



Announcements

- MP5 will be released today
- Excel Spreadsheets are available on course website

Lecture on Monday:

- `Match(lookup_value, lookup_array, [match_type])`
 - `Match_type` is always 0 (since we want exact match)
 - `Match()` isn't generally useful by itself
 - We need the `Index()` function – `Index(array, row_num, [column_num])`
 - Combine the two into 1 formula -
 - `Index(array, Match(lookup_value, lookup_array, 0), [column_num])`
-

Look at Spreadsheet – Excel Spreadsheet #1 from the schedule page

Q11. How many days did it snow in 2010? (Q11 is in row 153)

- We will use `SUMIFS()` as the number of snow days in every month are already given in the data and we just need to add them up in our range of interest i.e. for all the months in 2010
- So, sum range is L, criteria range is to find the year 2010, so the criteria range would be A2:A109 and criteria is "2010".
- Hence, use: `SUMIFS(L2:L109, A2:A109, "2010")`

Q12. What month had the greatest amount of snowfall?

- Snowfall data is in column H – in number of inches per month
- **D157**: Use `MAX()` function to find the maximum amount of snowfall – `MAX(H2:H109) = 20.4` inches
- **E157**: To find out which month and year had max snowfall, use `MATCH(D157, H2:H109, 0)` which will return 72 because it is the 72nd record (D157 is the cell that had the result of `MAX(H2:H109)` which now contains the value 20.4). Row 73 has the correct data but since our range is from row 2 to row 109, we will get 72 as the index, i.e. 72nd row in our range has the value we are trying to find.
- **C157**: Now you have to look up the month that corresponds to 72nd row, use `INDEX(B2:B109, E157, 1)` which will give you the value in the E157th (i.e. 72nd) position in the range B2:B109.
- You can also combine the steps to write:
`INDEX(B2:B109, Match(Max(H2:H109), H2:H109, 0), 1)`



Q13. How many days did it snow in the month with the greatest amount of snowfall?

- We already know that row 72 had the maximum snow and this result is in the cell E157
- $\text{INDEX}(H2:H109, E157, 1) = 10$ Days, we look in column H as that has the number of snow days
- It snowed for 10 days in the month that had the maximum snowfall

Q14. Which year did this happen in?

- Now we can get all sorts of information about the month that has maximum snowfall using the index function on different columns with the row value in E157
- C164: $\text{INDEX}(A2:A109, E157, 1) = 2010$

Q15. Which month did this happen?

- C167: $\text{INDEX}(B2:B109, E157, 1) = 12$

Q16. How many days did it snow in that year? (Q14)

- You already found the year in answering Q14. In the cell C163
- So just sum the number of snow days (column L) if the entry in column A is C163
- C170: $\text{SUMIFS}(L2:L109, A2:A109, C163) = 21$

Summary sheet: The second sheet in the same workbook, Excel Spreadsheet #1

- A workbook (i.e. excel file) can contain multiple sheets.
- Summary sheet in our case is used to summarize the information contained in the raw data sheet. It contains condensed analyses from your raw data.
- If we want to reference data from a different sheet, put the name of the sheet in quotes and follow with exclamation mark, example: 'Sheet1'!E57
- If the sheet name does not have a space, you need not put it in quotes.
- B2: $\text{SUMIFS}('Raw Data'!C2:C109, 'Raw Data'!A2:A109, Summary!B1)$
- So sum range and criteria range are from sheet 'Raw Data'
- Here Summary!B1 refers to 2005. The value is in the same sheet that we are working on.
- You do not have to include Summary! before B1 as the data is in the same sheet.
- So you could also write
 $=\text{SUMIFS}('Raw Data'!C2:C109, 'Raw Data'!A2:A109, B1)$
- To populate C2 to J2, you cannot just copy the formula in B2 as the columns in formulas change if you do that.
- We need a method to lock rows and/or columns so that when we copy and paste formulas the columns and rows in formula that you do not want to change, stay the same.
- We use the \$ symbol to lock rows and columns as appropriate, as shown below:



Copy – Paste Formulas and use of \$

- Formulas will get updated by how many rows and columns you move in copying and pasting
- If A1 is the formula on B1 and it is copied from B1 into B2 and C3, this is what it would amount to:

	A	B	C
1		=A1	
2		=A2	
3			=B3

In copying from B1 to C3, we changed the row by 1 and column by 2, you can see that in the formula, the row and column changed by 1 and 2 as well (A1 to B3).

- Use of the dollar (\$) sign – \$A will make sure that the column is locked

	A	B	C
1		=\$A1	
2		=\$A2	
3			=\$A3

The columns do not change but the rows do as they have not been locked.

The rows can similarly be locked if you put the \$ sign before row numbers in your formula

- \$ before A and 1 both will lock both the row and the column

	A	B	C
1		=\$A\$1	
2		=\$A\$1	
3			=\$A\$1

So to populate fields C2 through J2, we have to lock the columns from sum range and criteria range (as we are changing columns in going from C2 to J2) in our formula.

`=SUMIFS('Raw Data'!$C2:$C109,'Raw Data'!$A2:$A109,B1)`

But as you may have guessed, we also want to use the same formula for populating C3 through J3 and then C4 through J4, i.e. we have to change the rows as well. However, we do not want the rows to change in our formula, so we might as well fix them at this point.

`=SUMIFS('Raw Data'!C2:C109,'Raw Data'!A2:A109,B$1)`

So if we paste this into B3, we need to just change C to D in our formula as the sum range now is in column D (for rain days).



```
=SUMIFS('Raw Data'!$D$2:$D$109, 'Raw Data'!$A$2:$A$109, B$1)
```

Now this can be copied and pasted in cells C3 through J3 to get number of rain days in those years. The only quantity that does change in the formulas in any row is the column of the criteria in our SUMIFS () statement i.e. B in the formula above which is not preceded by a \$.

So you write the formula only a few times and then copying and pasting it intelligently will give you the power to analyze a large data set.

VLOOKUP

- We are working in the file, Excel Spreadsheet #2 from the website
- MATCH("Apples", A2:A100) did not work
- MATCH("Apples ", A2:A100) works, we will see a workaround for such cases where the data is not 'perfect' in the lectures to come
- Since "Apples " was typed with Apples and a space at the end
- Formula that combines index and match is called VLOOKUP ()
- VLOOKUP(lookup_value, table_array, col_index_num, FALSE)
- VLOOKUP("Apples ", A2: B100, 2, FALSE)
 - First argument is the one that is searched in the **first column** from your table_array i.e. Column A in this case
 - Takes in a 2D array A2:B100, i.e. multiple rows and columns as the second argument, the column where you are expecting to search has to be the first column
 - Col_index_num: is the third argument, it is the column number in table_array from which the matching value will be returned
 - The last argument is FALSE as we are trying to find an exact match.
 - This formula will find "Apples " in the first column and return the value that is present in the 2nd column (since Col_index_num is 2) corresponding to the row that contains "Apples " i.e. it will return 48.