

Microsoft Excel

In our first set of labs we will be getting familiar with Microsoft Excel, a tool used by many organisations to evaluate daily performance and to make critical strategic and operational decisions.

Spreadsheet software for personal computers has become an indispensable tool for business analysis, particularly for the manipulation of numerical data and the development and analysis of decision models. Please note that Mac versions of Excel do not have the full functionality that Windows versions have.

Although Excel has some flaws and limitations from a statistical perspective, its widespread availability makes it the software of choice for many business professionals.

This first lab will briefly review some of the fundamental skills needed to use Excel. This is not meant to be a complete tutorial; many good Excel tutorials can be found online.

The Basics

This lab assumes you are familiar with the most elementary spreadsheet concepts and procedures:

- Opening, saving and printing files.
- Moving around a spreadsheet.
- Selecting ranges.
- Inserting/deleting rows and columns
- Entering and editing text, numerical data, and formulas.
- Formatting data.
- Working with text strings.
- Performing basic arithmetic calculations.
- Modifying the appearance of the spreadsheet.
- Sorting data

Functions

We will begin this lab by mentioning some important functions that are applicable to analysis. We will then use the Lookup function, see how filtering data is useful and end the lab with an example of using pivot tables on categorical data.

Functions are used to perform special calculations in cells. Some of the more common functions used in statistical applications include:

- MIN(range)
- MAX(range)
- SUM(range)
- AVERAGE(range)
- COUNT(range)
- COUNTIF(range, criteria)
- AND(condition1, condition2...) returns true if all conditions are true and false if not.
- OR(condition1, condition2...) returns true if any condition is true and false if not.
- IF(condition, value if true, value if false) returns one value if the condition is true and another if it is false.
- VLOOKUP(value, table range, column number) looks up a value in a table.

The easiest way to locate a particular function is to select a cell and click on the Insert function button on the ribbon. This is useful when you know the name of the function but are unsure of the arguments.

VLOOKUP Function

In this lab, we will read in a data file and prepare it for its use in subsequent tasks.

Download [Summer Olympic medals 1976 to 2008](#). This file consists of all the medals received from the Montreal 1976 Olympics to Beijing 2008 Olympics.

The file contains:

	A	B	C	D	E	F	G	H	I	J
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Code	Event_gender	Medal
2	Sydney	2000	Cycling	Cycling Road	individual time trial	HOLDEN, Mari	Women	USA	W	Silver
3	Atlanta	1996	Judo	Judo	48 - 52kg (half-lightweight)	RESTOUX, Marie-Claire	Women	FRA	W	Gold
4	Athens	2004	Boxing	Boxing	54 - 57kg (featherweight)	TISHCHENKO, Alexey	Men	RUS	M	Gold
5	Montreal	1976	Athletics	Athletics	4x400m relay	STROPAHL-STREIDT, Ellen	Women	GDR	W	Gold
6	Barcelona	1992	Canoe / Kayaking	Canoe / Kayaking	C-2 1000m (canoe double)	FREDERIKSEN, Christian Wigum	Men	DEN	M	Silver
7	Seoul	1988	Volleyball	Volleyball	volleyball	ZHENG, Mei-Zhu	Women	CHN	W	Bronze
8	Moscow	1980	Aquatics	Water polo	water polo	ROJE, Zoran	Men	YUG	M	Silver
9	Atlanta	1996	Boxing	Boxing	51 - 54kg (bantamweight)	MESA, Arnaldo	Men	CUB	M	Silver
10	Atlanta	1996	Hockey	Hockey	hockey	LOMANS, Bram	Men	NED	M	Gold
11	Atlanta	1996	Badminton	Badminton	doubles	TANG, Yingshu	Women	CHN	W	Bronze
12	Montreal	1976	Boxing	Boxing	63.5 - 67kg (welterweight)	GAMARRO, Pedro José	Men	VEN	M	Silver
13	Beijing	2008	Athletics	Athletics	100m	SIMPSON, Sherone	Women	JAM	W	Silver
14	Beijing	2008	Sailing	Sailing	470 - Two Person Dinghy	WILMOT, Nathan	Men	AUS	M	Gold
15	Athens	2004	Volleyball	Volleyball	volleyball	SANCHEZ SALFRAN, Martha	Women	CUB	W	Bronze
16	Atlanta	1996	Badminton	Badminton	singles	SIDEK, Rashid	Men	MAS	M	Bronze
17	Atlanta	1996	Wrestling	Wrestling	Freestyle 62 - 68kg (lightweight)	ZOZIROV, Zaza	Men	UKR	M	Bronze
18	Atlanta	1996	Softball	Softball	softball	WANG, Ying	Women	CHN	W	Silver
19	Atlanta	1996	Softball	Softball	softball	HOLLIDAY, Jennifer	Women	AUS	W	Bronze
20	Los Angeles	1984	Cycling	Cycling Road	individual road race	SCHUMACHER, Sandra	Women	FRG	W	Bronze
21	Atlanta	1996	Fencing	Fencing	foil team	KRZESINSKI, Adam	Men	POL	M	Silver
22	Moscow	1980	Basketball	Basketball	basketball	VECCHIATO, Renzo	Men	ITA	M	Silver

- Column A is the city where the summer Olympics were held.
- B: The year they were held in.
- C: The particular sport to which the medal belongs to.
- D: The discipline within the sport for example, the sport aquatics may consist of diving, swimming, water polo and so on.
- E: The specific event for the medal.
- F: The person who won the medal.
- G: The gender of the person who won the medal.
- H: The country to which the person belongs, the country is given in terms of three-letter codes.
- I: The event gender, that is, that particular sport, that particular event, was it a mens' event, and womens' event, or was it a gender neutral event.
- J: And finally the medals; gold, silver, bronze.

Sort

- This file is not sorted in any particular order.
- Select the entire first row then do a Shift+Ctrl down arrow.
- And it selects the entire file. Notice this is a big file, spanning across 15,317 rows.
- Go to data ribbon, sort.
- Notice headers as my first row, so make sure headers box is checked in top right corner.
- Sort by year for first level, smallest to largest.
- Add level, sort by sport – A to Z
- Add level, discipline – A to Z. Click ok

	A	B	C	D	E	F	G	H	I	J
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Code	Event_gender	Medal
2	Sydney	2000	Cycling	Cycling Road	individual time trial	HOLDEN, Mari	Women	USA	W	Silver
3	Atlanta	1996	Judo	Judo	48 - 52kg (half-lightwe	RESTOUX, Marie-Claire	Women	FRA	W	Gold
4	Athens	2004	Boxing	Boxing	54 - 57kg (featherweig	TISHCHENKO, Alexey	Men	RUS	M	Gold
5	Montreal	1976	Athletics	Athletics	4x400m relay	STROPAHL-STREIDT, Ellen	Women	GDR	W	Gold
6	Barcelona	1992	Canoe / K	Canoe / K	C-2 1000m (canoe do	FREDERIKSEN, Christian Wige	Men	DEN	M	Silver
7	Seoul	1988					Women	CHN	W	Bronze
8	Moscow	1980					Men	YUG	M	Silver
9	Atlanta	1996					Men	CUB	M	Silver
10	Atlanta	1996					Men	NED	M	Gold
11	Atlanta	1996					Women	CHN	W	Bronze
12	Montreal	1976					Men	VEN	M	Silver
13	Beijing	2008					Women	JAM	W	Silver
14	Beijing	2008					Men	AUS	M	Gold
15	Athens	2004					Women	CUB	W	Bronze
16	Atlanta	1996	Badminton	Badminton	singles	SIDEK, Rashid	Men	MAS	M	Bronze
17	Atlanta	1996	Wrestling	Wrestling Fr	62 - 68kg (lightweight)	ZOIROV, Zaza	Men	UKR	M	Bronze
18	Atlanta	1996	Softball	Softball	softball	WANG, Ying	Women	CHN	W	Silver
19	Atlanta	1996	Softball	Softball	softball	HOLLIDAY, Jennifer	Women	AUS	W	Bronze

This file if you look at the bottom left-hand side has more than one worksheet. One worksheet, which is currently open, called Medalists, the other worksheet is Country Codes.

The advantage of having multiple worksheets is that it helps organize your data. For example, if you run a business, one worksheet could have all your sales data. The other worksheet could have your inventory data, the third worksheet could have your cost data, and so on.

The second worksheet that this file has is called Country Codes. This worksheet contains the three letter Country Code and the corresponding country.

We'll use this data to update data in our first worksheet, which is called Medalists. Remember, with multiple worksheets, you can do calculations across worksheets. That is a function or a command in one particular worksheet can access data in another worksheet.

	A	B
1	Country_Code	Country
2	AFG	Afghanistan
3	AHO	Netherlands Antilles*
4	ALB	Albania
5	ALG	Algeria
6	AND	Andorra
7	ANG	Angola
8	ANT	Antigua and Barbuda
9	ARG	Argentina
10	ARM	Armenia
11	ARU	Aruba*
12	ASA	American Samoa*
13	AUS	Australia
14	AUT	Austria
15	AZE	Azerbaijan
16	BAH	Bahamas
17	BAN	Bangladesh
18	BAR	Barbados
19	BDI	Burundi
20	BEL	Belgium
21	BEN	Benin
22	BER	Bermuda*
23	BHU	Bhutan
24	BIH	Bosnia and Herzegovina
25	BIZ	Belize
26	BLR	Belarus

- Open the Medalists worksheet. We need to replace this Country Code with the actual country.
- Insert a column here and call it Country.
- Use a VLOOKUP function in Excel to look up the corresponding country, which corresponds to this 3 letter code.
- In cell I2 under country column in Medalist worksheet, type =VLOOKUP, click H2 for country cell, select the range A1 to B211 in country codes tab, choose column 2 and then false

Description:

In cell I2, type equals, VLOOKUP, open parentheses, we wish to look up the Country Code, comma, where do I want to look it up (table array)? It's in the next worksheet. Remember, functions and commands can work across worksheets. Go to Country Codes, select this data. The first column is the three letter Country Code. Next column as the corresponding country. Add a comma, once my three letter code from the worksheet Medallist is matched up to this look up table. We need to read the next column. So I give a number two there because its the second column needs to be accessed. We want a fixed reference that is an exact match. Hence I give a false there.

=VLOOKUP(H2,'Country Codes'!A2:B211, 2,FALSE)

	A	B	C	D	E	F	G	H	I	J	K
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Code	Country	Event_gender	Medal
2	Montreal	1976	Aquatics	Diving	3m springboard	KÖHLER, Christa	Women	GDR	East Germany	W	Silver
3	Montreal	1976	Aquatics	Diving	3m springboard	KOSENKOV, Aleksandr	Men	URS		M	Bronze

- Since we will be copying and pasting this formula we want a fixed referencing to look up this table.
- Insert dollar signs (absolute referencing) in my range reference and my formula is ready.
- =VLOOKUP(H2,'Country Codes'!\$A\$2:\$B\$211, 2,FALSE)
- The very first country is Germany.
- Copy and paste it all the way down.

	A	B	C	D	E	F	G	H	I	J	K
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Code	Country	Event_gender	Medal
2	Montreal	1976	Aquatics	Diving	3m springboard	KÖHLER, Christa	Women	GDR	East Germany	W	Silver
3	Montreal	1976	Aquatics	Diving	3m springboard	KOSENKOV, Aleksandr	Men	URS	Soviet Union	M	Bronze
4	Montreal	1976	Aquatics	Diving	3m springboard	BOGGS, Philip George	Men	USA	United States	M	Gold
5	Montreal	1976	Aquatics	Diving	3m springboard	CAGNOTTO, Giorgio Franco	Men	ITA	Italy	M	Silver
6	Montreal	1976	Aquatics	Diving	10m platform	WILSON, Deborah Keplar	Women	USA	United States	W	Bronze
7	Montreal	1976	Aquatics	Diving	10m platform	LOUGANIS, Gregory	Men	USA	United States	M	Silver
8	Montreal	1976	Aquatics	Diving	10m platform	VAYTSEKHOVSKAYA, Elena	Women	URS	Soviet Union	W	Gold
9	Montreal	1976	Aquatics	Diving	3m springboard	POTTER-MCINGVALE, Cynthia	Women	USA	United States	W	Bronze
10	Montreal	1976	Aquatics	Diving	10m platform	DIBIASI, Klaus	Men	ITA	Italy	M	Gold
11	Montreal	1976	Aquatics	Diving	10m platform	ALEINIK, Vladimir	Men	URS	Soviet Union	M	Bronze

This file is now ready for further analysis, using the filters and pivot tables functionality of Excel.

Data Filtering

We will continue with the same data, Olympic medal winners to apply filters on this file.

- Put cursor on spreadsheet
- Click Data ribbon and choose filter icon.

	A	B	C	D	E	F	G	H	I	J	K
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Code	Country	Event_gender	Medal
2	Montreal	1976	Aquatics	Diving	3m springboard	KÖHLER, Christa	Women	GDR	East Germany	W	Silver
3	Montreal	1976	Aquatics	Diving	3m springboard	KOSENKOV, Aleksandr	Men	URS	Soviet Union	M	Bronze
4	Montreal	1976	Aquatics	Diving	3m springboard	BOGGS, Philip George	Men	USA	United States	M	Gold
5	Montreal	1976	Aquatics	Diving	3m springboard	CAGNOTTO, Giorgio Franco	Men	ITA	Italy	M	Silver
6	Montreal	1976	Aquatics	Diving	10m platform	WILSON, Deborah Keplar	Women	USA	United States	W	Bronze

- Drop down boxes have been placed in the columns.
- For example, if you click on drop down box for the column Year, it contains all possible values of years that occur in that column.
- Similarly, if you click on the drop down box against City, all the possible cities that occur in that column. If you click on sport, all possible sports that occur in that column and so on.
- How do I use the filters – if you want to see the medals won by Japan. Go to country column, click the drop down box, Click on the Select All cell there, it deselects all the countries. Now select Japan.
- Notice the small icon at filter box, so you can see filter has been applied.

	A	B	C	D	E	F	G	H	I
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Cod	Country
163	Montreal	1976	Archery	Archery	individual FITA round	MICHINAGA, Hiroshi	Men	JPN	Japan
187	Montreal	1976	Gymnastic Artistic G.		team competition	KAJIYAMA, Hiroshi	Men	JPN	Japan
189	Montreal	1976	Gymnastic Artistic G.		team competition	KENMOTSU, Eizo	Men	JPN	Japan
194	Montreal	1976	Gymnastic Artistic G.		parallel bars	KATO, Sawao	Men	JPN	Japan
195	Montreal	1976	Gymnastic Artistic G.		team competition	FUJIMOTO, Shun	Men	JPN	Japan
197	Montreal	1976	Gymnastic Artistic G.		pommel horse	KENMOTSU, Eizo	Men	JPN	Japan
198	Montreal	1976	Gymnastic Artistic G.		individual all-round	TSUKAHARA, Mitsuo	Men	JPN	Japan
11	Montreal	1976	Gymnastic Artistic G.		vault	KAJIYAMA, Hiroshi	Men	JPN	Japan

TASK

- Continuing here, now filter on Sydney Olympics... what do you select?
- Year – 2000
- Further filter you can look at gold medals... how many??

Result:

	A	B	C	D	E	F	G	H	I	J	K
1	City	Year	Sport	Discipline	Event	Athlete	Gender	Country_Cod	Country	Event_gender	Medal
9676	Sydney	2000	Athletics	Athletics	marathon	TAKAHASHI, Naoko	Women	JPN	Japan	W	Gold
10683	Sydney	2000	Judo	Judo	73 - 81kg (half-middle)	TAKIMOTO, Makoto	Men	JPN	Japan	M	Gold
10690	Sydney	2000	Judo	Judo	- 48kg (extra-lightweig	TANI, Ryoko	Women	JPN	Japan	W	Gold
10691	Sydney	2000	Judo	Judo	90 - 100kg (half-heavy	INOUE, Kosei	Men	JPN	Japan	M	Gold
10696	Sydney	2000	Judo	Judo	- 60 kg	NOMURA, Tadahiro	Men	JPN	Japan	M	Gold
15318											
15319											
15320											
15321											

- What about the women who won gold medals for Japan? How many?
- The data is still there, it's just hidden. To unfilter – go to data and click clear. It will clear all filtering.

TASK

Which country won maximum medals gold medals in boxing in Sydney 2000 Olympics?

Pivot tables

A pivot table is a program tool that allows you to reorganize and summarize selected columns and rows of data in a spreadsheet or database table to obtain a desired report. A pivot table doesn't actually change the spreadsheet or database itself. It's a very powerful data summarisation tool. It's a form of reports and online analytical processing (OLAP).

Be aware they are dynamic, so each time the data changes in your data file, the summary gets updated.

Using the same file, Olympics data. We will use Pivot tables to get some useful summaries from the data.

Answer the following question: "I wish to find out how many medals, bronze, silver, gold medals, have each of these countries won across all years?"

- Insert the Pivot Table (Insert menu)
- Make sure all data selected.
- Choose new worksheet
- The Pivot table is placed on the worksheet. You can see the fields on your right.
- You can drag fields to the boxes columns, rows underneath the field list.
- Follow this...

The screenshot shows an Excel spreadsheet with a PivotTable summarizing Olympic medal data. The PivotTable is located in the range E4:U37. The PivotTable Field List task pane is visible on the right side of the screen.

PivotTable Data:

Row Labels	Bronze	Silver	Gold	Grand Total
Alghanistan	1			1
Algeria	8	2	4	14
Argentina	70	37	46	153
Armenia	7	1	1	9
Australia	312	270	216	798
Austria	8	17	9	34
Azerbaijan	9	3	4	16
Bahamas	3	9	7	19
Barbados	1			1
Belarus	53	25	14	92
Belgium	20	15	6	41
Bermuda*	1			1
Brazil	125	137	56	318
Bulgaria	107	100	40	247
Burundi	1			1
Cameroon	1		20	21
Canada	117	111	76	304
Chile	19	2	3	24
China	193	252	234	679
Colombia	5	2	1	8
Costa Rica	2	1	1	4
Cote d'Ivoire		1		1
Croatia	18	30	31	79
Cuba	88	109	152	349
Czech Republic	13	18	10	41
Czechoslovakia	29	29	27	85
Denmark	50	20	78	148
Djibouti	1			1
Dominican Republic	1	1	2	4
East Germany	150	190	286	626
Ecuador		1	1	2
Egypt	4	2	1	7

PivotTable Field List:

- Choose fields to add to report: City, Year, Sport, Discipline, Event, Athlete, Gender, Country_Code, ☒ Country, Event_gender, ☒ Medal
- Drag fields between areas below:
 - Report Filter: Medal
 - Row Labels: Country
 - Values: Count of Medal
- Buttons: Defer Layout Update, Update

- In this example, Canada has won a total of 304 medals over this period of time, broken up into bronze, gold, silver as per the numbers given.
- One thing to keep in mind is that if there is a team sport, and that team sport consists of, let's say, 11 members, 10 members and so on, and if a country won the medal in that team sport, then all the individual members of the team are listed as medal winners in this file.

Sort

- You can sort the column labels by clicking outside of pivot table which disables the right hand side pivot table options.

1					
2	Year	2000			
3					
4	<div> <div>Column Labels</div> <div> <div>Sort A to Z</div> <div>Sort Z to A</div> <div>More Sort Options...</div> <div>Clear Filter From "Medal"</div> <div>Label Filters</div> <div>Value Filters</div> <div>Search</div> <div> <input checked="" type="checkbox"/> (Select All) <input checked="" type="checkbox"/> Bronze <input checked="" type="checkbox"/> Silver <input checked="" type="checkbox"/> Gold </div> <div>OK Cancel</div> </div> </div>				
				Grand Total	
				1	5
				20	
				1	
				60	183
				3	4
				2	3
				6	6
				1	
				3	22
				7	
				48	
				5	13
				18	18
				4	31
				18	
				39	79
				1	1
				2	
				1	10
				22	69
				2	9
27	Denmark	4	3	18	25
28	Estonia	2		1	3
29	Ethiopia	3	1	4	8

- Choose a reverse sorting option
- You can change order also by manually selecting a column, e.g. silver, drop the column on top of one of the other to show a different sequence:

Count of Medal	Column Labels			Grand Total
Row Labels	Bronze	Silver	Gold	
Afghanistan	1			1
Algeria	8	2	4	14
Argentina	70	37	46	153
Armenia	7	1	1	9
Australia	312	270	216	798
Austria	8	17	9	34
Azerbaijan	9	3	4	16
Bahamas	3	9	7	19
Barbados	1			1
Belarus	53	25	14	92
Belgium	20	15	6	41
Bermuda	1			1
Brazil	125	137	56	318
Bulgaria	107	120	40	247
Burundi			1	1

TASK:

- What if we only want the middle tally for the Sydney Olympics in year 2000?
- We can do that by using the report filter option on the right hand side.
- We wish to filter this pivot table using that year, so select the year variable. Drag it and place it in the report filter area.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Year	2000																			
2																					
3	Count of Medal	Column Labels																			
4	Row Labels	Bronze	Silver	Gold	Total																
5	Algeria	3	1	1	5																
6	Argentina	3	17		20																
7	Armenia	1			1																
8	Australia	54	69	60	183																
9	Austria		1	3	4																
10	Azerbaijan																				
11	Bahamas	1		2	3																
12	Barbados			6	6																
13	Belarus	1			1																
14	Belgium	4	3		7																
15	Brazil	36	12		48																
16	Bulgaria	2	6	5	13																
17	Cameroon			18	18																
18	Canada	23	4	4	31																
19	Chile	18			18																
20	China	17	23	39	79																
21	Colombia			1	1																
22	Costa Rica	2			2																
23	Croatia	5		1	10																
24	Cuba	12	35	22	69																
25	Czech Republic	4	3	2	9																
26	Denmark	4	3	18	25																
27	Estonia	2		1	3																
28	Ethiopia	3	1	4	8																
29	Finland	1	1	3	5																
30	France	14	30	22	66																
31	Georgia	6			6																
32	Germany	65	23	31	119																
33	Greece	8	6	4	18																
34	Hungary	4	24	25	53																

PivotTable Field List

Choose fields to add to report:

☐ City
☒ Year
☐ Sport
☐ Discipline
☐ Event
☐ Athlete
☐ Gender
☐ Country_Code
☒ Country
☐ Event_gender
☒ Medal

Drag fields between areas below:

☒ Report Filter
☐ Column Labels

Year
Medal

☐ Row Labels
☒ Values

Country
Count of Medal

☐ Defer Layout Update

Task:

- Do a Sydney Olympics medal tally only for the sport of boxing:

	A	B	C	D	E
1	Year	2000			
2	Sport	Boxing			
3					
4	Count of Medal	Column Labels			
5	Row Labels	Bronze	Silver	Gold	Total
6	Algeria		1		1
7	Azerbaijan				
8	Cuba	2		4	6
9	Czech Republic		1		1
10	France	1		1	2
11	Georgia	1			1
12	Germany	1			1
13	Hungary	1			1
14	Italy	1			1
15	Kazakhstan		2	2	4
16	North Korea	1			1
17	Mexico	1			1
18	Moldova	1			1
19	Morocco	1			1
20	Romania	1	1		2
21	Russia	2	3	2	7
22	Spain		1		1
23	Thailand	1		1	2
24	Ukraine	3	2		5
25	United Kingdom			1	1
26	United States	2	2		4
27	Uzbekistan				
28	Total	24	12	12	48

- What if we want the medal tally broken up across all the sports?
- Show the medals for each sport in each country in 2000 Sydney Olympics?