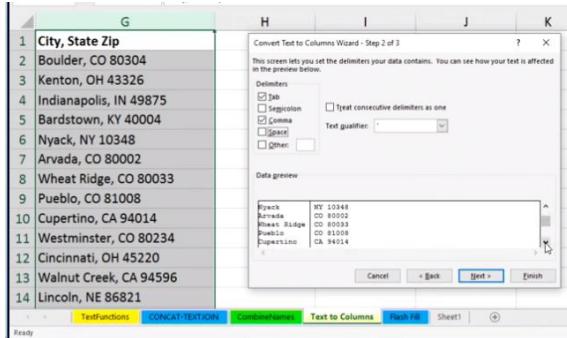
- Open sheet, **Text to Columns**
- In column A, we have some names, in some cases there could be a larger amount of text. We cannot sort this list by name in its current state because the names are in first name, last name order. We could, however, split these into separate columns, and then we could either rejoin them in reverse order, or maybe just leave them in separate columns, then we can sort our list by last name. Text to Columns usually begins with the idea that we probably will need an extra column, typically to the right. In this case, we're going to split these names. Notice that none of them have middle initials, they're all single names and we don't have any Mary Jo's or anything like that, multiple names, just first name, last name.
- Select column A, data tab – in the ribbon, choose Text to Columns, choose delimited, then next. The delimiter in this case, the characters that separates the names is space.

- Check out the preview to see... Click finish and see the data split.
- In column D it's a little bit more problematic, some of the people have middle initials. So we've got at least two extra columns to the right, so that should work okay. But the way these are separated might be a little bit different here.
- Begin my using the "," to split these in column D. It depends on what your looking for, we might have last name first, first name next, then initials...
- Select column D, data tab ribbon – text to columns, this time we use comma
- Check the preview, try comma and space together or space without comma to compare.

- Click finish.

- In column G, we have a bigger list of contact information. We cannot sort this in this state by state or zip code, so we'll split these into different columns.
- Highlight G, Data ribbon, Text to Columns

- If we use space, it might look good at first, but some of those city names are multiple, like Wheat Ridge, San Francisco. Is there a North Las Vegas in there? But if we use space and comma both, it looks like we might have some problems. Let's take out space. Let's choose comma. we can't do everything at once at all times here, but it looks like comma will at least isolate the cities for us. So we're using comma this time, Finish



G	H
1 City	State Zip
2 Boulder	CO 80304
3 Kenton	OH 43326
4 Indianapolis	IN 49875
5 Bardstown	KY 40004
6 Nyack	NY 10348
7 Arvada	CO 80002
8 Wheat Ridge	CO 80033
9 Pueblo	CO 81008
10 Cupertino	CA 94014
11 Westminster	CO 80234
12 Cincinnati	OH 45220
13 Walnut Creek	CA 94596
14 Lincoln	NE 86821

- Our list looks like this, our cities are isolated. Readjust the column width here, let's do the same thing all over again.
- Highlight H column, looks like we have a leading space. Ribbon – Text to Columns – Delimiters – choose space. See preview. Separated the state and zip.

G	H	I	J	K
1 City		State	Zip	
2 Boulder		CO	80304	
3 Kenton		OH	43326	
4 Indianapolis		IN	49875	
5 Bardstown		KY	40004	
6 Nyack		NY	10348	
7 Arvada		CO	80002	
8 Wheat Ridge		CO	80033	
9 Pueblo		CO	81008	
10 Cupertino		CA	94014	
11 Westminster		CO	80234	
12 Cincinnati		OH	45220	
13 Walnut Creek		CA	94596	
14 Lincoln		NE	86821	

- Now we have a blank column H. We can remove it or move over I and J on top of H.

City	State	Zip
Boulder	CO	80304
Kenton	OH	43326
Indianapolis	IN	49875
Bardstown	KY	40004
Nyack	NY	10348
Arvada	CO	80002
Wheat Ridge	CO	80033
Pueblo	CO	81008
Cupertino	CA	94014
Westminster	CO	80234
Cincinnati	OH	45220
Walnut Creek	CA	94596
Lincoln	NE	86821

- We can sort it by zip or state or possibly city first, then by state to get them in alphabetical order within each state. These techniques are really necessary at times.

Flash fill

Flash fill is a relatively new feature in Excel, introduced in Excel 2013. It is a valuable tool, and a great addition to any situation where you need to restructure data that's in columns

You'll find it on the Data tab in the Data Tools group, Flash Fill, it automatically fill in values.

- Open sheet **Flashfill**
- Look at the data in column A
- In Cell B2, If we wanted to have the name, Amy Ryan, formatted with the first letters of each name capitalized, and the order reversed, type in cell B2 “Ryan comma space Amy”.
- One of three different ways to use Flash Fill is to complete one entry, start the next one, starting typing the capital W, the first letter of the last name of the next person, capital W. Excel sees what your doing...

	A	B
1	Contact	
2	AMY RYAN	Ryan, Amy
3	MAX WAGNER	Wagner, Max
4	JACKIE FITZGERALD	Fitzgerald, Jackie
5	SHEILA HANSEN	Hansen, Sheila
6	MARY TODD-JONES	Jones, Mary
7	ERIC O'BRIEN	O'brien, Eric
8	AMY TIDWELL	Tidwell, Amy
9	JO MCDONALD	Mcdonald, Jo
10	ROBERT CATALANO	Catalano, Robert
11	KAREN CHAMBERS	Chambers, Karen
12	JEFF TODD-JONES	Jones, Jeff
13	HARLON HARVEY	Harvey, Harlon
14	ROBERT KONOPKA	Konopka, Robert
15	MARK MORRIS	Morris, Mark
16	DON NICHOLS	Nichols, Don
17	DONNA O'BRIEN	O'brien, Donna
18	DANIELY REBECCA	Rebecca, Daniely

- You can see clearly what's happening. Press enter to accept the entries.
- Notice a few things it did not do. Hyphenated names didn't work here. The B in O'Brien is not capitalized, that occurs later. Everything else seems to be okay. Excel is not yet programmed the first D in McDonald to be capitalized. But in other respects, this works very well.
It's best to either have them in separate columns like D and E, or to have them in reverse order the way we see over in Column B. Column A is a huge disadvantage because we cannot sort the list by last name.
- In column F, cell F2, type Mark Baker, press Control + Enter, so the active cell doesn't move, and then slide up to the feature Flash Fill on the data ribbon, click it. Check out the hyphenated names – it worked perfectly, see row 11. Ted O'Brien in row 4 worked.
- Column H – this is similar to what we did in column A but this time we have issues. You can see the spaces in row 4 too many spaces between Hansen and Sheila. There are leading spaces there.
- In column I, start typing, Ryan, Amy. Use shortcut for flashfill, Ctrl+E
- The spaces are not being dealt with row 4. They are in row 15 with Cathy Watanuki, they are. So perhaps a little bit of unpredictability about how space issues. We're not sure what happens to trailing spaces. Is there one behind Marilyn Fier in row 5? If you check column I the trailing space was taken care of.
- Undo. Drag Mary Todd-Jones up to top of list H (before Ryan, Amy) Start typing Todd-Jones, Mary in I2:

F	G	H	I
		Contact	Todd-Jones, Mary
Mark Baker	MARY TODD-JONES		
Sheila Hansen	RYAN, AMY		
Ted O'Brien	BAKER, MARK		
Marilyn Fier	HANSEN, SHEILA		
Mark Morris	FIER, MARILYN		
Jussi Björling	MORRIS, MARK		
Ryan Long	BJÖRLING, JUSSI		
Jackie Fitzgerald	FITZGERALD, JACKIE		
Riccardo Muti	MUTI, RICCARDO		
Mary Todd-Jones	TIDWELL, LIESL		
Karen Chambers	EATON, JEFFREY		
Barney Perez	CHAMBERS, KAREN		
Cathy Watanuki	PEREZ, BARNEY		
George Porter	WATANUKI, CATHY		
Max Wagner	PORTER, GEORGE		
Robert Konopka	WAGNER, MAX		
Don Nichols	KONOPKA, ROBERT		

TextFunctions CONCAT-TEXTJOIN CombineNames Text to Columns Flash Fill

- Press Ctrl+E or flashfill.
- The result is a disaster – see how it repeats names in an odd way.
- Undo, change the order.
- Drag Amy Ryan to top of the list.
- Retype the entry in I2, Ryan,Amy. Ctrl+Enter Flashfill
- In general it works well, but we shouldn't have spaces.
- In column K- we should never see City, State Zip in same column. You cannot search for either of them attributes.
- Try the following, type in L, M, N:

K	L	M	N	O
1 City, State Zip				
2 Kenton, OH 43326	Kenton	OH	43326	
3 Boulder, CO 80304				
4 San Francisco, CA 94111				
5 Nyack, NY 10348				
6 Indianapolis, IN 49875				
7 Colorado Springs, CO 80033				
8 Bardstown, KY 40004				
9 Arvada, CO 80002				
10 Pueblo, CO 81008				
11 Cupertino, CA 94014				
12 Westminster, CO 80234				
13 Cincinnati, OH 45220				
14 Walnut Creek, CA 94596				
15 Lincoln, NE 86821				
16 Cincinnati, OH 45219				
17				

TextFunctions CONCAT-TEXTJOIN CombineNames Text to Columns

- Click in Kenton, Ctrl+enter, flashfill, it picks up the cities. It's recognized that we've typed everything to the left of the comma, so it does pick up the two city names, like San Francisco, Colorado Springs.
- Flash fill column M and N. So we can get rid of column K
- Column Q- type in the state code CO. Now you recognize in the list to the left, everything is upper case. Will Excel read this as picking up the first two letters of the city name, or will it pick

up the state code? If I complete this with Control + Enter, and press Flash Fill, looks like it's picking up the first two letters of the city name. If you type two rows of States, CO, CO, Excel does not pick up on it in Flashfill.

- Type Arvade in Q2, State CO in R2, in column S, if you were to type the zip Flashfill may get confused and look to the right, so its good to have a blank column to the right

	M	N	O	P	Q	R	S	T	U	V	W
1	State Zip	City, State Zip			State Z Zip		City	State Zip			
2	OH	43326	ARVADA, CO 80002		Arvada	CO	80002	Boulder	CO	80304	
3	CO	80304	KENTON, OH 43326		Kenton	OH	43326	Nyack	NY	10348	
4	CA	94111	COLORADO SPRINGS, CO 80033		Colorado Springs	CO	80033	Colorado Springs	CO	80033	
5	NY	10348	BOULDER, CO 80304		Boulder	CO	80304	Indianapolis	IN	49875	
6	IN	49875	NYACK, NY 10348		Nyack	NY	10348	Kenton	OH	43326	
7	CO	80033	INDIANAPOLIS, IN 49875		Indianapolis	IN	49875	Bardstown	KY	40004	
8	KY	40004	BARDSTOWN, KY 40004		Bardstown	KY	40004	Arvada	CO	80002	
9	CO	80002	PUEBLO, CO 81008		Pueblo	CO	81008	Pueblo	CO	81008	
10	CO	81008	CUPERTINO, CA 94014		Cupertino	CA	94014	Cupertino	CA	94014	
11	CA	94014	WESTMINSTER, CO 80234		Westminster	CO	80234	Westminster	CO	80234	
12	CO	80234	CINCINNATI, OH 45220		Cincinnati	OH	45220	Cincinnati	OH	45220	
13	OH	45220	WALNUT CREEK, CA 94596		Walnut Creek	CA	94596	Walnut Creek	CA	94596	
14	CA	94596	LINCOLN, NE 86821		Lincoln	NE	86821	Lincoln	NE	86821	
15	NE	86821	SAN FRANCISCO, CA 94111		San Francisco	CA	94111	San Francisco	CA	94111	
16	OH	45219	CINCINNATI, OH 45219		Cincinnati	OH	45219				
17											

Fill in Missing Title Information

- Open Ch-06 Workbook, Sheet: **Missing Titles**
- In Column A as you scroll down the list you see a new department name here that applies to the person in column B, and the list of the persons below it. There's a lot of names in some cases. Whether this is considered good or bad the sheet has arrived to you for cleaning. The issue with this is if you wanted to sort by Compensation or Employee Name, you going to lose track of which person is in each department. Try sorting just to see? It can move an employee out of the wrong department. Undo whatever you sorted!
- There's a list of about 25 departments. There are a couple of shortcuts that we can use. Before we try a technique for selecting blank cells, we will deal with column A. Select the cells as follows:

A	B	C	D	E	F	G	H
1 Department	Employee Name	Building	Status	Hire Date	Benefits	Compensation	Job Rating
2 ADC	Anderson, Terry	Taft	Contract	10-Aug-15		59,423	3
3	Bailey, Vic	North	Contract	24-Jun-98		78,988	2
4	Bishop, Juan	South	Half-Time	15-Mar-04	DMR	62,251	1
5	Carr, Susan	West	Half-Time	8-Jan-02	R	49,537	2
6	Hardy, Svetlana	North	Half-Time	5-Mar-04	DM	59,475	5
7 Admin Training	Day, David	North	Contract	9-Sep-02		64,389	4
8	Diaz, David	North	Full Time	22-Nov-11	DMR	95,628	3
9	Frost, Adam	Main	Contract	7-Mar-04		61,776	1
10	Hood, Renee	North	Full Time	6-Oct-07	DMR	37,661	3
11	Hoover, Evangeline	South	Contract	31-Oct-96		103,194	1
12	Malone, Daniel	North	Contract	16-Feb-96		30,342	4
13	McGee, Carol	West	Contract	22-Jul-98		83,759	4
14	Moss, Chan	Main	Hourly	26-Mar-96		19,126	5
15	Noble, Michael	North	Full Time	24-Jun-04	M	60,268	5
16	Randall, Yvonne	Main	Full Time	28-May-06	DMR	93,366	2
17	Rodriguez, Scott	South	Full Time	12-Jul-16	M	89,583	5

- And press Control + enter. All four cells will get data from the cell above.
- Undo, this time, select all of column A and use an unusual feature in Excel found on the Home tab, far right button, Find & Select, Go To Special.

The screenshot shows a Microsoft Excel spreadsheet titled 'CH-06 - Excel'. The table has columns labeled 'Department', 'Employee Name', 'Building', 'Status', 'Hire Date', 'Benefits', 'Compensation', and 'Job Rating'. The 'Compensation' column contains numerical values like 59,423, 78,988, etc. The 'Job Rating' column contains integers like 3, 2, 1, etc. The 'Benefits' column is empty. The 'Blanks' option is selected in the 'Go To Special...' dialog box, which is overlaid on the spreadsheet. The dialog also lists other options like 'Comments', 'Constants', 'Formulas', 'Numbers', 'Text', 'Specials', 'Errors', and 'Blanks'.

- Or instead to get there faster, press function key f5, then 'Go to Special'.

This screenshot is similar to the previous one, showing the 'CH-06 - Excel' spreadsheet with the same table structure. The 'Go To Special...' dialog box is open again, with 'Blanks' selected in the 'Select' section. Other options like 'Comments', 'Constants', 'Formulas', etc., are visible in the list.

- Blanks, ok.
- Just the blank cells get selected in column A down to the last row 742 to check!!
- In cell A3 write a formula =A2 then press Ctrl+enter, which should fill in all the blank cells.

- Not quite finished as most of these are formulas, except the original entries.
- Like before lets copy column A over itself. Select Column A, and now with the right mouse button, I'm simply going to drag this temporarily into Column B and then right back on top of itself. In one smooth motion, drag into Column B then right back into Column A, let go of the right mouse button, Copy Here as Values Only.
- Now we can work with the data. And if we happen to go into Column B and sort by employee name, it keeps its department title in column A.

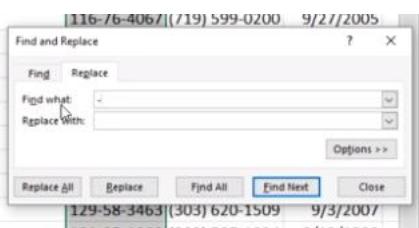
Add phone no's and SS# consistently

When you're trying to match up different lists, you sometimes run into problems if you're trying to compare lists based on either Social Security numbers or phone numbers.

- Open Sheet: **SS# and Phone**
- Column A has Social Security Numbers. And also in column F. Imagine these are different lists and possibly they're on different worksheets, but for ease of use, here, we've put them on the same worksheet.
- Here's what we'd like to be able to do. For the list on the right, we'd like to pick up the employee's name and put it in the list. And we can see that there's a Social Security number in the list on the right, in column F, it's also in the list on the left, column A. We don't necessarily know if every Social Security number that's here is going to be found in the other list. The list on the left is organized by employee name, it's in alphabetical order. The list on the right is in order by Social Security number. If we use the function called match, to see, for example, if this Social Security number in cell F2 is found in the other list.
- The match function allows us to compare data in one location, in this case it's cell F2, with the list over in column A. Widen column E and start typing the following:

EZ	A	B	C	D	E	F
1	SS#	Employee Name	Department		=match(F2,A:A,0)	SS#
2	123-77-7777	Abbott, James	Quality Control			114-86-0179 (30)
3	439-68-8270	Acosta, Robert	Project & Contract Services		MATCH([lookup_value,value,lookup_array,[match_type]])	115-34-2325 (97)
4	714-51-5891	Adkins, Michael	Quality Control			116-54-7612 (30)
5	402-32-5823	Aguilar, Kevin	Quality Assurance			116-59-0210 (30)
6	226-88-4598	Alexander, Charles	Manufacturing			116-76-4067 (71)
7	115-34-2325	Allen, Thomas	Engineering/Maintenance			118-37-9993 (30)
8	942-23-0210	Allison, Timothy	Quality Assurance			121-19-7081 (71)
9	688-78-2634	Alvarado, Sonia	Manufacturing			121-26-6920 (30)
10	785-17-2967	Alvarez, Steven	Quality Control			122-57-8487 (97)
11	191-92-0925	Anderson, Terry	Logistics			123-77-7777 (71)
12	619-07-3294	Andrews, Diane	Operations			125-10-0297 (41)
13	249-46-2796	Anthony, Robert	Quality Control			125-84-1307 (97)
14	116-76-4067	Armstrong, David	Quality Control			129-58-3463 (30)
15	580-67-7315	Arnold, Cole	Manufacturing			131-05-1060 (30)
16	511-16-6939	Atkins, Kevin	Engineering/Maintenance			133-02-4065 (97)
17	608-46-5650	Atkinson, Danielle	Executive Education			133-51-0438 (71)
18	120-77-0411	Austin, William	Quality Assurance			133-50-7251 (71)

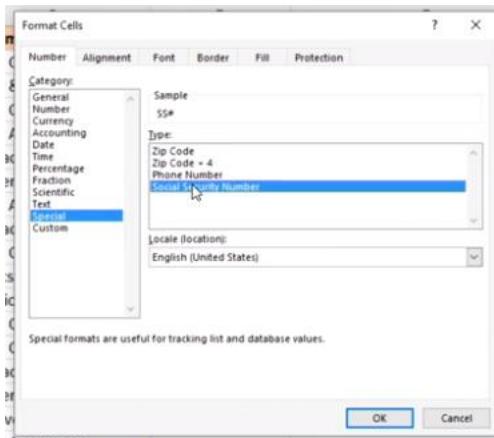
- Note: chose ‘exact match’ hit enter. N/A – that SS# is not found. Double click the lower right hand corner to fill the cells in column E. All cells have N/A- not working. Now you might notice in cell F11 a SS# that is the same as A2 so why didn’t they match up?
- Look in cell A2 and look in formula bar – that is the content 123777777, no dashes and yet dashes appear in cell A2. That’s because this cell and all the cells in column A have been formatted with a special format. Over in column F, click one of these, we see actual dashes up in the formula bar. So the reason we can’t match these up is because cell F12 and the cell over in A2 are not equal to one another.
- Remove the dashes in column F – select column F: Home tab, Find and select (far right), Replace, Replace Dash with nothing.



- They're all done. 1,396 replacements. Look what's happened in the background. The column F has found matches. In column E, the answer in each cell corresponds with the row number.

A	B	C	D	E	F	G
SS#	Employee Name	Department		TRUE	SS#	PI
123-77-7777	Abbott, James	Quality Control			461	114860179 (303)
439-68-8270	Acosta, Robert	Project & Contract Services			7	115342325 (970)
714-51-5891	Adkins, Michael	Quality Control			459	116547612 (303)
402-32-5823	Aguilar, Kevin	Quality Assurance			469	116590210 (303)
226-88-4598	Alexander, Charles	Manufacturing			14	116764067 (719)
115-34-2325	Allen, Thomas	Engineering/Maintenance			215	118379993 (303)
942-23-0210	Allison, Timothy	Quality Assurance			674	121197081 (719)
688-78-2634	Alvarado, Sonia	Manufacturing			270	121266920 (303)
785-17-2967	Alvarez, Steven	Quality Control			501	122578487 (970)
191-92-0925	Anderson, Terry	Logistics			2	123-77-7777 (719)
619-07-3294	Andrews, Diane	Operations			79	125100297 (415)
249-46-2796	Anthony, Robert	Quality Control			100	125841307 (970)
116-76-4067	Armstrong, David	Quality Control			509	129583463 (303)
580-67-7315	Arnold, Cole	Manufacturing			515	131051060 (303)
511-16-6939	Atkins, Kevin	Engineering/Maintenance			684	133024065 (970)
608-46-5650	Atkinson, Danielle	Executive Education			375	133510438 (719)
570-77-0473	Austin, William	Quality Assurance			207	122507253 (719)

- Cell E11 has matched to ‘2’.
- Next stage is to select column F and fix the dashes. – Highlight column F, Ctrl+1 to go to format cells. Number tab, Special, Social Security Number.



- This applies the format, the dashes. Our matches work properly, we're all set.
- You could also use Vlookup.

A	B	C	D	E	F
SS#	Employee Name	Department		TRUE	SS#
123-77-7777	Abbott, James	Quality Control		=vlookup(F2,A:B,2,false)	114-86-0179
439-68-8270	Acosta, Robert	Project & Contract Services		VLOOKUP[lookup_value,table_array,col_index_num,[range_lookup]]	5 (4)
714-51-5891	Adkins, Michael	Quality Control		TRUE - Appropriate match	5-54-7612
402-32-5823	Aguilar, Kevin	Quality Assurance		FALSE - Exact match	469 116-59-0210
226-88-4598	Alexander, Charles	Manufacturing			14 116-76-4067
115-34-2325	Allen, Thomas	Engineering/Maintenance			215 118-37-9993
942-23-0210	Allison, Timothy	Quality Assurance			674 121-19-7081
688-78-2634	Alvarado, Sonia	Manufacturing			270 121-26-6920
785-17-2967	Alvarez, Steven	Quality Control			501 122-57-8487
191-92-0925	Anderson, Terry	Logistics			2 123-77-7777

- You might not be familiar with vlookup, here we indicate what it is we're looking up, this is the location of the table array, over in columns A and B. Then we put in the column number, reading from left to right. This case it's the second column that we're trying to get data out of. Comma, and if it is an exact match, we put in the word false, or you type zero, either way. You could tab in false or type, it doesn't have to be capitalized. Or put in zero, either way. Press enter.

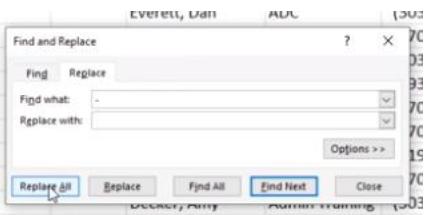
A	B	C	D	E	F
SS#	Employee Name	Department		TRUE	SS#
123-77-7777	Abbott, James	Quality Control		Morrow, Richard	114-86-0179
439-68-8270	Acosta, Robert	Project & Contract Services			7 115-34-2325
714-51-5891	Adkins, Michael	Quality Control			459 116-54-7612
402-32-5823	Aguilar, Kevin	Quality Assurance			469 116-59-0210
226-88-4598	Alexander, Charles	Manufacturing			14 116-76-4067
115-34-2325	Allen, Thomas	Engineering/Maintenance			215 118-37-9993
942-23-0210	Allison, Timothy	Quality Assurance			674 121-19-7081
688-78-2634	Alvarado, Sonia	Manufacturing			270 121-26-6920
785-17-2967	Alvarez, Steven	Quality Control			501 122-57-8487
191-92-0925	Anderson, Terry	Logistics			2 123-77-7777
619-07-3294	Andrews, Diane	Operations			79 125-10-0297

- It finds 'Richard Morrow'. Check to see if it matches?
- Over to the right, suppose we are comparing the list from columns F through I and then M through O. What they have in common is a phone number. They are formatted with a special format. In column O, this has been applied previously. We can jump to Format Cells, Control + One and see that it has been applied.
- If we try to match up these phone numbers, we will run into the same issue as before and ultimately, it's going to be the same kind of solution. We're trying to match up a phone

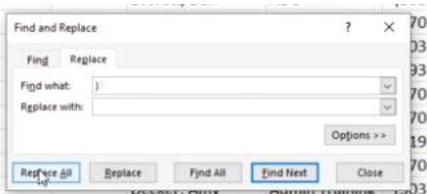
number from G2, comma, comparing it with a list over in column O, comma, we need an exact match, zero. We're not going to get it. Copying down the list, that's not going to work at all.

F	G	H	I	J	K	L
SS#	Phone	Hire Date	Years	Salary		
114-86-0179	(303) 474-3535	11/15/2015	1	116,868	=match(G2,O:O,	MATCH([lookup_value, lookup_array, [match_type]])
115-34-2325	(970) 171-5499	11/19/1995	20	107,935	Finds the largest value that is less than or equal to lookup_value. Lookup_array must be placed in ascending order	1 - Less than 0 - Exact match -1 - Greater than
116-59-0210	(303) 585-8234	4/16/1996	20	98,011		

But if we change the data in column G, We need to get rid of the left parenthesis, the parenthesis and space, and the actual dash. That's going to be three replaces, but it's going to happen pretty fast. Find and Select, again on home ribbon.



- Replace all dashes with nothing.
- Now replace the right parenthesis



- Now all the left parenthesis, repeat like above with left parenthesis.
- See the formulas pop into place in column K.
- Now we can compare...
- Note in column G, the work involved in typing those entries with spaces, () parenthesis. Any change to this could affect it. Be aware any time you're working with phone numbers and Social Security numbers, be sensitive to the idea, they could be entered with actual dashes, and parentheses, and spaces, and so on, but ideally, you want to have those entered strictly as numerical information and then apply a format.
- It's advisable that you set up a format at the beginning from Control + one- format cells, special, phone or whatever your typing.

Remove duplicate rows from a list

- Open sheet **HR List Duplicates**
- You can tell by its name what's involved in this sheet. There are some duplicate records and that's a common problem when you work with lists, particularly if different people have been

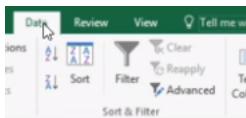
involved in updating the list, copying the list, moving them back and forth and so on. And very often, when you do see a duplicate record you recognise it's adjacent to the current one. Slightly different in this case, a little bit unusual, but here's Susan Carr here in row two, and in row 7, if you hold down the Control key, just for emphasis here, here's another Susan Carr. It's a duplicate record across.

	A	B	C	D	E	F	G	H	I	J	K
1	Employee Name	Building	Department	SS#	Phone	Status	Hire Date	Years	Benefits	Comp.	Job Rating
2	Carr, Susan	West	ADC	914-42-8485	(612) 622-8199	Half-Time	11/27/2006	9	DM	26,795	4
3	Bailey, Victor	West	ADC	948-19-5711	(619) 129-9076	Contract	12/3/2003	12		42,540	5
4	Hardy, Svetlana	Taft	ADC	311-52-6157	(709) 491-9418	Hourly	4/17/1999	17		35,680	2
5	Bishop, Juan	Watson	ADC	991-65-6720	(512) 749-1979	Full Time	5/13/2007	9	M	72,830	2
6	Anderson, Terry	West	ADC	100-43-2924	(415) 585-8234	Full Time	7/27/2001	15	DMR	24,550	1
7	Carr, Susan	West	ADC	914-42-8485	(612) 622-8199	Half-Time	11/27/2006	9	DM	26,795	4
8	Frost, Adam	North	Admin Training	608-79-6012	(309) 752-8456	Full Time	6/17/2016	-	DMR	79,760	5
9	Hood, Renee	Main	Admin Training	202-81-5919	(711) 165-8481	Contract	3/17/2005	11		66,580	5
10	Moss, Chan	Taft	Admin Training	356-11-0882	(907) 252-0526	Half-Time	1/24/2000	16	DMR	15,240	1
11	Vega, Alexandra	Watson	Admin Training	840-31-3216	(717) 184-7141	Full Time	9/5/2004	12	M	37,670	3
12	McGee, Carol	Taft	Admin Training	481-33-6564	(903) 848-8350	Full Time	8/3/2015	1	Q	72,090	5
13	Sullivan, Robert	Main	Admin Training	638-27-1383	(715) 526-7252	Full Time	10/26/2002	14	DMR	49,350	4
14	Malone, Daniel	South	Admin Training	475-25-6935	(511) 224-4880	Full Time	11/18/2001	14	R	85,300	2
15	Wolf, Debbie	West	Admin Training	771-27-7493	(704) 490-0514	Hourly	11/25/2001	14		10,636	4
16	Hoover, Evangeline	North	Admin Training	768-68-1542	(213) 223-8535	Full Time	5/30/1998	18	DMR	60,830	1

- We sorted this list by Employee Name, these would have come together. But the feature to get rid of duplicate rows doesn't require that these be adjacent to one another. If it was just the one record repeated, you could just remove the duplicate row. But the concern is how many others do we have? There's 700 rows there...
- In row 758 and 760 there's Edward Hayes twice.
- Click in column A, data ribbon, AZ in sort and filter and you see some duplicates
- This is not the best way to track them down.
- Click in the data and press Control + A to highlight all the contiguous data. You will need to make sure your list is in a contiguous chunk of cells. When you scroll down you can check. You don't need to do this every single time if you've worked with the list and you know its contiguous.
- Data ribbon, remove duplicates.
- Lets do this for all the columns here. In other words when all the column entries are the same, that's when we want to get rid of the duplicate. Select All. Click ok.

Employee Name	Building	Department	SS#	Phone	Status	Hire Date	Years
Abbott, James	Main	Engineering/Main	513-14-0687	(417) 312-2603	Contract	11/5/2004	12
Acosta, Robert	Main	Project & Contrac	456-80-9622	(619) 320-4992	Full Time	11/21/2010	5
Adams, David	West	Research C				7/16/2002	14
Adams, David	West	Research C				7/16/2002	14
Adkins, Michael	North	Manufactur				2/21/1998	18
Aguilar, Kevin	Main	Manufactur				3/7/1998	18
Alexander, Charles	North	Project & C				2/16/2004	12
Allen, Thomas	West	Process De				6/1/1998	18
Allen, Thomas	West	Project & C				6/1/1998	18
Alvarado, Sonia	Main	Manufactur				4/14/2012	4
Alvarez, Steven	West	Engineering				7/31/2001	15
Anderson, Terry	West	ADC				7/27/2001	15
Andrews, Diane	West	Operations	738-94-6277	(300) 804-6670	Full Time	8/18/2011	5
Anthony, Robert	West	Manufacturing	467-03-0396	(703) 665-7361	Full Time	6/20/2015	1
Armstrong, David	North	Quality Assurance	693-05-5639	(515) 363-1883	Full Time	1/27/2003	13

- The wording might be a little bit strange, 21 duplicate values found and removed, that makes it sound like it's found 21 different cells where there were duplicates, actually it means rows within this data. We've got 738 unique values. Remember there were 760 rows, 759 plus title row, we got rid of 21, so left with 738. Click ok
- Got rid of the duplicate with David Adams in row 4. Check out Thomas Allens in 8 and 9. These are not duplicates if you look across the rows. Control+end and zip to the bottom. You can see the end is row 739.
- There could be times when cleaning up the data, you don't necessarily want to get rid of this, you might want to create a new list, and that would help you identify which records were duplicated. Press Control + Z here.
- Data tab, choose Advanced



- Excel does figure out the extent of the list. if we want to create a new and unique list that does not have duplicates, we're going to be copying this list to another location, unique records only, we don't need a criteria range, but we need to copy it to a location, scroll to the right in cell M1, and simply put its upper left corner in cell M1, and click OK
- The list appearing in the far right of the sheet from M1 position will not have duplicate records in the list, if you zoom back a little bit and scroll right, you can see the new list. Click to the end.
- The difference is we did not disrupt the initial list, we created a new list without duplicates, and sometimes that's the approach you take. I think most people probably prefer, let's just get rid of the bad data and move on, and so the first way we used that feature, removing duplicates, is probably the one used most of the time, and relatively easy to use too, as we saw in this example.

