The Basics of Excel

Part III

Monday, April 17th 2017
D-Lab | University of California, Berkeley

Contents

Last class we learned about…

- Types of functions
- Text functions:
  - Concatenate
  - Extracting substrings of text (mid/left/right)
- Logical functions:
  - If
  - Countif
  - Sumproduct
  - Vlookup
  - Index & Match

Today, we will go over

- Databases
  - Formatting
  - Conditional Formatting
  - Data validation
- Modeling 101 & best practices
- Pivot Tables

Contents

Database format

- To analyze your data, you want to ensure your data is in the database format: a separate column for each field and each row corresponding to each record.

	A	B	C	D	E	F
1	Hospital	Drug	Strength	Size	Dollars	
2	Brownville Hospital - 85242	BRONCHOMED	100MCG/ML	4X1ML	$ 3,111	
3	Brownville Hospital - 85242	BRONCHOMED	200MCG/0.4ML	.4 ML	$ 25,406	
4	Brownville Hospital - 85242	BRONCHOMED	300MCG/0.6ML	.6 ML	$ 5,444	
5	Brownville Hospital - 85242	BRONCHOMED	300MCG/ML	1 ML	$ 36,553	
6	Brownville Hospital - 85242	BRONCHOMED	500MCG/ML	1 ML	$ 15,555	
7	Brownville Hospital - 85242	CONVULSOFIN	200MCG/0.4ML	.4 ML	$ 519	
8	Brownville Hospital - 85242	CONVULSOFIN	500MCG/ML	1 ML	$ 1,297	
9	Carleton Regional Medical Center - 69037	BRONCHOMED	100MCG/0.5ML	4X.5ML	$ 206	
10	Carleton Regional Medical Center - 69037	BRONCHOMED	100MCG/ML	4X1ML	$ 27,307	
11	Carleton Regional Medical Center - 69037	DURANIFIN	10MU/ML	6X2ML	$ 1,056	
12	Carleton Regional Medical Center - 69037	DURANIFIN	40MU/ML	4X1ML	$ 22,452	
13	Carleton Regional Medical Center - 69037	FARMITREXAT	300MCG/ML	10X1ML	$ 7,340	
14	Chilton Regional Hospital - 16013	DURANIFIN	10MU/ML	6X1ML	$ 6,107	
15	Chilton Regional Hospital - 16013	DURANIFIN	20MU/ML	6X1ML	$ 16,380	
16	Chilton Regional Hospital - 16013	DURANIFIN	2MU/ML	6X1ML	$ 647	
17	Chilton Regional Hospital - 16013	DURANIFIN	40MU/ML	4X1ML	$ 7,531	
18	Chilton Regional Hospital - 16013	DURANIFIN	4MU/ML	6X1ML	$ 1,079	

Examples of problem data

- Data that is imported to Excel via a third party application often appears in a layout that requires conversion to database format. Examples include….

	A	B	C	D	E
1					
2		Drug Sales Report			
3		Run Date: 2/3/10			
4	Source: Drugstat				Page 1
5					
6	Hospital	Drug	Strength	Size	Dollars
7	Brownville Hospital - 85242	BRONCHOMED	100MCG/ML	4X1ML	3,110.92
8	Brownville Hospital - 85242	BRONCHOMED	200MCG/0.4ML	.4 ML	25,405.85
9	Brownville Hospital - 85242	BRONCHOMED	300MCG/0.6ML	.6 ML	5,444.28
10	Brownville Hospital - 85242	BRONCHOMED	300MCG/ML	1 ML	36,552.78
11	Brownville Hospital - 85242	BRONCHOMED	500MCG/ML	1 ML	15,554.55
12	Brownville Hospital - 85242	CONVULSOFIN	200MCG/0.4ML	.4 ML	518.59
13	Brownville Hospital - 85242	CONVULSOFIN	500MCG/ML	1 ML	1,296.51
14	Carleton Regional Medical Center - 69037	BRONCHOMED	100MCG/0.5ML	4X.5ML	205.90
15	Carleton Regional Medical Center - 69037	BRONCHOMED	100MCG/ML	4X1ML	27,306.62
16	Carleton Regional Medical Center - 69037	DURANIFIN	10MU/ML	6X2ML	1,056.26
17	Carleton Regional Medical Center - 69037	DURANIFIN	40MU/ML	4X1ML	22,451.53
18	Carleton Regional Medical Center - 69037	FARMITREXAT	300MCG/ML	10X1ML	7,340.26
19	Chilton Regional Hospital - 16013	DURANIFIN	10MU/ML	6X1ML	6,107.36
20	Chilton Regional Hospital - 16013	DURANIFIN	20MU/ML	6X1ML	16,380.14
21					
22		Drug Sales Report			
23		Run Date: 2/3/10			
24	Source: Drugstat				Page 1
25					
26	Hospital	Drug	Strength	Size	Dollars
27	College Park Medical Center - 86045	DURANIFIN	10MU/ML	10X1ML	9,804.81
28	College Park Medical Center - 86045	DURANIFIN	20MU/ML	10X1ML	206.79

Segmented data with repeated headings

Examples of problem data

- Data that is imported to Excel via a third party application often appears in a layout that requires conversion to database format. Examples include….

	A	B	C	D	E
1	Hospital	Generic Name	Strength	Size	Dollar
2	BROWNVILLE HOSPITAL - 85242	BRONCHOMED	100MCG/ML	4X1ML	3,111
3			200MCG/0.4ML	.4 ML	25,406
4			300MCG/0.6ML	.6 ML	5,444
5			300MCG/ML	1 ML	36,553
6			500MCG/ML	1 ML	15,555
7					86,068
8		CONVULSOFIN	200MCG/0.4ML	.4 ML	519
9			500MCG/ML	1 ML	1,297
10					1,815
11					
12	CARLETON REGIONAL MEDICAL CENTER - 69037	BRONCHOMED	100MCG/0.5ML	4X.5ML	206
13			100MCG/ML	4X1ML	27,307
14					27,513
15		DURANIFIN	10MU/ML	6X2ML	1,056
16			40MU/ML	4X1ML	22,452
17					23,508
18		FARMITREXAT	300MCG/ML	10X1ML	7,340
19					7,340
20					
21	CHILTON REGIONAL HOSPITAL - 16013	DURANIFIN	10MU/ML	6X1ML	6,107

Stepped data

Examples of problem data

- Data that is imported to Excel via a third party application often appears in a layout that requires conversion to database format. Examples include….

	A	B	C	D
1	Hospital and Generic Name	Strength	Size	Dollars
2	BROWNVILLE HOSPITAL - 85242			
3	BRONCHOMED	100MCG/ML	4X1ML	3,111
4		200MCG/0.4ML	.4 ML	25,406
5		300MCG/0.6ML	.6 ML	5,444
6		300MCG/ML	1 ML	36,553
7		500MCG/ML	1 ML	15,555
8				86,068
9	CONVULSOFIN	200MCG/0.4ML	.4 ML	519
10		500MCG/ML	1 ML	1,297
11				1,815
12				
13	CARLETON REGIONAL MEDICAL CENTER - 69037			
14	BRONCHOMED	100MCG/0.5ML	4X.5ML	206
15		100MCG/ML	4X1ML	27,307
16				27,513
17	DURANIFIN	10MU/ML	6X2ML	1,056
18		40MU/ML	4X1ML	22,452
19				23,508
20	FARMITREXAT	300MCG/ML	10X1ML	7,340
21				7,340

Stepped data with a twist

Now you try it…

In the Excel exercise file
Go to the **Stepped Data** worksheet:

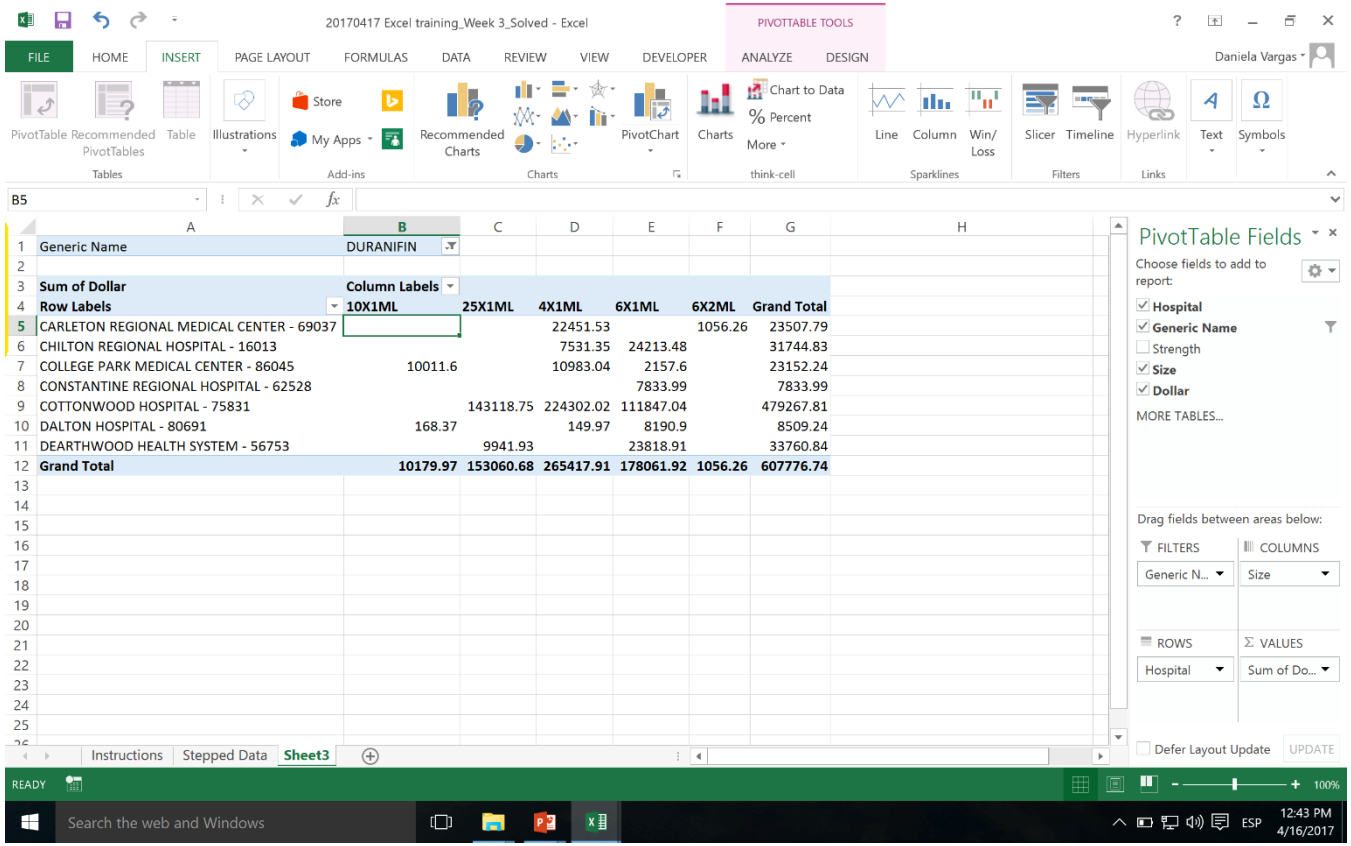
- Highlight cells A7:B95
- Use the F5 method to select blank cells
- Type “=B2” then press CTRL + Enter
- Remove the formulas and remove unneeded rows

Contents

Pivot Tables - allows you to extract the significance from a large, detailed data set

- Being able to analyze all the data in your worksheet can help you better understand it, but sometimes it's hard to know where to start, especially when you have a lot of data

- PivotTables are a great way to summarize, analyze, explore, and present your data



Now you try it…

In the Excel exercise file
Go to the **Stepped Data** worksheet:

- Highlight all the cells in your table
- Go to the Insert tab
- Select Pivot Table
- Answer the questions in the instructions file

Contents

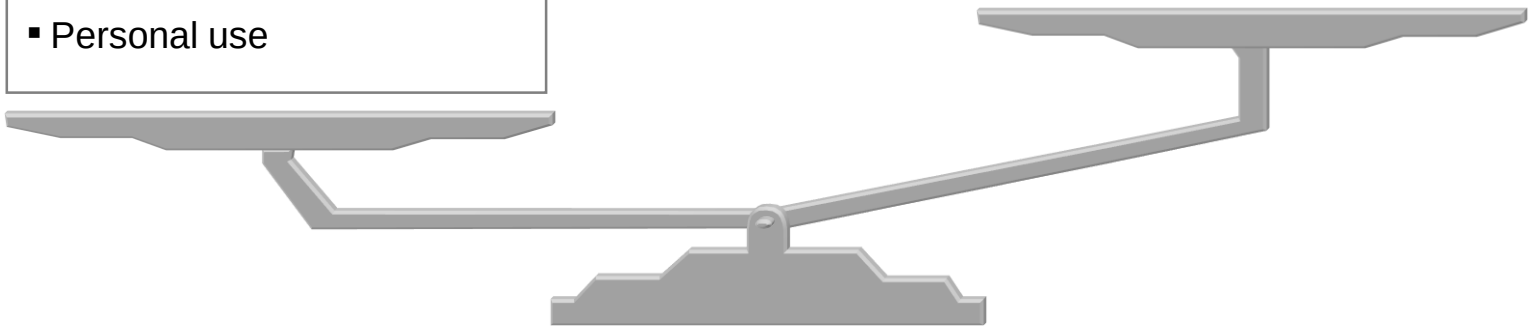
When to create a model

Worksheet

- Relatively simple calculations
- Used once
- Personal use

Model

- Complex calculations
- Possible need for revisions and extensions in the future
- Other people will make use of it and will need to understand the logic

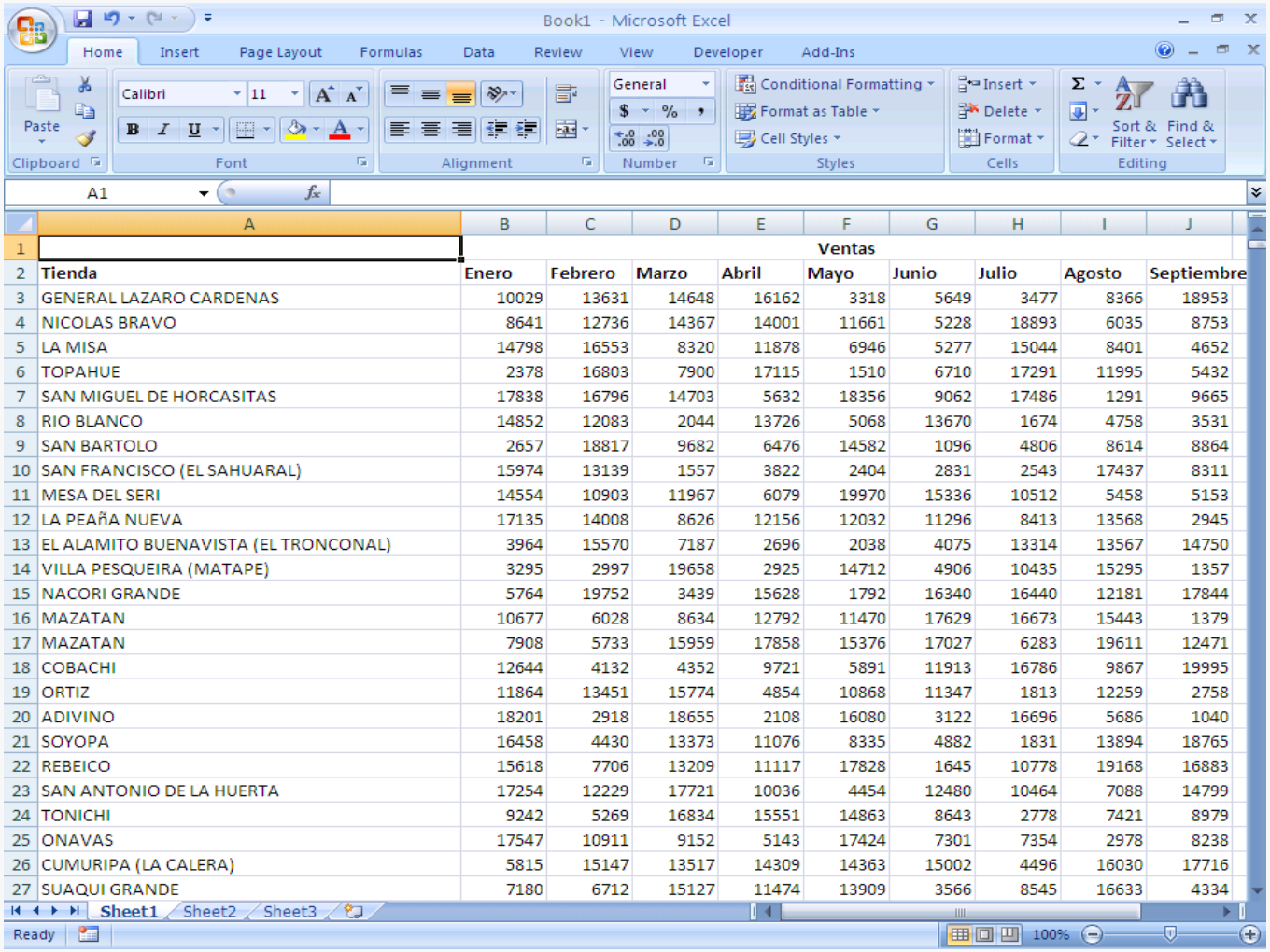


Planning

	Relevant questions
Purpose	▪ What is the purpose? ▪ Who will be the users of the model?
Solution method	▪ Which software to use? ▪ What calculations are necessary? ▪ How many scenarios will be required? Which ones?
Inputs	▪ What information is required? ▪ What are the relevant sources? ▪ How you can deal with missing or incomplete information?
Outputs	▪ What results should the model produce? ▪ How will users interact with the model?

Excel Best practices (1/4)

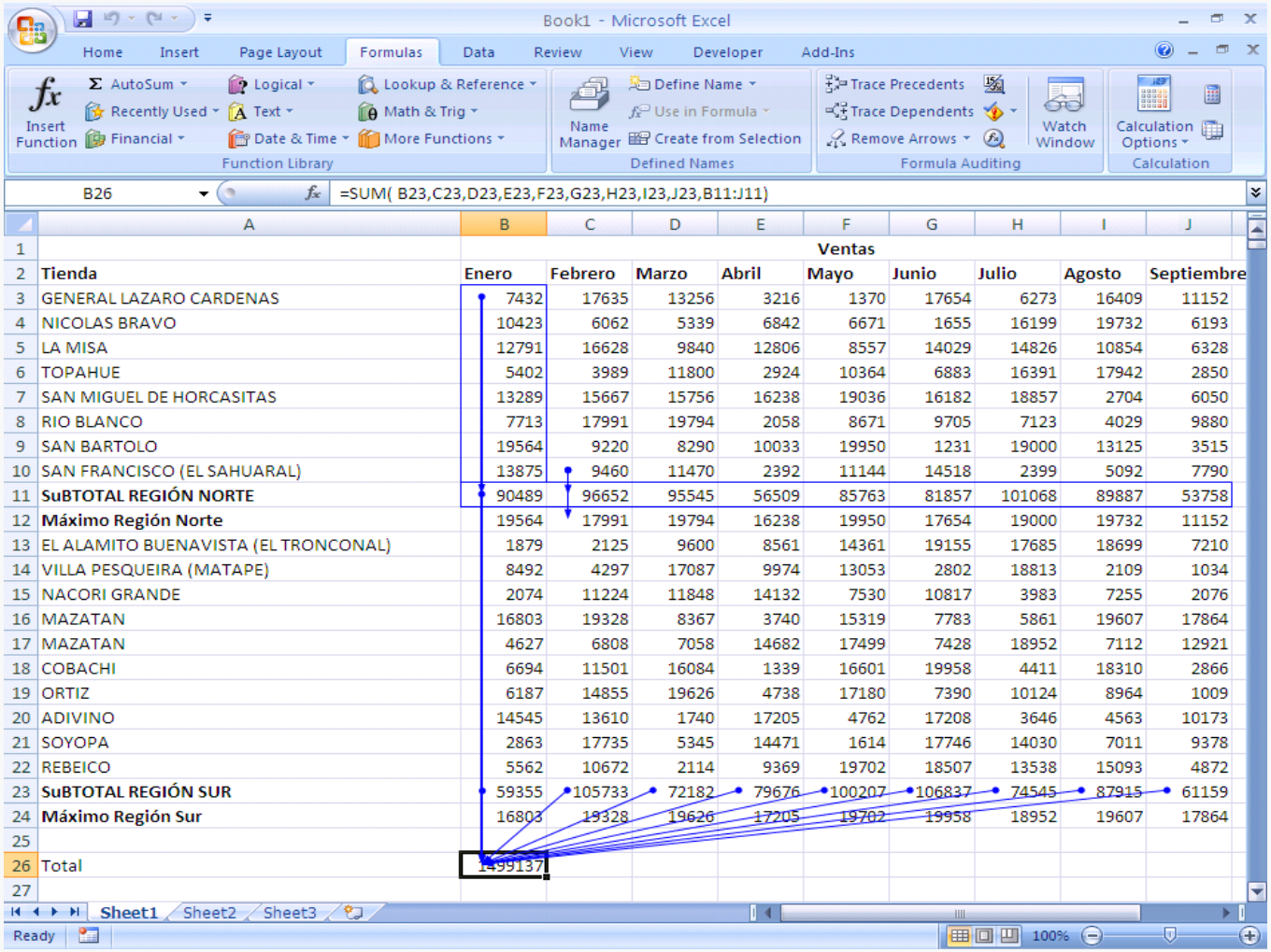
- Organize information vertically



	A	B	C	D	E	F	G	H	I	J
1						Ventas				
2	Tienda	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre
3	GENERAL LAZARO CARDENAS	10029	13631	14648	16162	3318	5649	3477	8366	18953
4	NICOLAS BRAVO	8641	12736	14367	14001	11661	5228	18893	6035	8753
5	LA MISA	14798	16553	8320	11878	6946	5277	15044	8401	4652
6	TOPAHUE	2378	16803	7900	17115	1510	6710	17291	11995	5432
7	SAN MIGUEL DE HORCASITAS	17838	16796	14703	5632	18356	9062	17486	1291	9665
8	RIO BLANCO	14852	12083	2044	13726	5068	13670	1674	4758	3531
9	SAN BARTOLO	2657	18817	9682	6476	14582	1096	4806	8614	8864
10	SAN FRANCISCO (EL SAHUARAL)	15974	13139	1557	3822	2404	2831	2543	17437	8311
11	MESA DEL SERI	14554	10903	11967	6079	19970	15336	10512	5458	5153
12	LA PEAÑA NUEVA	17135	14008	8626	12156	12032	11296	8413	13568	2945
13	EL ALAMITO BUENAVISTA (EL TRONCONAL)	3964	15570	7187	2696	2038	4075	13314	13567	14750
14	VILLA PESQUEIRA (MATAPE)	3295	2997	19658	2925	14712	4906	10435	15295	1357
15	NACORI GRANDE	5764	19752	3439	15628	1792	16340	16440	12181	17844
16	MAZATAN	10677	6028	8634	12792	11470	17629	16673	15443	1379
17	MAZATAN	7908	5733	15959	17858	15376	17027	6283	19611	12471
18	COBACHI	12644	4132	4352	9721	5891	11913	16786	9867	19995
19	ORTIZ	11864	13451	15774	4854	10868	11347	1813	12259	2758
20	ADIVINO	18201	2918	18655	2108	16080	3122	16696	5686	1040
21	SOYOPA	16458	4430	13373	11076	8335	4882	1831	13894	18765
22	REBEICO	15618	7706	13209	11117	17828	1645	10778	19168	16883
23	SAN ANTONIO DE LA HUERTA	17254	12229	17721	10036	4454	12480	10464	7088	14799
24	TONICHI	9242	5269	16834	15551	14863	8643	2778	7421	8979
25	ONAVAS	17547	10911	9152	5143	17424	7301	7354	2978	8238
26	CUMURIPA (LA CALERA)	5815	15147	13517	14309	14363	15002	4496	16030	17716
27	SUAQUI GRANDE	7180	6712	15127	11474	13909	3566	8545	16633	4334

Excel Best practices (2/4)

- A formula should only reference cells above it



Excel Best practices (3/4)

- Formulas should be in the simplest possible form, to prevent unnecessary calculations:

= (A1*1000+B1*1000+C1*1000)/10

= (A1+B1+C1)*100

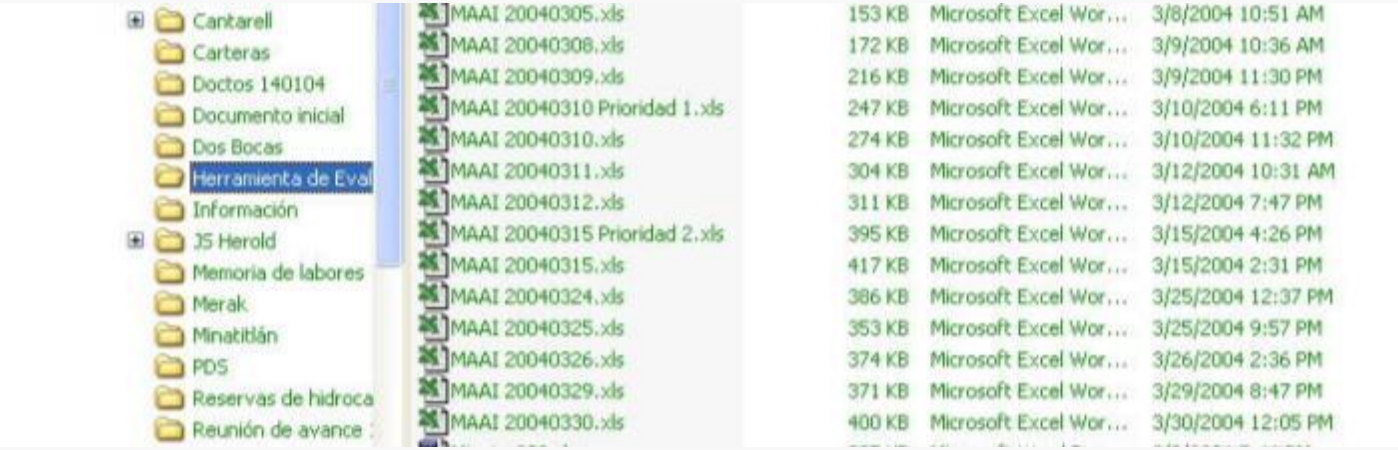
- Setting parameters is preferable to inputting numbers directly into the formula, since it will later be easier to make changes and add comments to these parameters

= (A1+B1+C1)*100

= (A1+B1+C1)*D5

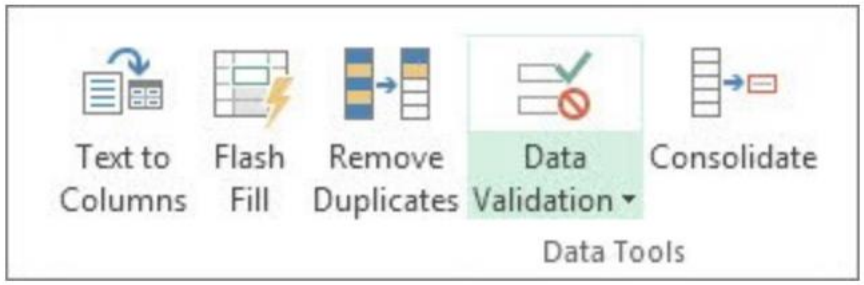
Excel Best practices (4/4)

- Organize the spreadsheet forecasting potential changes
- Do not waste much time on formatting until the spreadsheet is final
- Save information frequently but carefully: do NOT rely on Auto Save
  - It is best to store files with consecutive names to keep working (file01.xls, file02.xls, file03.xls) and then erase earlier versions, than to find out you made irreparable changes to your file!

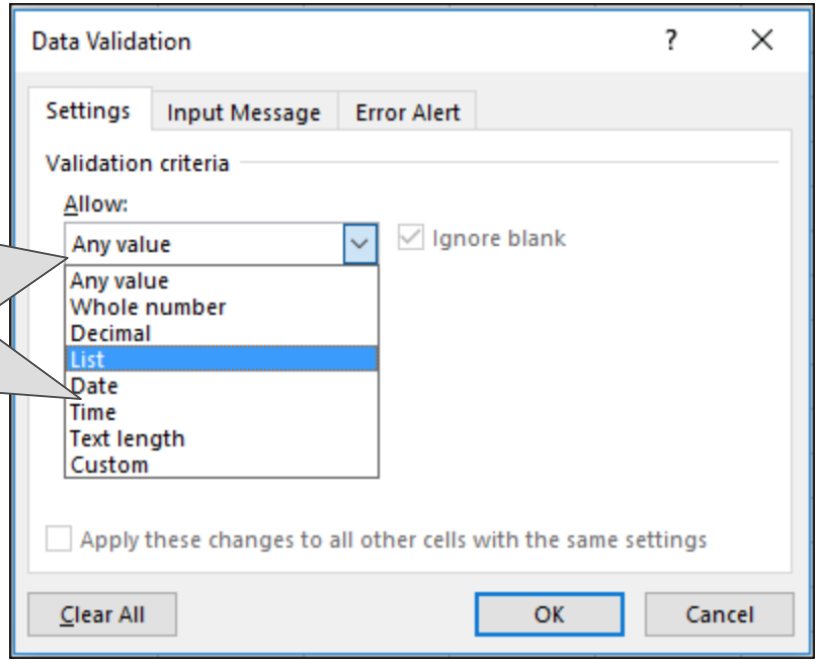


Contents

Data validation - restrict the type of data or the values that users enter into a cell

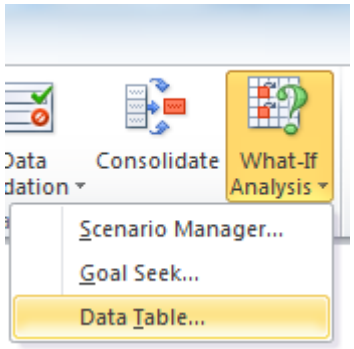


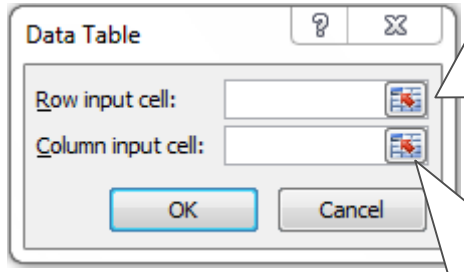
One of the most common data validation uses is to create a drop-down list



Contents

Sensitivity Analysis or "data table" - range of cells in which you can change values in some in some of the cells and come up with different answers to a problem





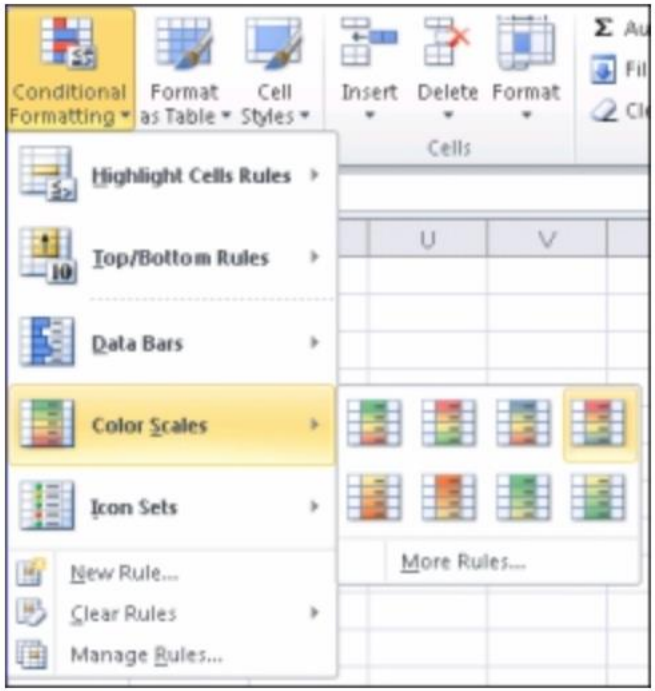
The "Row input cell" is the cell where you would enter the different values on the x axis of your table

The "Column input cell" is the cell where you would enter the different values on the y axis of your table

You need to have the top left cell on your table be a reference to the value you want to play with

Contents

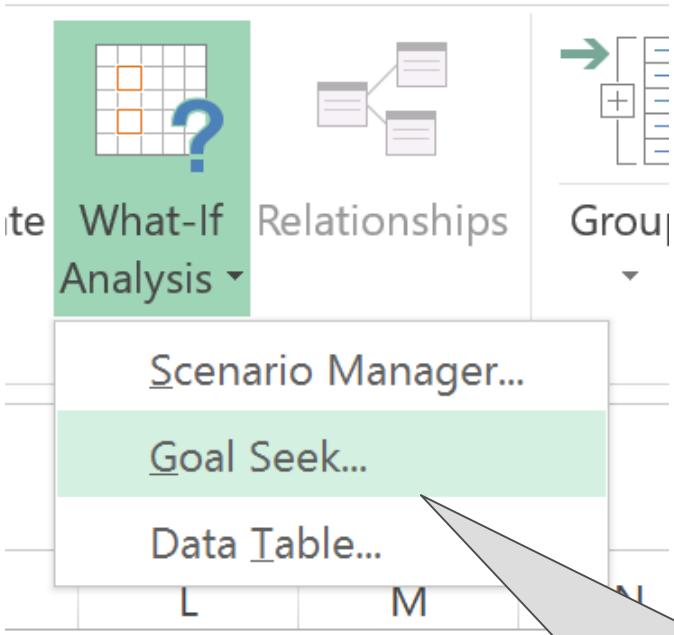
Conditional Formatting - quickly identify variances in a range of values with a quick glance

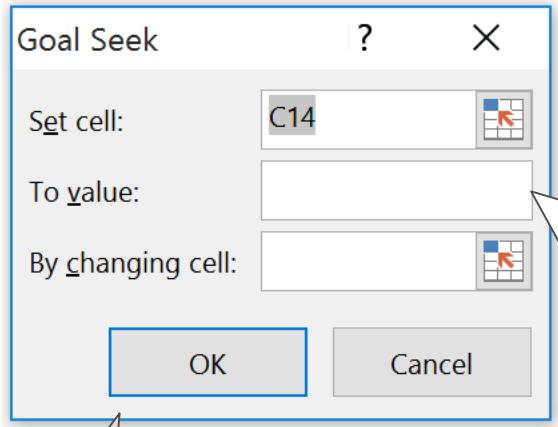


	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2	Avg High	40	38	44	46	51	56	67	72	70	59	45	41
3	Avg Low	34	33	38	41	45	48	51	55	54	45	41	38
4	Record High	61	69	79	83	95	97	100	101	94	87	72	66
5	Record Low	0	2	9	24	28	32	36	39	35	21	12	4

Contents

Goal Seek - find the result you want by adjusting an input value





You need to enter a manual value here

If you know the result you want from a formula, but aren't sure of which input value will producte that result, use Goal Seek