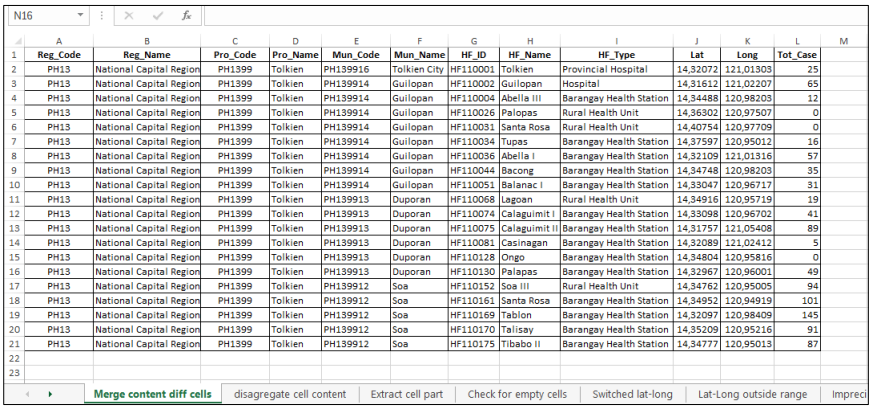
Guidance for the management and use of geospatial data and technologies in health

Part 2 - Implementing the geospatial data management cycle: 2.5 Cleaning, validating, and documenting the data - 2.5.2 Using advanced Microsoft Excel functions and tools

Version 1.4 (last update: 21.02.2024)



N16

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	HF_ID	HF_Name	HF_Type	Lat	Long	Tot_Case	
2	PH13	National Capital Region	PH1399	Tolkien	PH139916	Tolkien City	HF110001	Tolkien	Provincial Hospital	14,32072	121,01303	25	
3	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110002	Guilopan	Hospital	14,31612	121,02207	65	
4	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110004	Abella III	Barangay Health Station	14,34488	120,98203	12	
5	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110026	Palopas	Rural Health Unit	14,36302	120,97507	0	
6	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110031	Santa Rosa	Rural Health Unit	14,40754	120,97709	0	
7	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110034	Tupas	Barangay Health Station	14,37597	120,95012	16	
8	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110036	Abella I	Barangay Health Station	14,32109	121,01316	57	
9	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110044	Bacong	Barangay Health Station	14,34748	120,98203	35	
10	PH13	National Capital Region	PH1399	Tolkien	PH139914	Guilopan	HF110051	Balanac I	Barangay Health Station	14,33047	120,96717	31	
11	PH13	National Capital Region	PH1399	Tolkien	PH139913	Duporan	HF110068	Lagoan	Rural Health Unit	14,34916	120,95719	19	
12	PH13	National Capital Region	PH1399	Tolkien	PH139913	Duporan	HF110074	Calaguimit I	Barangay Health Station	14,33098	120,96702	41	
13	PH13	National Capital Region	PH1399	Tolkien	PH139913	Duporan	HF110075	Calaguimit II	Barangay Health Station	14,31757	121,05408	89	
14	PH13	National Capital Region	PH1399	Tolkien	PH139913	Duporan	HF110081	Casinagan	Barangay Health Station	14,32089	121,02412	5	
15	PH13	National Capital Region	PH1399	Tolkien	PH139913	Duporan	HF110128	Ongo	Barangay Health Station	14,34804	120,95816	0	
16	PH13	National Capital Region	PH1399	Tolkien	PH139913	Duporan	HF110130	Palapas	Barangay Health Station	14,32967	120,96001	49	
17	PH13	National Capital Region	PH1399	Tolkien	PH139912	Soa	HF110152	Soa III	Rural Health Unit	14,34762	120,95005	94	
18	PH13	National Capital Region	PH1399	Tolkien	PH139912	Soa	HF110161	Santa Rosa	Barangay Health Station	14,34952	120,94919	101	
19	PH13	National Capital Region	PH1399	Tolkien	PH139912	Soa	HF110169	Tablon	Barangay Health Station	14,32097	120,98409	145	
20	PH13	National Capital Region	PH1399	Tolkien	PH139912	Soa	HF110170	Talisay	Barangay Health Station	14,35209	120,95216	91	
21	PH13	National Capital Region	PH1399	Tolkien	PH139912	Soa	HF110175	Tibabo II	Barangay Health Station	14,34777	120,95013	87	
22													
23													

Merge content diff cells | disagregate cell content | Extract cell part | Check for empty cells | Switched lat-long | Lat-Long outside range | Impreci

In collaboration and with the support of:

THE GLOBAL FUND

Revision History

Revision	Revision Date	Comment	By
1.0	28 May 2019	Document created	Izay Pantanilla, Steeve Ebener
1.1	04 March 2020	Adjustment of the terminology associated to the concepts of geographical features, geographic objects and master lists to align with other volumes of the series	Steeve Ebener
1.2	10 January 2022	Inclusion of the reference to the new HGL guidance document	Izay Pantanilla
1.3	08 April 2022	Modified the steps for checking for empty cells	Izay Pantanilla
1.4	21 February 2024	Improvement of some of the methods and addition of new functions. Update of the layout and URLs across the document	Steeve Ebener, Abdoul Bassit Sawadogo

Authors

Izay Pantanilla [1]
Steve Ebener [1]

1. Health GeoLab, Manila, Philippines

Acknowledgments

Our gratitude goes to the Global Fund for having supported the revision of this document. MORU is funded by Wellcome.

Table of Content

Purpose and audience

The purpose of the Health GeoLab series of guidance is to inform concerned practitioners about the key elements they need to be aware of when it comes to managing and using geospatial data and technologies in public health and guide them through the processes to be followed in that regard.

The audience for this guidance includes geospatial data managers, technical advisors, and any other practitioners that are directly or indirectly involved in the collection and use of geospatial data and technologies in public health.

Please note that some of the sections in the present guidance require a basic understanding of concepts pertaining to the use of Microsoft Excel.

Abbreviations

GIS	Geographic Information System
GNSS	Global Navigation Satellite System
HGL	Health GeoLab
HIS	Health Information System
MS	Microsoft
SDG	Sustainable Development Goal

1. Background

The Health GeoLab (HGL) is a regional resource supporting low- and middle-income countries in Asia and the Pacific to fully benefit from the power of geography, geospatial data and technology to achieve the Sustainable Development Goals health-related (SDG 3), namely enabling healthy lives for all and promoting well-being for all at all ages.[1]

The HGL uses the Health Information System (HIS) geo-enabling framework to strengthen the capacity of the health sector in countries related to the management and use of geospatial data and technologies. This document was developed as part of this approach and with the aim of being used by as many users as possible, including those outside Asia.

This volume is part of a series of guidance that are organized as follows:

- Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management cycle [1]
- Part 2 - Implementing the geospatial data management cycle:
    - 2.1 Documenting the process and defining the data needs [2]
    - 2.2 Defining the terminology, data specifications, and the ground reference [3]
    - 2.3 Compiling existing data and identifying gaps [4]
    - 2.4 Creating geospatial data
        - 2.4.1 Extracting vector format geospatial data from basemaps [5]
        - 2.4.2 Collecting data in the field [6]
    - 2.5 Cleaning, validating, and documenting the data
        - 2.5.1 Documenting the data using a metadata profile [7]
        - 2.5.2 Using advanced Microsoft Excel functions (the present document)
    - 2.6 Distributing, using, and updating the data
        - 2.6.1 Creating good thematic maps using desktop GIS software [8]
        - 2.6.2 Using thematic maps for decision making [9]
        - 2.6.3 Developing and implementing the appropriate data policy [10]

This document is designed to evolve based on input received from users. Do not hesitate to contact us if you have any suggestions for improvement.

The terms used in the present guidance are defined in the following glossary of terms maintained by the Health GeoLab: http://tinyurl.com/2dbu7ypf .

Please also contact us at the same email address if you use this document as part of your activities and would like your institution to be recognized as one of its users.

[1] https://www.un.org/sustainabledevelopment/fr/health/

2. Introduction

While most of the work done by a geospatial data manager[2] is done using geospatial technology (Geographic Information System (GIS), Global Navigation Satellite System (GNSS), and Remote Sensing (RS)), several key operations must be performed in a spreadsheet. This is the case when creating geospatial data but more often when verifying, cleaning, or validating data compiled or collected in the field.

Without being exhaustive, this document describes these operations and how they can be carried out using the advanced functions and tools of Microsoft Excel 365 [3]. The functions in question are briefly described in Annex 1 and accompanied by a link to a web page describing how to use them in more detail.

The decimal separator used in the document is the point (.) in accordance with the ISO 80000-1 standard.

Examples based on fictitious data are provided to illustrate each of these operations and allow them to be practiced. These examples are accessible in the following MS Excel files:

- Original file: http://tinyurl.com/3k7keth9
- File with results: http://tinyurl.com/2vxcnerd

3. Operations performed using advanced Microsoft Excel functions and tools

The operations discussed in this section are those that will most likely be used when verifying, cleaning, or validating data that has been compiled or collected in the field.

The steps presented here are performed on example data contained in the Excel file mentioned here above.

3.1 Merge different cells' content

Information to be managed in a master list or displayed on a map may require the content of different cells to be merged.

For example, we may need the type of health facility to be included with the name of this facility while these two pieces of information are stored in separate cells (name of health facility in one cell, type of facility health in another).

Merging the two pieces of information into a single cell can be achieved using the CONCATENATE function.

The formula used is as follows:

= CONCATENATE (text 1, text 2, text 3, ...)

[2]See Annex J of the guidelines on the use of geospatial data and technologies in immunization programs: https://www.unicef.org/media/58181/file

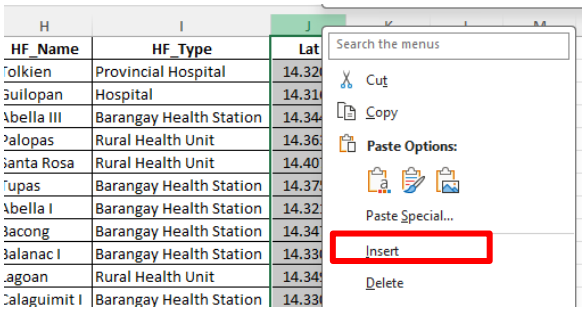
[3]The functions presented here are also available in older versions of MS Excel. The way they are worded can nevertheless be a little different depending on the version (use of ";" instead of ",")

With :

- *text 1*: the first part of the text to merge. This can also be the coordinate of the cell in which the text is located (example: A2)
- *text 2*: the second part of the text to merge.
- *text 3:* the third part of the text to merge.

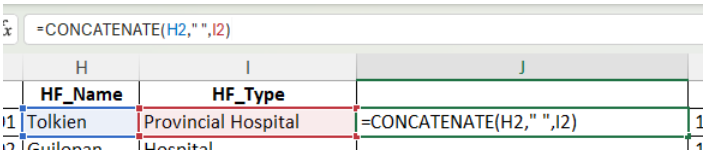
Here is the process to be used to merge the content of two cells using the content of the "Merge content diff cells" worksheet from the MS Excel examples file mentioned in the introduction:

1. Identify the columns for which you want to combine information. In this case, these are columns H and I (*HF_Name* and *HF_Type*).
2. Insert a column to the right of these columns by right clicking the column letter *J* and choosing *Insert* .

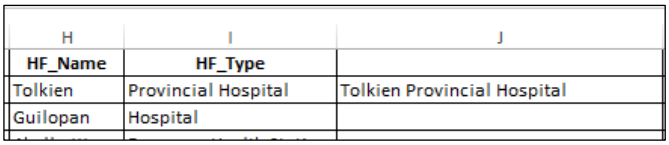


3. In cell J2, use the CONCATENATE function as follows to combine the health facility name contained in cell H2 and the health facility type contained in cell I2 with a space in between :

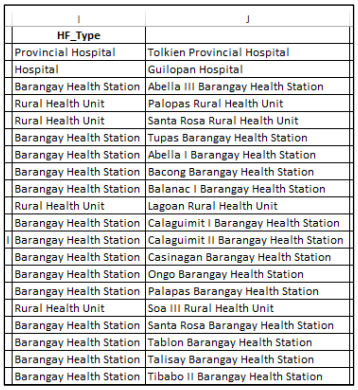
= CONCATENATE(H2," ",I2)



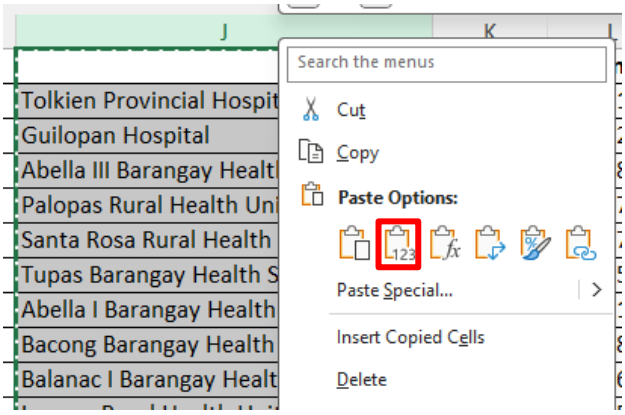
4. Press the *Enter key* on your keyboard. This will bring up the function result for cell J2.



5. To extend the function to the rest of column J, double-click the small square in the lower right corner of the cell where you entered the function ( ). This will fill the rest of the column with the same formula and thus the merge result (you can also click and hold the bottom right corner of the cell and drag it to the last row of the table).



6. If you want to freeze the content of the column in which you have just applied the function, select the column in question, copy its content, and paste it on itself using the Paste *Special>Values* from the bar function from the main menu or by right-clicking on the letter in column J.



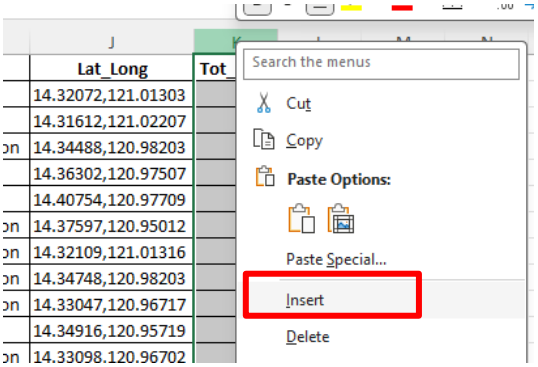
3.2 Disaggregate the content of a cell into separated ones

Storing multiple information in a single cell (such as address, geographic coordinates, etc.) may prevent correct representation in a GIS software or result in poor analysis. This information should be captured into separate cells. This task can be performed using the Text-to-Column Conversion Wizard.

In this example, the latitude and longitude of each record are in the same cell and will be separated into individual cells.

Here is the process to achieve this result using the content of the "Disaggregate cell content" worksheet from the MS Excel examples file mentioned in the introduction:

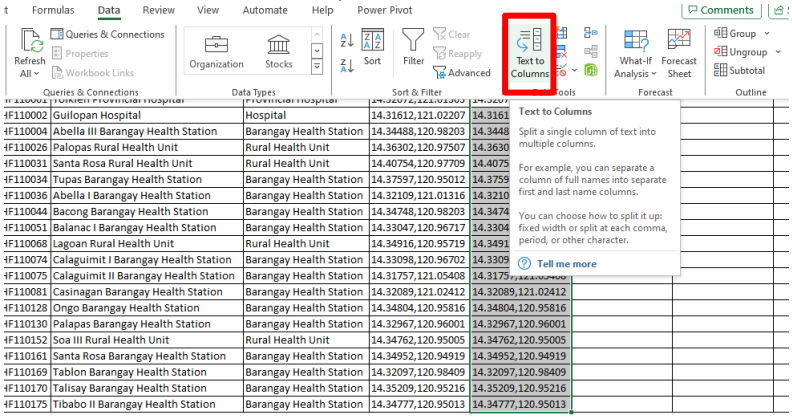
1. Identify the column containing the combined information and identify the character used to separate the information (comma, space, tab, etc.). In this example, the information is in column J (*Lat_Long*). In this column, the latitude and longitude of each record are separated by a comma.
2. Insert three columns after the one containing the combined information (*Lat_Long*) by right-clicking the letter of the next column (*K)* and choosing *Insert*. Repeat this to add the second and third columns.



3. Copy the content of column J (*Lat_Long)* into column *K.* This will ensure that you have a copy of the data in case you make a mistake.

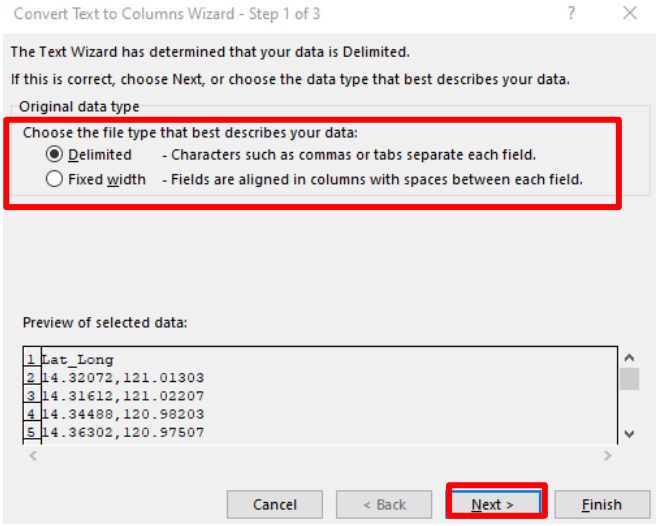
	J	K
ıe	Lat_Long	
ıital	14.32072,121.01303	14.32072,121.01303
	14.31612,121.02207	14.31612,121.02207
h Station	14.34488,120.98203	14.34488,120.98203
ıit	14.36302,120.97507	14.36302,120.97507
ıit	14.40754,120.97709	14.40754,120.97709
h Station	14.37597,120.95012	14.37597,120.95012
h Station	14.32109,121.01316	14.32109,121.01316
h Station	14.34748,120.98203	14.34748,120.98203
h Station	14.33047,120.96717	14.33047,120.96717
ıit	14.34916,120.95719	14.34916,120.95719

4. Select the column with the copied data (column *K*).
5. Under the *Data* menu tab , click *Text-to-Column* in the *Data Tools section*.

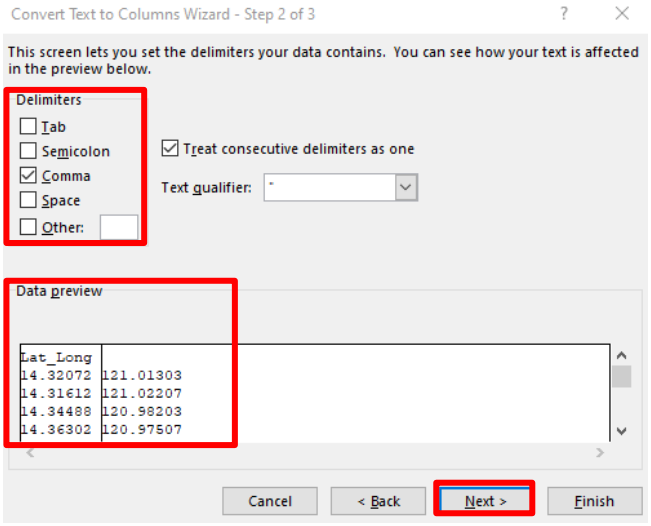


4F110002	Guilopan Hospital	Hospital	14.31612,121.02207	14.3161
4F110004	Abella III Barangay Health Station	Barangay Health Station	14.34488,120.98203	14.3448
4F110026	Palopas Rural Health Unit	Rural Health Unit	14.36302,120.97507	14.3630
4F110031	Santa Rosa Rural Health Unit	Rural Health Unit	14.40754,120.97709	14.4075
4F110034	Tupas Barangay Health Station	Barangay Health Station	14.37597,120.95012	14.3759
4F110036	Abella I Barangay Health Station	Barangay Health Station	14.32109,121.01316	14.3210
4F110044	Bacong Barangay Health Station	Barangay Health Station	14.34748,120.98203	14.3474
4F110051	Balanac I Barangay Health Station	Barangay Health Station	14.33047,120.96717	14.3304
4F110068	Lagoan Rural Health Unit	Rural Health Unit	14.34916,120.95719	14.3491
4F110074	Calaguimit I Barangay Health Station	Barangay Health Station	14.33098,120.96702	14.3309
4F110075	Calaguimit II Barangay Health Station	Barangay Health Station	14.31757,121.05408	14.3175
4F110081	Casinagan Barangay Health Station	Barangay Health Station	14.32089,121.02412	14.32089,121.02412
4F110128	Ongo Barangay Health Station	Barangay Health Station	14.34804,120.95816	14.34804,120.95816
4F110130	Palapas Barangay Health Station	Barangay Health Station	14.32967,120.96001	14.32967,120.96001
4F110152	Soa III Rural Health Unit	Rural Health Unit	14.34762,120.95005	14.34762,120.95005
4F110161	Santa Rosa Barangay Health Station	Barangay Health Station	14.34952,120.94919	14.34952,120.94919
4F110169	Tablon Barangay Health Station	Barangay Health Station	14.32097,120.98409	14.32097,120.98409
4F110170	Talisay Barangay Health Station	Barangay Health Station	14.35209,120.95216	14.35209,120.95216
4F110175	Tibabo II Barangay Health Station	Barangay Health Station	14.34777,120.95013	14.34777,120.95013

6. The Text-to-Column Conversion Wizard opens. In the first window, you will be asked to choose the file type that best describes your data. Choose *Delimited* because a comma separates latitude and longitude. Click *Next*.

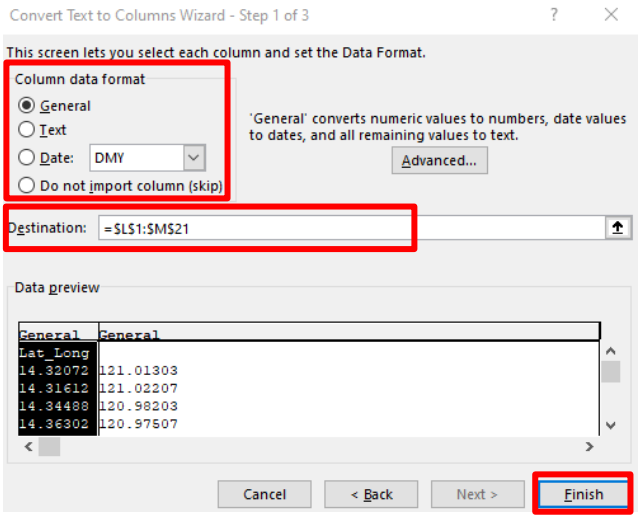


7. In the next window, choose *Comma* as separator. Verify that the expected result is obtained in the Data Preview part of the window. Click *Next if* this is the case.

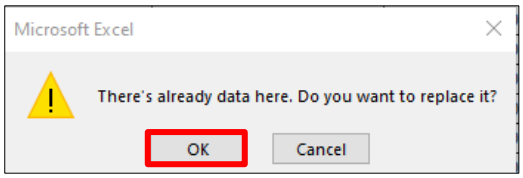


8. In the third window:
    a. *Column data format*: General
    b. *Destination*: Select the cells in columns L and M in which the disaggregation result will be captured

    vs. Click *Finish*.



9. When asked if you want to replace the content of the destination cells, click *OK.*



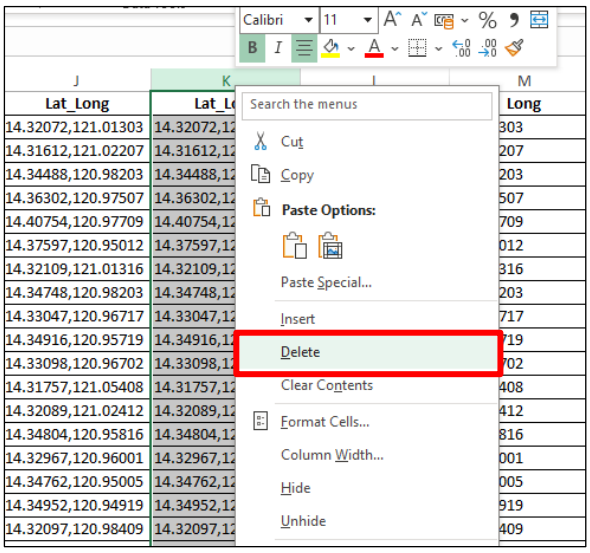
This will cause the latitude and longitude of each record to appear separately in columns L and M. Check that this is the case.

K	L	M
Lat_Long	Lat_Long	
14.32072,121.01303	14.32072	121.01303
14.31612,121.02207	14.31612	121.02207
14.34488,120.98203	14.34488	120.98203
14.36302,120.97507	14.36302	120.97507
14.40754,120.97709	14.40754	120.97709
14.37597,120.95012	14.37597	120.95012
14.32109,121.01316	14.32109	121.01316
14.34748,120.98203	14.34748	120.98203
14.33047,120.96717	14.33047	120.96717
14.34916,120.95719	14.34916	120.95719

10. Name the column L *Lat* and column M *Long* ..

K	L	M
Lat_Long	Lat	Long
14.32072,121.01303	14.32072	121.01303
14.31612,121.02207	14.31612	121.02207
14.34488,120.98203	14.34488	120.98203
14.36302,120.97507	14.36302	120.97507
14.40754,120.97709	14.40754	120.97709

11. You can now delete column K containing information by right clicking the letter in column *K* and choosing *Delete.*



3.3 Extract part of a cell's content

It is sometimes necessary to extract part of the content of a cell to place it in another cell. This task can be performed using the RIGHT or LEFT functions.

These two functions will extract a specified number of characters from the right or left side of a provided text string. The formula used is as follows:

=RIGHT(text, [num_chars])

Or

=LEFT(text, [num_chars])

With:

- *RIGHT* or *LEFT*: The side from which the count of characters to extract will begin
- *text:* The text from which to extract characters from the right or left. This can also be the coordinate of the cell in which the text is located (example: A2)
- *num_chars:* The number of characters to extract, starting from the right or the left. Optional, default = 1.

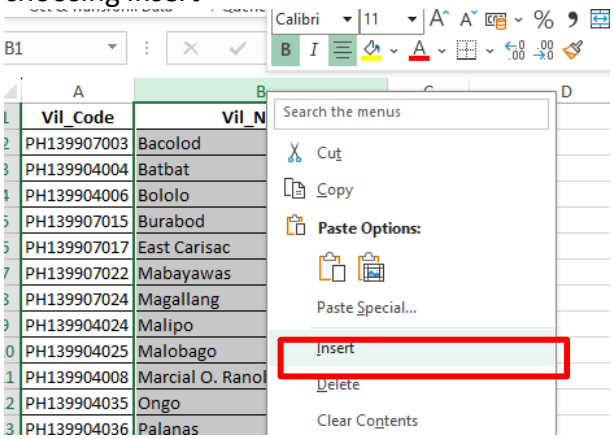
In the following example, we will extract from the unique identifier of each village the part corresponding to the unique identifier of the municipality in which they are located.

The process is as follows using the content of the "Extract cell part" worksheet from the MS Excel examples file mentioned in the introduction:

1. Identify the column containing the information you want to extract and the number of characters from the right or left you want to extract. In this example, the information is in column A (*Vil_Code)* and the part of the text to extract is made up of the first eight (8) characters from the left.

	A	B
1	**Vil_Code**	**Vil_Name**
2	PH139907003	Bacolod
3	PH139904004	Batbat
4	PH139904006	Bololo
5	PH139907015	Burabod
6	PH139907017	East Carisac
7	PH139907022	Mabayawas
8	PH139907024	Magallang

2. Insert a column to the right of column (*Vil_Code)* by right-clicking the letter of column B and choosing *Insert*

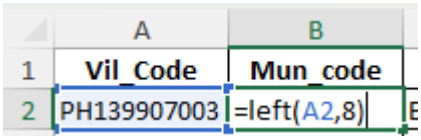


3. Name this new column Mun_ code .

	A	B	C
1	**Vil_Code**	**Mun_code**	**Vil_Name**
2	PH139907003		Bacolod
3	PH139904004		Batbat
4	PH139904006		Bololo
5	PH139907015		Burabod
6	PH139907017		East Carisac
7	PH139907022		Mabayawas
8	PH139907024		Magallang

4. In cell B2, enter the following formula based on the LEFT function to extract the first 8 characters:

   =LEFT(A2,8)



5. Press the *Enter key* on your keyboard. This will bring up the result of the LEFT function in the cell.

6.

	A	B
1	Vil_Code	Mun_code
2	PH139907003	PH139907
3	PH139904004	

7. Click and hold the small square in the lower right corner of the cell where you entered the function ( ) and drag the cursor down to the last row of the table. This will fill the rest of the column with the same formula and therefore the result for the LEFT function.

	A	B	C
1	Vil_Code	Mun_code	Vil_Name
2	PH139907003	PH139907	Bacolod
3	PH139904004	PH139904	Batbat
4	PH139904006	PH139904	Bololo
5	PH139907015	PH139907	Burabod
6	PH139907017	PH139907	East Carisac
7	PH139907022	PH139907	Mabayawas
8	PH139907024	PH139907	Magallang
9	PH139904024	PH139904	Malipo
10	PH139904025	PH139904	Malobago
11	PH139904008	PH139904	Marcial O. Ranola
12	PH139904035	PH139904	Ongo
13	PH139904036	PH139904	Palanas
14	PH139904046	PH139904	Sinungtan
15	PH139907018	PH139907	West Carisac
16	PH139904003	PH139904	Zone I (Pob.)
17	PH139907004	PH139907	Zone I (Pob.)
18	PH139907008	PH139907	Zone V (Pob.)
19	PH139907009	PH139907	Zone VI (Pob.)

8. To freeze the content of the column in which you have just applied the function, copy its content and paste it on itself using the *Paste Special>Values function.*

3.4 Identify empty cells

Empty cells in an MS Excel spreadsheet are not uncommon and a situation that might need to be corrected because these kinds of cells could mean that:

1. The value or information is missing, perhaps because it was not collected or is not coded correctly.
2. The value in the cell should be null/zero.
3. The value or information is unknown.

Regardless of the reason, these situations must be identified, and a specific value entered in the concerned cells to avoid any confusion or analysis error.

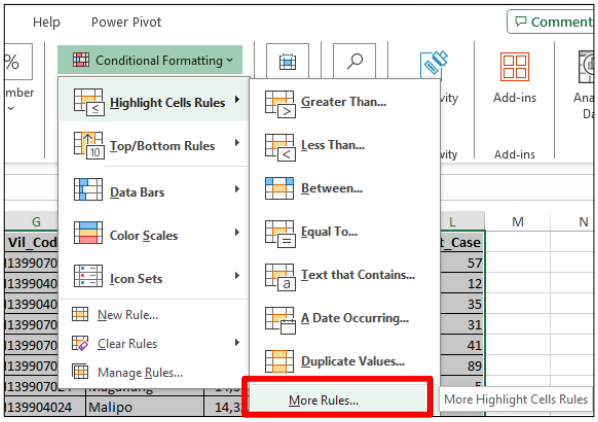
The easiest way to identify such cells in an MS Excel spreadsheet is to highlight these cells.

The process is as follows using the content of the "Check for empty cells" worksheet from the MS Excel examples file mentioned in the introduction:

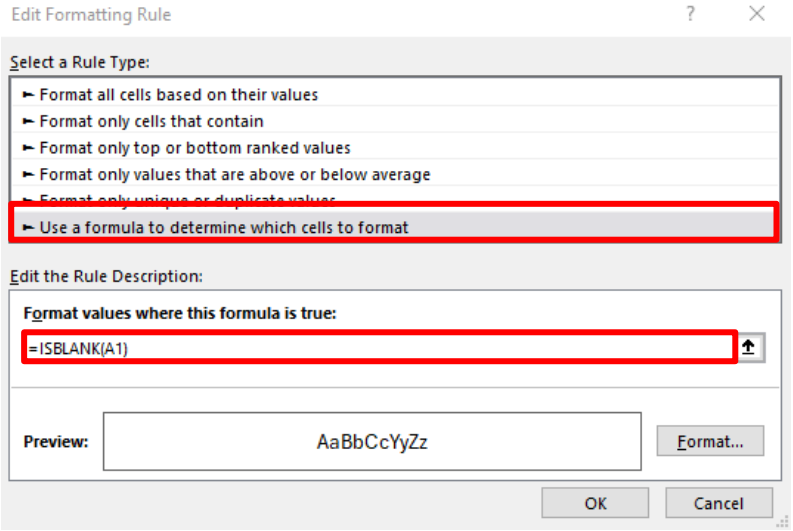
1. Select the content of the entire table containing the empty cells to identify.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
2		National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	14.38984	120.98114	724	57
3	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12
4	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
5	PH13	National Capital Region (NCR)			PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
6	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	14.33051	121.07202	682	41
7	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	14.38790	121.02925	2019	89
8	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	14.34245	121.03314	633	5
9	PH13		PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
10	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994		PH139904025	Malobago	14.35216	120.95519	441	19
11	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465	
12	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
13		National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
14	PH13		PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	14.34113	121.08804	284	
15	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
16	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	14.38956	121.00637	1145	145
17	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.41479	120.99752		91
18	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	14.39777	120.98938	932	87
19	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904036		14.33679	120.98815	421	

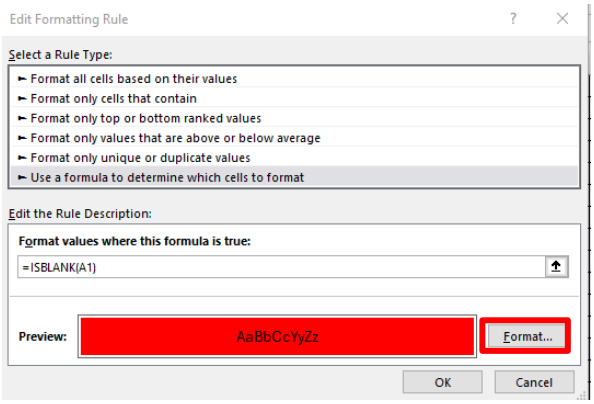
2. In the *Home* tab of the menu bar, click the *Conditional Formatting icon* in the *Styles section.* Then choose the *Highlight Cells Rules* and *More Rules* item



3. In the window that opens, choose the option *Use a formula to determine for which cells to format* and specify the following formula in the *Format values where this formula is true* field: =ISBLANK(A1) (Note: A1 corresponds to the upper left corner of the selected table)



4. Click on the *Format* button to set the color of empty cells. Red in the example here:



5. Click on Ok. This will make the empty cells appear in red in the table

	A	B	C	D	E	F	G	H	I	J	K	L
1	Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
2		National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	14.38984	120.98114	724	57
3	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12
4	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
5	PH13	National Capital Region (NCR)			PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
6	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	14.33051	121.07202	682	41
7	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	14.38790	121.02925	2019	89
8	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	14.34245	121.03314	633	5
9	PH13		PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
10	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994		PH139904025	Malobago	14.35216	120.95519	441	19
11	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465	
12	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
13		National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
14	PH13		PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	14.34113	121.08804	284	
15	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
16	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	14.38956	121.00637	1145	145
17	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.41479	120.99752		91
18	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	14.39777	120.98938	932	87
19	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904036		14.33679	120.98815	421	

3.5 Identify switched latitudes and longitudes

Differentiating latitude from longitude in a pair of geographic coordinates can sometimes be confusing, especially for people who are unfamiliar with how these pairs work.

As a result, they can sometimes be switched up in a database. This must be identified and corrected before the data is used in a GIS software to ensure that all points are in the correct place on the map.

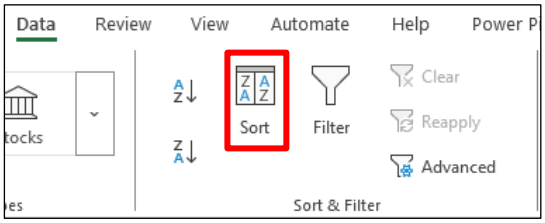
This task can be done using the *Sort* function in MS Excel.

Here is the process to be used to check for switched "Lat -long" spreadsheet from the MS Excel examples file mentioned in the introduction:

1. Select the content of the entire table containing the geographic coordinates to analyze

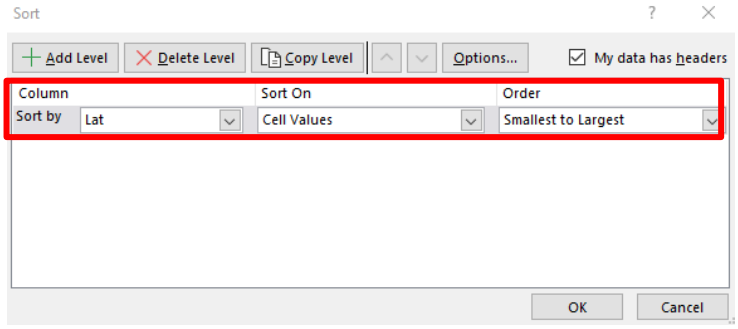
	A	B	C	D	E	F	G	H	I	J	K	L
1	Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
2	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	14.38984	120.98114	724	57
3	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12
4	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
5	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
6	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	14.33051	121.07202	682	41
7	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	14.38790	121.02925	2019	89
8	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	14.34245	121.03314	633	5
9	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
10	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
11	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	120.97810	14.36196	465	0
12	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
13	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904036	Palanas	14.33679	120.98815	421	0
14	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
15	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	14.34113	121.08804	284	0
16	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
17	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	121.00637	14.38956	1145	145
18	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.41479	120.99752	947	91
19	PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	14.39777	120.98938	932	87

2. Under the *Data* menu, click the *Sort Icon* in the *Sort & Filter section*.



3. In the *Sort* dialog window that opens, check the *My data has headers* option. This will make it easier to select the column on which the sorting will be based.

4. In the second part of the window choose:
    a. Column: Sort by *Lat*
    b. Sort on *: Cell Values*
    c. Order: *Smallest to largest*



5. Click *OK*. This will sort the table's content by placing the smallest latitude at the top.
6. Look at the records at the top of the table to see if any of the coordinate pairs appear switched. This is not the case in the example here:

G	H	I	J
Vil_Code	**Vil_Name**	**Lat**	**Long**
PH139904024	Malipo	14.32559	120.96507
PH139907017	East Carisac	14.33051	121.07202
PH139904036	Palanas	14.33679	120.98815
PH139904046	Sinungtan	14.34103	120.97309
PH139907018	West Carisac	14.34113	121.08804

7. Look at the records at the bottom of the table to see if any pairs appear switched. This is the case for the last two records in the example here:

Zone VI (Pob.)	14.39777	120.98938
Burabod	14.40817	121.03665
Zone V (Pob.)	14.41479	120.99752
Marcial O. Ranola	120.97810	14.36196
Zone I (Pob.)	121.00637	14.38956

8. Correct the inverted coordinates and save the resulting file.

3.6 Identify if latitudes and longitudes fall within a specific zone

A quick way to check whether geographic coordinates have been correctly collected in the field before even using them, including in GIS software, is to identify whether they are in the administrative unit indicated in the master list.

This task can be done using one of the following functions, each of which first requires knowing the minimum and maximum latitude and longitude (rectangle) of the area encompassing the administrative unit in question [4]:

- MIN and MAX functions
- Sort function
- Conditional formatting function

Lat -Long Out of Area" spreadsheet in the MS Excel examples file mentioned in the introduction:

Mun_Code	Name_Mun	LAT_MIN	LAT_MAX	LONG_MIN	LONG_MAX
PH139904000	Guinotaban	14.31	14.39	120.93	121.00
PH139907000	Bilon	15.32	15.42	121.00	122.00

Using the MIN and MAX functions

The MIN function returns the smallest numeric value in a range of values while the MAX function returns the largest numeric value in that same range of values.

The formulas used are as follows in the case where the coordinates are contained in cells of an MS Excel spreadsheet:

= MIN(Cell1:Cell2)

And

= MAX(Cell1:Cell2)

With:

- *MIN* or *MAX* : The function that specifies whether to return the smallest or largest value.
- *Cell1* : Starting cell of the range that contains the numeric values for which we want to identify the minimum or maximum value.
- *Cell2* : End cell of the range that contains the numeric values for which we want to identify the minimum or maximum value.

This method does not allow each record to be analyzed separately but to identify whether one or more records attached to a given administrative unit are located outside this area.

As each administrative unit has its own extension (minimum-maximum latitude-longitude), the MIN/MAX functions must be applied separately to the sets of geographic coordinates associated with each of these units.

The following steps should be followed in this regard:

1. Label four empty columns to the right of the table as follows: *Lat_ Min* , *Lat_Max* , *Long_Min* and *Long_Max* . This is where we will put the MIN/MAX formulas for latitudes and longitudes.

[4] For this method to be effective, you must have access to a vector format GIS layer of quality containing the delimitation of the administrative divisions in question.

Lat	Long	Pop_2010	Tot_Case	Lat_Min	Lat_Max	Long_Min	Long_Max
14.32559	120.96507	877	65				
14.33679	120.98815	421	0				
14.34103	120.97309	1015	101				
14.34439	120.94905	592	35				
14.35216	120.95519	441	19				
14.35493	120.99304	1204	94				
14.35847	120.97619	694	49				
14.36196	120.97810	465	0				

2. In cell N4 (under *Lat_Min*), enter the following formula based on the MIN function to identify the minimum latitude for all records associated with the municipality of Guinotaban :

   = MIN(I4:I12)

Lat	Long	Pop_2010	Tot_Case	Lat_Min
14.32559	120.96507	877	65	=MIN(I4:I12)
14.33679	120.98815	421	0	
14.34103	120.97309	1015	101	
14.34439	120.94905	592	35	
14.35216	120.95519	441	19	
14.35493	120.99304	1204	94	
14.35847	120.97619	694	49	
14.36196	120.97810	465	0	
14.37597	120.95012	835	12	

3. Press the Enter key on your keyboard. This will bring up the result of the MIN function in the cell.

Lat	Long	Pop_2010	Tot_Case	Lat_Min	Lat_Max
14.32559	120.96507	877	65	14.32559	
14.33679	120.98815	421	0		
14.34103	120.97309	1015	101		
14.34439	120.94905	592	35		
14.35216	120.95519	441	19		
14.35493	120.99304	1204	94		
14.35847	120.97619	694	49		
14.36196	120.97810	465	0		
14.37597	120.95012	835	12		

4. In cell O2 (under *Lat_Max*), enter the following formula based on the MAX function to identify the maximum latitude for all records associated with the municipality of Guinotaban :

   = MAX(I4:I12)

Lat	Long	Pop_2010	Tot_Case	Lat_Min	Lat_Max
14.32559	120.96507	877	65	14.32559	=MAX(I4:I12)
14.33679	120.98815	421	0		
14.34103	120.97309	1015	101		
14.34439	120.94905	592	35		
14.35216	120.95519	441	19		
14.35493	120.99304	1204	94		
14.35847	120.97619	694	49		

5. Press the Enter key on your keyboard. This will bring up the result of the MAX function in the cell:

Lat	Long	Pop_2010	Tot_Case	Lat_Min	Lat_Max
14.32559	120.96507	877	65	14.32559	14.37597
14.33679	120.98815	421	0		
14.34103	120.97309	1015	101		
14.34439	120.94905	592	35		
14.35216	120.95519	441	19		

6. Repeat steps 2 to 5 to obtain the minimum and maximum longitude for records associated with the municipality of Guinotaban in cells P4 and Q4.

J	K	L	M	N	O	P	Q
Long	Pop_2010	Tot_Case		Lat_Min	Lat_Max	Long_Min	Long_Max
120.96507	877	65		14.32559	14.37597	120.94905	120.99304
120.98815	421	0					
120.97309	1015	101					

7. Repeat steps 2 through 6 to obtain the minimum and maximum latitudes and longitudes for records associated with the Municipality of Bilon. Enter the formula in the cell:
    a. N13 for minimum latitude
    b. O13 for maximum latitude
    c. P13 for minimum longitude
    d. Q13 for maximum longitude

I	J	K	L	M	N	O	P	Q
Lat	Long	Pop_2010	Tot_Case		Lat_Min	Lat_Max	Long_Min	Long_Max
14.32559	120.96507	877	65		14.32559	14.37597	120.94905	120.99304
14.33679	120.98815	421	0					
14.34103	120.97309	1015	101					
14.34439	120.94905	592	35					
14.35216	120.95519	441	19					
14.35493	120.99304	1204	94					
14.35847	120.97619	694	49					
14.36196	120.97810	465	0					
14.37597	120.95012	835	12					
15.30051	121.07202	682	41		14.42479	15.40817	120.98114	121.08804
15.34113	121.08804	284	0					
15.34245	121.03314	633	5					
15.38790	121.02925	2019	89					
15.38956	121.00637	1145	145					
15.38984	120.98114	724	57					
15.39777	120.98938	932	87					
15.40817	121.03665	348	31					
14.42479	120.99752	947	91					

8. Check if the minimum and maximum latitudes and longitudes obtained are within the zone that has been defined for each municipality:
    a. If yes, this means that all municipality records are in the area. This is the case here for the Municipality of Guinotaban.
    b. If no, this means that some of the municipality's records fall outside the bounding box. This is the case for the Municipality of Bilon because at least one of the records present a latitude below 15.32 and three a longitude below 121, value defined as the minimum latitude and longitude for this municipality .
9. Correct the affected records and repeat the test to ensure that all records are now in the zone for each of the municipalities.

Using the Sort function:

Like the MIN/MAX functions, this method does not allow you to directly identify each record that is outside the area that has been defined, but to identify whether at least one of the records within a given administrative unit is located outside the area in question.

As each municipality has its minimum-maximum latitude-longitude, records are sorted by municipality and latitude/longitude.

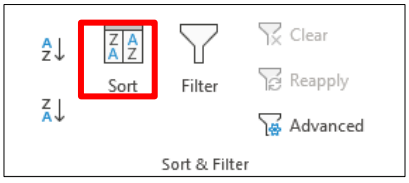
The following steps should be followed in this regard:

1. Select the content of the table.

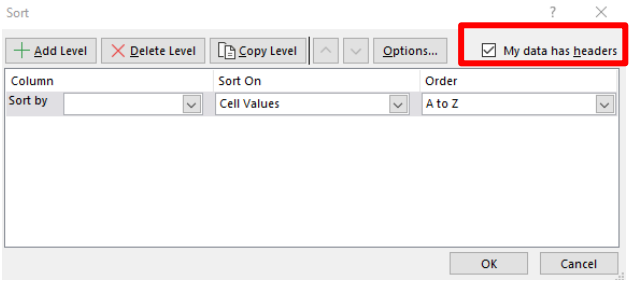
Sort function

Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	15.38984	120.98114	724	57
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	15.40817	121.03665	348	31
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	15.30051	121.07202	682	41
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	15.38790	121.02925	2019	89
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	15.34245	121.03314	633	5
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465	0
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904036	Palanas	14.33679	120.98815	421	0
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	15.34113	121.08804	284	0
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	15.38956	121.00637	1145	145
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.42479	120.99752	947	91
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	15.39777	120.98938	932	87

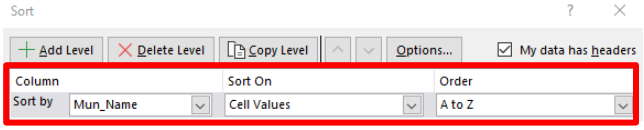
2. Under the *Data menu* , click the *Sort Icon* in the *Sort & Filter section* .



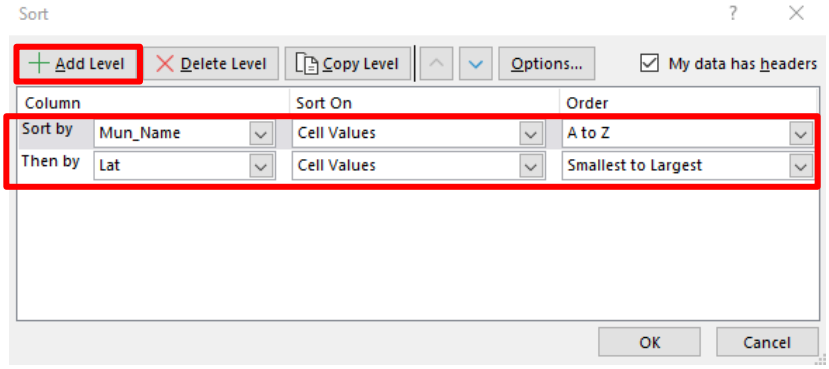
3. *Sort* dialog window opens, check the *My data has headers option* at the top right. This will allow you to select the columns on which the sorting will be based.



4. In the second part of the window choose:
    a. Column: Sort by *Mun_Name*
    b. Sort by: *Cell Values*
    c. Order: *A to Z*



5. Click the *Add Level button*. In the second part of the window choose:
    a. Column: Sort by *Lat*
    b. Sort by: *Cell Values*
    c. Order: *Smallest to largest*



6. Click *OK.* This will cause the table to sort by municipality and then an increasing value for latitude. At this point, it may be helpful to highlight the first record for each municipality, as shown below, to make it easier to locate the first entry for each municipality.

Fonction Trier

Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.42479	120.95752	947	91
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	15.30051	121.07202	682	41
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	15.34113	121.08804	284	0
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	15.34245	121.03314	633	5
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	15.38790	121.02925	2019	89
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	15.38956	121.00637	1145	145
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	15.38984	120.98114	724	57
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	15.39777	120.98938	932	87
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	15.40817	121.03665	348	31
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904036	Palanas	14.33679	120.98815	421	0
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465	0
PH13	National Capital Region (NCR)	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12

7. Note the minimum and maximum latitude values observed for each municipality. In this example, these values are as follows:

Municipality	Minimum latitude of step 6	Maximum latitude of step 6
Bilon	14.42479	15.4087
Guinotaban	14.32559	14.37597

8. Check if the latitudes noted in step 7 are within the zone defined for each municipality:
    a. If yes, this means that all municipality records are in the area. This is the case here for the municipality of Guinotaban.
    b. If not, this means that some of the municipality's records are outside the area. This is the case of the Municipality of Bilon because two of the records have a latitude below 15.32, value defined as the minimum latitude for this municipality.
9. Repeat steps 1-8 for longitudes to identify if other records are also outside the area. This is the case for 3 records of the Bilon's municipality
10. Correct the affected records and repeat the test to ensure that all records are now in the zone for each of the municipalities.

Using the conditional formatting feature

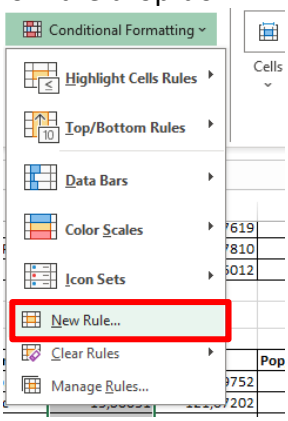
Although a little more complicated than the previous methods, this one makes it possible to identify each of the records that would fall outside the zone that has been defined for each administrative unit.

Here are the steps to follow in this regard:

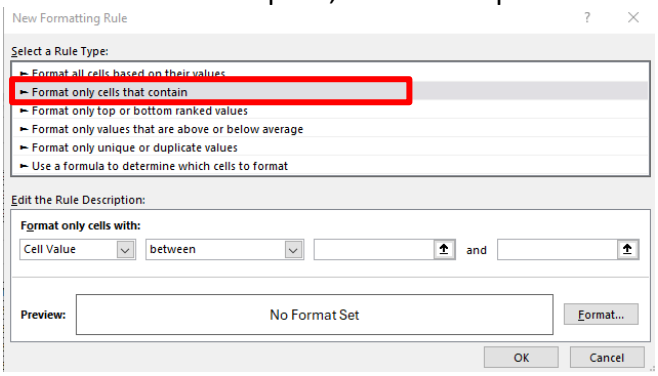
1. Select the content of the table and sort it in the same way as for the previous method
2. Select the cells containing the latitude for the records of the Bilon Municipality (cells I48-I56).

F	G	H	I
Mun_Name	Vil_Code	Vil_Name	Lat
Bilon	PH139907008	Zone V (Pob.)	14.42479
Bilon	PH139907017	East Carisac	15.30051
Bilon	PH139907018	West Carisac	15.34113
Bilon	PH139907024	Magallang	15.34245
Bilon	PH139907022	Mabayawas	15.38790
Bilon	PH139907004	Zone I (Pob.)	15.38956
Bilon	PH139907003	Bacolod	15.38984
Bilon	PH139907009	Zone VI (Pob.)	15.39777
Bilon	PH139907015	Burabod	15.40817
Guinotaban	PH139904024	Malipo	14.32559

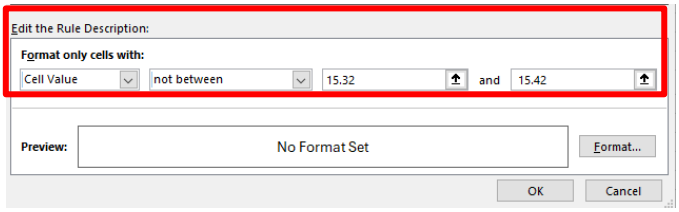
3. In the *Home* menu tab, click *Conditional Formatting* in the *Styles section*. Choose *New Rule…* from the drop-down menu.



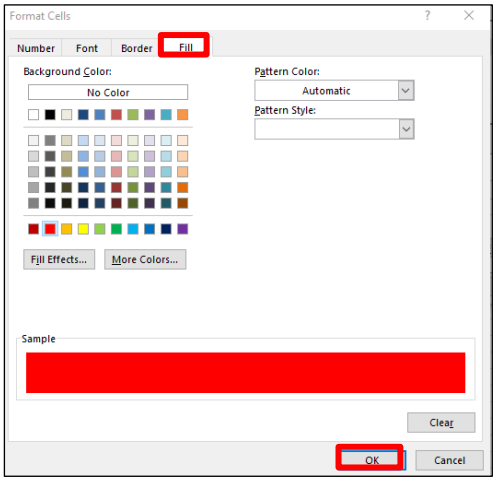
4. In the window that opens, select the option *Format only cells that contain.*



5. In the second part of the window (*Edit the Rule Description*)
    a. Keep *Cell Value* in first field.
    b. Use *Not between* in the second field.
    c. Enter the minimum and maximum values for the latitudes associated with the municipality of Bilon (15.32 and 15.42)
    d. Click the *Format button*.



6. Click the *Fill* tab and choose a color that will highlight cells whose latitude is outside the range (red for example). Click *OK* .



7. Click *OK in the New Formatting Rule* window. Any latitude outside the defined range will be highlighted in red. This is the case for two records from the municipality of Bilon:

Mun_Name	Vil_Code	Vil_Name	Lat
Bilon	PH139907008	Zone V (Pob.)	14.42479
Bilon	PH139907017	East Carisac	15.30051
Bilon	PH139907018	West Carisac	15.34113
Bilon	PH139907024	Magallang	15.34245
Bilon	PH139907022	Mabayawas	15.38790
Bilon	PH139907004	Zone I (Pob.)	15.38956
Bilon	PH139907003	Bacolod	15.38984
Bilon	PH139907009	Zone VI (Pob.)	15.39777
Bilon	PH139907015	Burabod	15.40817

8. Repeat steps 2 to 7 for the longitudes of the records associated with the municipality of Bilon, taking care to select the cells containing the longitudes for this municipality and entering the minimum and maximum longitude values for the area. As we can see this is the case for three records:

Mun_Name	Vil_Code	Vil_Name	Lat	Long
Bilon	PH139907008	Zone V (Pob.)	14.42479	120.99752
Bilon	PH139907017	East Carisac	15.30051	121.07202
Bilon	PH139907018	West Carisac	15.34113	121.08804
Bilon	PH139907024	Magallang	15.34245	121.03314
Bilon	PH139907022	Mabayawas	15.38790	121.02925
Bilon	PH139907004	Zone I (Pob.)	15.38956	121.00637
Bilon	PH139907003	Bacolod	15.38984	120.98114
Bilon	PH139907009	Zone VI (Pob.)	15.39777	120.98938
Bilon	PH139907015	Burabod	15.40817	121.03665

9. Repeat steps 2 to 8 for the latitudes and longitudes of the municipality of Guinotaban. In this case, all coordinates fall in the area:

Guinotaban	PH139904024	Malipo	14.32559	120.96507
Guinotaban	PH139904036	Palanas	14.33679	120.98815
Guinotaban	PH139904046	Sinungtan	14.34103	120.97309
Guinotaban	PH139904006	Bololo	14.34439	120.94905
Guinotaban	PH139904025	Malobago	14.35216	120.95519
Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304
Guinotaban	PH139904035	Ongo	14.35847	120.97619
Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810
Guinotaban	PH139904004	Batbat	14.37597	120.95012

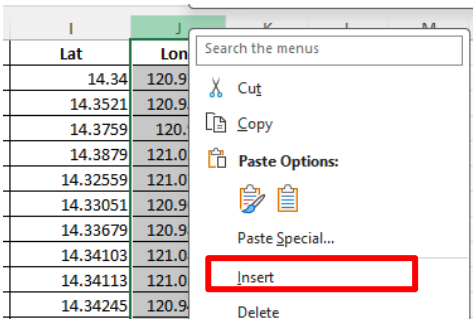
3.7 Identify imprecise latitudes/longitudes

The recommended level of precision for geographic coordinates is the meter [3]. To obtain this level of precision, the coordinates in question must have a minimum of 5 digits after the decimal separator. It is therefore important to identify coordinates with less digits and correct them.

This task can be performed using the LEN function.

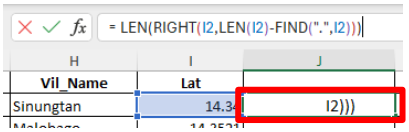
Here is the process to be used to identify latitude/longitude with fewer than 5 decimal digits using the example reported in the “imprecise Lat-long” spreadsheet from the MS Excel example file mentioned in the introduction:

1. Insert a column to the right of column I (*Lat*) by right clicking column J and choosing *Insert*.

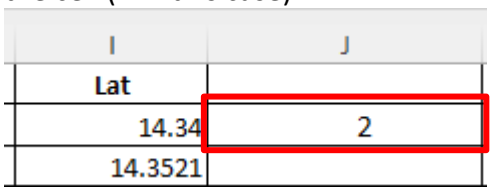


2. Repeat this to add a column to the right of the column containing the longitudes (column K after step 1).
3. In cell J2, use the LEN function as follows to count the number of digits to the right of the decimal point for the latitude contained in cell I2:

   = LEN(RIGHT(I2,LEN(I2)-FIND(".",I2)))



4. Press the *Enter* key on your keyboard. This will bring up the result of the BBCAR function in the cell (2 in this case).



5. To fill the rest of the column with the same formula, double-click the small square in the lower right corner of the cell where you entered the function ( ). You can also click and hold the bottom right corner of the cell and drag it to the last row in the table.
6. Repeat steps 3 to 5 to obtain the number of digits after the decimal separator of the longitude of each record in column L. The formula for cell L2 would be, for example:

   = LEN(RIGHT(K2,LEN(K2)-FIND(".",K2)))

   The number of digits after the decimal point is now visible for the latitude and longitude of each record. You can then always either:
   - Sort the content of the entire table in ascending order by the number of digits after the decimal separator for latitude and then longitude.
   - Use the conditional formatting function to identify cells containing a number of digits lower than 5.

Using the second approach will give the following result where we can see that latitude of 4 records and the longitude of 1 record do not have the required level of accuracy:

H	I	J	K	L
Vil_Name	Lat		Long	
Sinungtan	14.34	2	120.97309	5
Malobago	14.3521	4	120.95519	5
Batbat	14.3759	4	120.9501	4
Mabayawas	14.3879	4	121.02925	5
East Carisac	14.32559	5	121.07202	5
Malipo	14.33051	5	120.96507	5
Palanas	14.33679	5	120.98815	5
West Carisac	14.34103	5	121.08804	5
Magallang	14.34113	5	121.03314	5
Bololo	14.34245	5	120.94905	5
Zone I (Pob.)	14.35493	5	120.99304	5
Ongo	14.35847	5	120.97619	5
Marcial O. Ranola	14.36196	5	120.9781	4
Zone I (Pob.)	14.38956	5	121.00637	5
Bacolod	14.38984	5	120.98114	5
Zone VI (Pob.)	14.39777	5	120.98938	5
Burabod	14.40817	5	121.03665	5
Zone V (Pob.)	14.414791	6	120.99752	5

It is important to note that the 4th and 5th decimal digits in a geographic coordinate reading can be zeros. However, this is rarely the case for both the latitude and longitude at the same time and/or all numbers from the 2nd decimal place.

Consequently, any geographic coordinate containing a latitude and/or longitude with only one (1) or two (2) decimal places should not be considered. This is for example the case of the first record in the table above.

For other records, the best practice is to reject geographic coordinates for which both the latitude and longitude have less than 5 digits after the decimal separator. This is for example the case of the third record from the top in the table above.

3.8 Identify duplicate records

One of the most common errors encountered in tabular data is the presence of record duplicates.

Identifying and then removing these duplicates is important, especially in the case of master lists.

This can be done in two ways

- Using the Conditional Formatting function
- Using the IF and COUNTIF functions together

Both options show not only duplicate records but also the first instance in which the value appeared.

The next sections describe the process for these two options using the examples contained in the “Identify duplicates” worksheet in the MS Excel examples file mentioned in the introduction.

Using Conditional Formatting

The quickest way to identify duplicates is to use the conditional formatting feature. This function is a built-in tool in MS Excel that highlights all cells with the same values. This approach does not require the user to enter a formula.

Here is the process to follow using the top table in the spreadsheet mentioned previously:

1. Select the content of the entire table.

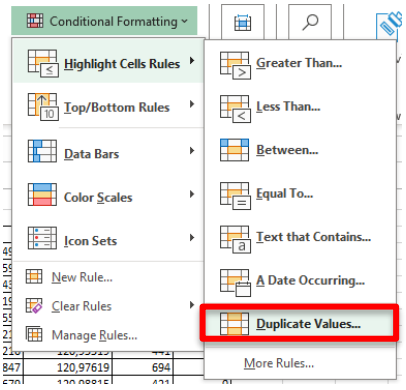
Conditional Formating function

Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465	0
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904036	Palanas	14.33679	120.98815	421	0
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	14.38984	120.98114	724	57
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	14.38956	121.00637	1145	145
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.41479	120.99752	947	91
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	14.39777	120.98938	932	87
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	14.33051	121.07202	682	41
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	14.34113	121.08804	284	0
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	14.38790	121.02925	2019	89
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	14.34245	121.03314	633	5

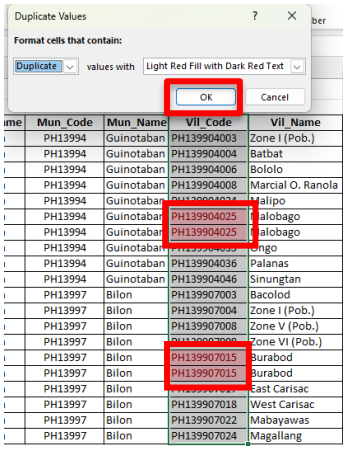
2. Sort the content of the table according to the column containing the information you want to use to identify duplicates (please refer to section 3.5 for details on how to sort). In this example, we will use column G containing the unique identifier of each village (*Vil_Code*). Doing this will allow the duplicates to be next to each other.
3. Once the table is sorted, select the content of the table located in column H (*Vil_Code*)

Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name
Tolkien	PH13994	Guinotaba	PH139904003	Zone I (Pob.)
Tolkien	PH13994	Guinotaba	PH139904004	Batbat
Tolkien	PH13994	Guinotaba	PH139904006	Bololo
Tolkien	PH13994	Guinotaba	PH139904008	Marcial O. Ranol
Tolkien	PH13994	Guinotaba	PH139904024	Malipo
Tolkien	PH13994	Guinotaba	PH139904025	Malobago
Tolkien	PH13994	Guinotaba	PH139904025	Malobago
Tolkien	PH13994	Guinotaba	PH139904035	Ongo
Tolkien	PH13994	Guinotaba	PH139904036	Palanas
Tolkien	PH13994	Guinotaba	PH139904046	Sinungtan
Tolkien	PH13997	Bilon	PH139907003	Bacolod
Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)
Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)
Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)
Tolkien	PH13997	Bilon	PH139907015	Burabod
Tolkien	PH13997	Bilon	PH139907015	Burabod
Tolkien	PH13997	Bilon	PH139907017	East Carisac
Tolkien	PH13997	Bilon	PH139907018	West Carisac
Tolkien	PH13997	Bilon	PH139907022	Mabayawas
Tolkien	PH13997	Bilon	PH139907024	Magallang

4. Under the *Home menu*, click *Conditional Formatting* in the *Styles section and then click Cell Highlighting* Rules in the first drop-down menu, and then choose *Duplicate Values* in the second



5. *The Duplicate Values* dialog box opens and the cells containing the duplicates are directly highlighted in light red with dark red text. Click *OK* to accept the default.



Mun_Code	Mun_Name	Vil_Code	Vil_Name
PH13994	Guinotaban	PH139904003	Zone I (Pob.)
PH13994	Guinotaban	PH139904004	Batbat
PH13994	Guinotaban	PH139904006	Bololo
PH13994	Guinotaban	PH139904008	Marcial O. Ranola
PH13994	Guinotaban	PH139904024	Malipo
PH13994	Guinotaban	PH139904025	Malobago
PH13994	Guinotaban	PH139904025	Malobago
PH13994	Guinotaban	PH139904035	Ongo
PH13994	Guinotaban	PH139904036	Palanas
PH13994	Guinotaban	PH139904046	Sinungtan
PH13997	Bilon	PH139907003	Bacolod
PH13997	Bilon	PH139907004	Zone I (Pob.)
PH13997	Bilon	PH139907008	Zone V (Pob.)
PH13997	Bilon	PH139907009	Zone VI (Pob.)
PH13997	Bilon	PH139907015	Burabod
PH13997	Bilon	PH139907015	Burabod
PH13997	Bilon	PH139907017	East Carisac
PH13997	Bilon	PH139907018	West Carisac
PH13997	Bilon	PH139907022	Mabayawas
PH13997	Bilon	PH139907024	Magallang

Using the IF and COUNTIF functions

The combination of IF and COUNTIF functions can also be used together to identify duplicates in a particular column. More precisely:

- The COUNTIF function is used to identify if given information appears more than once among the selected records
- The IF function itself is used to highlight cells containing information appearing multiple times.

The formula used based on these two functions to identify duplicates is as follows:

= IF(COUNTIF (range, criteria)>1,"Duplicate","")

With:

- *range*: The set of cells for which we are trying to identify duplicates. In this example, this corresponds to the cells containing the unique identifier of each village.
- *criteria*: The value we analyze in the range. This corresponds here to the unique identifier of the village for the record on which the formula is applied.
- *>1*: The condition to which the COUNTIF result must conform for the IF function to return " *Duplicate"* in the cell. If this condition is not met, the IF function will return *an* empty cell.

Here is the process to follow using the bottom table in the spreadsheet mentioned previously:

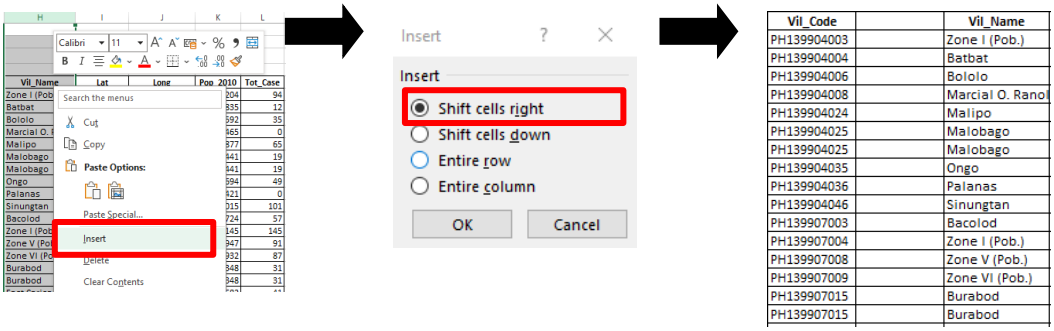
1. Select the entire content of the table.

IF and COUNTIF functions

Reg_Code	Reg_Name	Pro_Code	Pro_Name	Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010	Tot_Case
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204	94
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835	12
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592	35
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465	0
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877	65
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441	19
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694	49
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904036	Palanas	14.33679	120.98815	421	0
PH13	National Capital Region	PH1399	Tolkien	PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015	101
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907003	Bacolod	14.38984	120.98114	724	57
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907004	Zone I (Pob.)	14.38956	121.00637	1145	145
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907008	Zone V (Pob.)	14.41479	120.99752	947	91
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907009	Zone VI (Pob.)	14.39777	120.98938	932	87
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348	31
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907017	East Carisac	14.33051	121.07202	682	41
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907018	West Carisac	14.34113	121.08804	284	0
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907022	Mabayawas	14.38790	121.02925	2019	89
PH13	National Capital Region	PH1399	Tolkien	PH13997	Bilon	PH139907024	Magallang	14.34245	121.03314	633	5

2. Sort the content of the table according to the column containing the information you want to use to identify duplicates (please refer to section 3.5 for details on how to sort). In this example we will use column G containing the unique identifier of each village (*Vil_Code*). Doing this will allow the duplicates to be next to each other.

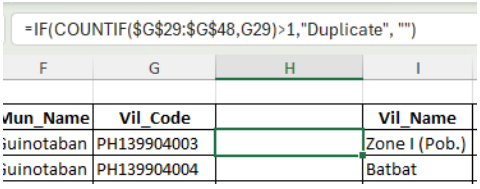
3. Insert cells to the right of those containing the unique identifier of each village by selecting the cells (not the column) containing the *Vil_Name* field and those containing the name of each village, then right-clicking and choosing *Insert*. In the window that opens choose the option *Shift cells right*.

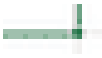


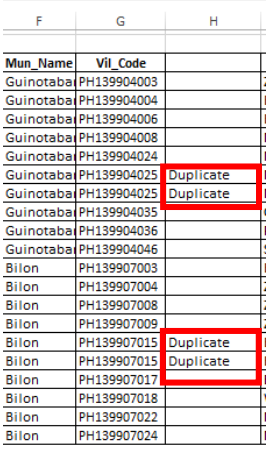
Vil_Code		Vil_Name
PH139904003		Zone I (Pob.)
PH139904004		Batbat
PH139904006		Bololo
PH139904008		Marcial O. Ranol
PH139904024		Malipo
PH139904025		Malobago
PH139904025		Malobago
PH139904035		Ongo
PH139904036		Palanas
PH139904046		Sinungtan
PH139907003		Bacolod
PH139907004		Zone I (Pob.)
PH139907008		Zone V (Pob.)
PH139907009		Zone VI (Pob.)
PH139907015		Burabod
PH139907015		Burabod

4. In cell H29, enter the following formula based on the IF and COUNTIF functions (using the "$" sign in front of the character and cell number will freeze the formula on that range of cells when copying the formula in other cells):

   =IF(COUNTIF($G$29:$G$48,G29)>1,"Duplicate", "")



5. Press the Enter key on your keyboard. This will execute the formula and return the results of the function.
6. Double-click the small square in the lower right corner of the cell where you entered the function (). This will fill the rest of the column with the same formula (or click and hold on the bottom right corner of the cell and drag to the last row of the table). The word " *Duplicate* " will now appear to the right of each record for which one or more unique identifier duplicates exist.



Mun_Name	Vil_Code	
Guinotaba	PH139904003	
Guinotaba	PH139904004	
Guinotaba	PH139904006	
Guinotaba	PH139904008	
Guinotaba	PH139904024	
Guinotaba	PH139904025	Duplicate
Guinotaba	PH139904025	Duplicate
Guinotaba	PH139904035	
Guinotaba	PH139904036	
Guinotaba	PH139904046	
Bilon	PH139907003	
Bilon	PH139907004	
Bilon	PH139907008	
Bilon	PH139907009	
Bilon	PH139907015	Duplicate
Bilon	PH139907015	Duplicate
Bilon	PH139907017	
Bilon	PH139907018	
Bilon	PH139907022	
Bilon	PH139907024	

3.9 Join the content of one table to another

Data you may want to have in a single table can sometimes be stored in different tables. If the tables in question contain the same unique identifier for each record, then it is possible to join their content.

This task can be performed using the VLOOKUP function. This function is called via the following formula:

= VLOOKUP (search_value,matrix_table,col_index_no,[near_value])

With:

- *searched_value* : Value you want to search for.
- *Table_matrix* : Range of cells in which the VLOOKUP function will search for the search_value argument and the return value.
- *no _index_col* : Column number which contains the return value (the leftmost column of the matrix_table corresponding to 1).
- *Close_value :* [optional] Represents a logical value indicating whether you want the VLOOKUP function to search for an exact or an approximate match. TRUE = approximate match (default). FALSE = exact match.

In the example used here, we will use this function to attach the population of the villages to the other information contained in the master list of villages.

Here is the process to use to achieve this result using the contents of the "Join tables" worksheet from the MS Excel examples file mentioned in the introduction:

1. Label the column to the right of the *Long column* in the left table (master list (primary table)) *Pop_2010*.

G	H	I	J	K
Vil_Code	Vil_Name	Lat	Long	POP_2010
PH139907003	Bacolod	14,38984	120,98114	
PH139904004	Batbat	14,37597	120,95012	
PH139904006	Bololo	14,34439	120,94905	
PH139907015	Burabod	14,40817	121,03665	
PH139907017	East Carisac	14,33051	121,07202	

2. In cell K4, enter the following formula based on the VLOOKUP function:

   =VLOOKUP(G4,P:Q,2,FALSE)

   What this formula allows you to do is take the value contained in cell G4 (PH139907003) and see if this value is found in the first column of the range made up of columns P and Q (P:Q).
   - If this is the case, the value contained in the second column of the range (column Q) will be reported in cell K4
   - If it is not, #N/A (FALSE) will be indicated in cell K4

   In the present case, the value PH139907003 is indeed in column P:

P	Q
Secondary table	
Vil_Code	Pop_2010
PH139904003	1204
PH139904004	835
PH139904006	592
PH139907003	724
PH139904008	465
PH139904024	877
PH139904025	441

3. By pressing the Enter key on your keyboard, the value associated with the unique identifier PH139907003 in column Q will appear in cell K4:

Lat	Long	POP_2010
14.38984	120.98114	724
14.37597	120.95012	
14.34439	120.94905	

4. Double-click the small square in the lower right corner of the cell where you entered the function (). This will fill the rest of the column with the same formula and thus the VLOOKUP result (or click and hold on the bottom right corner of the cell and drag to the last row of the table).

Vil_Name	Lat	Long	Pop_2010
Bacolod	14.38984	120.98114	724
Batbat	14.37597	120.95012	835
Bololo	14.34439	120.94905	592
Burabod	14.40817	121.03665	348
East Carisac	14.33051	121.07202	682
Mabayawas	14.38790	121.02925	2019
Magallang	14.34245	121.03314	633
Malipo	14.32559	120.96507	877
Malobago	14.35216	120.95519	441
Marcial O. Ranola	14.36196	120.97810	465
Ongo	14.35847	120.97619	694
Palanas	14.33679	120.98815	421
Sinungtan	14.34103	120.97309	1015
West Carisac	14.34113	121.08804	284
Zone I (Pob.)	14.35493	120.99304	1204
Zone I (Pob.)	14.38956	121.00637	1145
Zone V (Pob.)	14.41479	120.99752	947
Zone VI (Pob.)	14.39777	120.98938	932

Note: the XLOOKUP function allows to do the same as what is presented here above but with more flexibility than the VLOOKUP function as it allows to search in any direction (left or right) and defaults to an exact match, while VLOOKUP primarily searches right from the leftmost column and requires an explicit setting for exact matches. Please refer to the webpage mentioned in Annex 1 for more details about this function

3.10 Check completeness against the master list

The process, and therefore the same function as described in section 3.9 (VLOOKUP), can also be used to check whether a table contains all the geographic features reported in the corresponding master list [5]. This test is important to ensure the completeness of your data set.

In this example, we will check if the table containing the number of cases per village contains all the villages reported in the master list. This example is included in the "Verify exhaustivity" worksheet in the MS Excel Examples file mentioned in the introduction:

1. Label one of the columns next to column H (*Vil_Name*) of the master list as *Ver_Completeness*

G	H	I	J
Vil_Code	Vil_Name		Ver_Completeness
PH139904003	Zone I (Pob.)		
PH139904004	Batbat		
PH139904006	Bololo		
PH139904008	Marcial O. Ranola		

2. In cell J5, enter the following formula based on the VLOOKUP function:

   =VLOOKUP(G5,N:O,2,FALSE)

[5] Single, authoritative list maintained by the officially mandated agency, complete, up-to-date and uniquely encoded of all active (and past) records for a type of entity/geographic object (e.g. health facilities, divisions administrative, villages)

What this formula allows you to do is take the value contained in cell G5 (PH139904003) and see if this value is found in the first column of the range made up of columns N and O (N:O).

- If this is the case, the value contained in the second column of the range (column O) will be reported in cell J5
- If it is not, #N/A (FALSE) will be indicated in cell J5

In this case, the value PH139904003 is indeed in column N:

N	O
Number of cases per village (	
Vil_Code	**Vil_Name**
PH139904003	Zone I (Pob.)
PH139904004	Batbat
PH139904006	Bololo
PH139904008	Marcial O. Ranola

3. By pressing the Enter key on your keyboard, the value associated with the unique identifier PH139904003 in column O will appear in cell J5:

Vil_Code	Vil_Name		Ver_Completeness
PH139904003	Zone I (Pob.)		Zone I (Pob.)
PH139904004	Batbat		

4. Click and hold the small square in the lower right corner of the cell where you entered the function ( ) and drag the cursor down to the last row of the table. This will fill the rest of the column with the same formula and the VLOOKUP result.

Vil_Code	Vil_Name		Ver_Completeness
PH139904003	Zone I (Pob.)		Zone I (Pob.)
PH139904004	Batbat		Batbat
PH139904006	Bololo		Bololo
PH139904008	Marcial O. Ranola		Marcial O. Ranola
PH139904024	Malipo		Malipo
PH139904025	Malobago		Malobago
PH139904035	Ongo		Ongo
PH139904036	Palanas		Palanas
PH139904046	Sinungtan		Sinungtan
PH139907003	Bacolod		Bacolud
PH139907004	Zone I (Pob.)		Zone I (Pob.)
PH139907008	Zone V (Pob.)		Zone V (Pob.)
PH139907009	Zone VI (Pob.)		Zone VI (Pob.)
PH139907015	Burabod		Burabod
PH139907017	East Carisac		East Carisac
PH139907018	West Carisac		West Carisac
PH139907021	Linao		#N/A
PH139907022	Mabayawas		Mabayawas
PH139907024	Magallang		Magallang
PH139907026	Marayag		#N/A

The value "#N/A" in the *Ver_Completeness column* indicates records in the master list for which there is no value in the table containing the number of cases.

This is the case of two villages in this example (Linao and Marayag). The table of the number of cases at the village level is therefore incomplete compared to the master list.

It is important to note that to be valid this test must be carried out based on the master list having the same temporal validity as the programmatic data (number of cases per village in the present case). In other words, if programmatic data captures the number of cases in villages for January 2023, the master list of villages to be used should be representative of the situation observed for that same month. If this is not the case, the identified difference may be due to actual changes that occurred between the time the master list was last updated and the programmatic data collected.

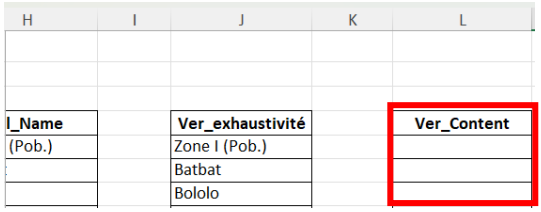
3.11 Identify differences of content between two columns

Identifying potential differences of content between two columns, especially after performing a join using the VLOOKUP function, helps ensure consistency between different tables. Such identification is carried out using the IF function.

In this example, we will check if the spelling of the name of the village indicated in the table of cases per village used in section 3.9 matches the official spelling indicated in the master list of villages.

Here is the process to use to compare the content of these two columns using the example included in the "Verify exhaustivity" worksheet in the MS Excel examples file mentioned in the introduction:

1. Label one of the columns next to the *Ver_completeness column* from the master list *Ver_ content*.



H	I	J	K	L
l_Name		**Ver_exhaustivité**		**Ver_Content**
(Pob.)		Zone I (Pob.)		
		Batbat		
		Bololo		

2. In cell L5, enter the following formula based on the IF function:

   = IF(J5=H5,"","Different")

3. Press the Enter key on your keyboard. This will run the IF function which will test if the content of cell J5 is the same or different from the content of cell H5. If so, the function will return " " (empty cell). Otherwise, the text "Different" will appear in the cell.
4. Click and hold the small square in the lower right corner of the cell where you entered the function ( ) and drag the cursor down to the last row of the table. This will fill the rest of the column with the same formula and therefore the result of the IF-based formula.

H	I	J	K	L
Vil_Name		**Ver_Completeness**		**Ver_Content**
Zone I (Pob.)		Zone I (Pob.)		
Batbat		Batbat		
Bololo		Bololo		
Marcial O. Ranola		Marcial O. Ranola		
Malipo		Malipo		
Malobago		Malobago		
Ongo		Ongo		
Palanas		Palanas		
Sinungtan		Sinungtan		
Bacolod		Bacolud		Different
Zone I (Pob.)		Zone I (Pob.)		
Zone V (Pob.)		Zone V (Pob.)		
Zone VI (Pob.)		Zone VI (Pob.)		
Burabod		Burabod		
East Carisac		East Carisac		
West Carisac		West Carisac		
Linao		#N/A		#N/A
Mabayawas		Mabayawas		
Magallang		Magallang		
Marayag		#N/A		#N/A

In the present case, we can observe that the use of the IF function returned:

- A blank cell for most villages meaning that, in these cases, the spelling of the village name contained in the number of cases table matches the spelling contained in the master list.
- The "Different" text for a record. In this case, the spelling of the name of the village in the table of case numbers (Bacolud) is different from that reported in the master list (Bacolod)
- #N/A for the two villages that are in the master list but not in the case count table.

3.12 Summarize the content of a given column

It is sometimes useful to check whether the number of geographic features in an administrative unit is correct or to calculate how many of these features are of a certain type according to a predefined classification and place this information in a table.

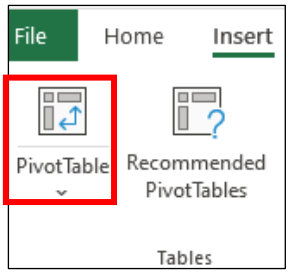
This task can be done using the *PivotTable* tool in MS Excel.

In this example, we will count how many villages are in each of the municipalities included in the example reported in the "Summary column content" worksheet of the MS Excel examples file mentioned in the introduction:

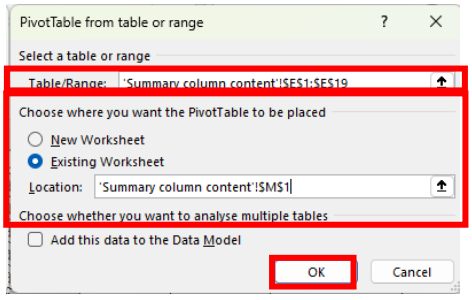
1. Select the column, or range, containing the information to summarize. In this case, it is the cells containing either the name or the unique identifier of the municipalities (*Mun_code* or *Mun_Name*) because the table contains the list of villages for each municipality. By calculating the number of records per municipality we will indirectly obtain the number of villages in each of them.

D	E	F
Pro_Name	Mun_Code	Mun_Name
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13994	Guinotaban
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon
Tolkien	PH13997	Bilon

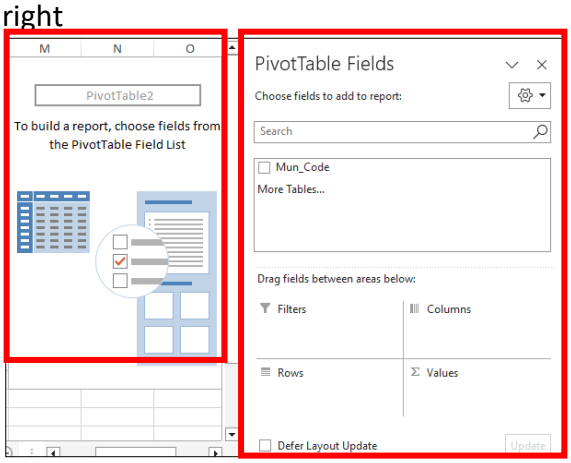
2. Under the *Insert menu*, click on the *Pivot Table icon* in the *Tables* section



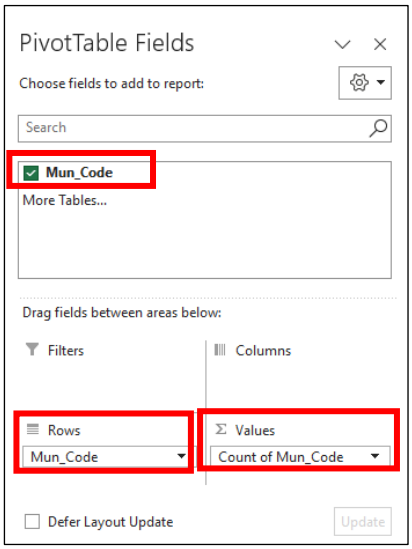
3. In the window that opens
    a. Table/Range: keep the information as such (column you selected)
    b. Choose where you want to place the pivot table: choose the *Existing Worksheet* option
    c. Location: click in the field then on cell M1 in the spreadsheet



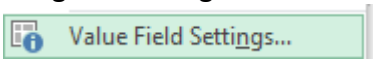
4. Click OK. This will bring up the pivot table in the worksheet and the panel for this tool on the right



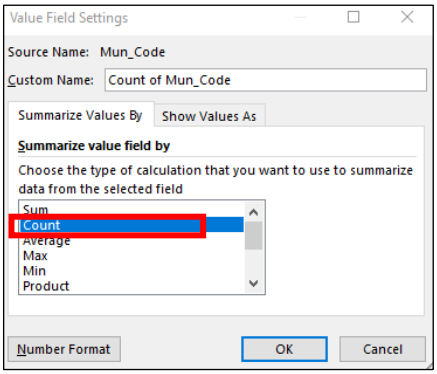
5. In this panel:
    a. Check the *Mun_Code* box in the top window. This will make this field appear in the *Row area* at the bottom left.
    b. *Right click* on *Mun_Code* In the top window, hold click and drag to the *Values* box in the bottom right area.



6. Check that the parameter for the value field is "Count" as is the case in the example above. If not, click on the small triangle to the right of the field 

and select the *Value Field Settings* option. 

. In the window that opens, select the *Count* option and click OK



7. Click in the spreadsheet to see the resulting pivot table

M	N
Row Labels	**Count of Mun_Code**
PH13994	9
PH13997	9
Grand Total	**18**

As we can see there are 9 registrations, and therefore villages, in each municipaly.

3.13 Aggregate the content of a given column

The values you are working with may sometimes need to be aggregated to be represented on a map at a different level of the administrative structure for example (e.g. province instead of district).

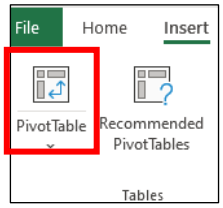
This task can also be done using the PivotTable tool.

We will aggregate the population by village at the municipality level for the year 2010 using the example reported in the "Aggregate column content" worksheet of the MS Excel examples file mentioned in the introduction:

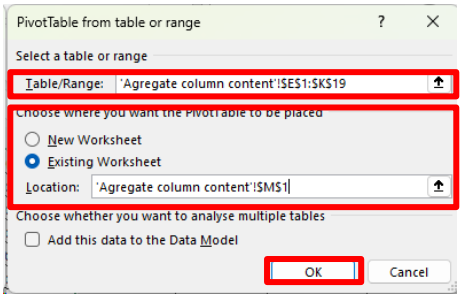
1. Select columns E to K in the spreadsheet (we will use the unique identifier of each municipality in the summary table because this will facilitate joining with other tables (see section 3.9) or the corresponding GIS layer).

Mun_Code	Mun_Name	Vil_Code	Vil_Name	Lat	Long	Pop_2010
PH13994	Guinotaban	PH139904003	Zone I (Pob.)	14.35493	120.99304	1204
PH13994	Guinotaban	PH139904004	Batbat	14.37597	120.95012	835
PH13994	Guinotaban	PH139904006	Bololo	14.34439	120.94905	592
PH13994	Guinotaban	PH139904008	Marcial O. Ranola	14.36196	120.97810	465
PH13994	Guinotaban	PH139904024	Malipo	14.32559	120.96507	877
PH13994	Guinotaban	PH139904025	Malobago	14.35216	120.95519	441
PH13994	Guinotaban	PH139904035	Ongo	14.35847	120.97619	694
PH13994	Guinotaban	PH139904036	Palanas	14.33679	120.98815	421
PH13994	Guinotaban	PH139904046	Sinungtan	14.34103	120.97309	1015
PH13997	Bilon	PH139907003	Bacolod	14.38984	120.98114	724
PH13997	Bilon	PH139907004	Zone I (Pob.)	14.38956	121.00637	1145
PH13997	Bilon	PH139907008	Zone V (Pob.)	14.41479	120.99752	947
PH13997	Bilon	PH139907009	Zone VI (Pob.)	14.39777	120.98938	932
PH13997	Bilon	PH139907015	Burabod	14.40817	121.03665	348
PH13997	Bilon	PH139907017	East Carisac	14.33051	121.07202	682
PH13997	Bilon	PH139907018	West Carisac	14.34113	121.08804	284
PH13997	Bilon	PH139907022	Mabayawas	14.38790	121.02925	2019
PH13997	Bilon	PH139907024	Magallang	14.34245	121.03314	633

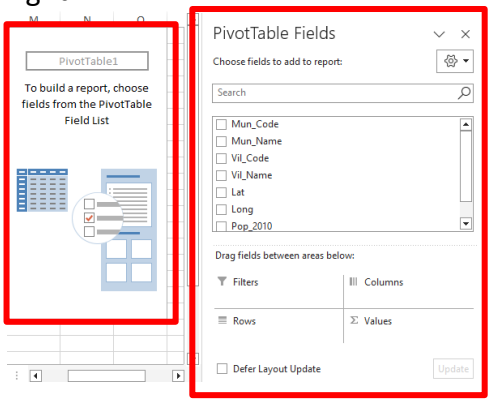
8. Under the *Insert menu*, click on the *Pivot Table icon* in the *Tables section*



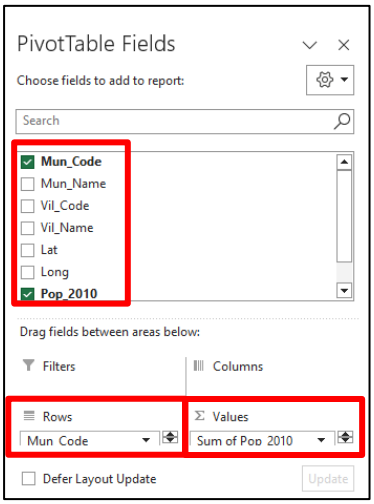
9. In the window that opens
    a. Table/Range: keep the information as is (range you selected)
    b. Choose where you want to place the pivot table: choose the *Existing Worksheet option*
    c. Location: click in the field then on cell M1 in the spreadsheet

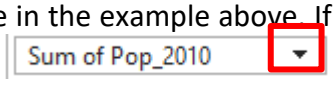


2. Click OK. This will bring up the pivot table in the worksheet and the panel for this tool on the right



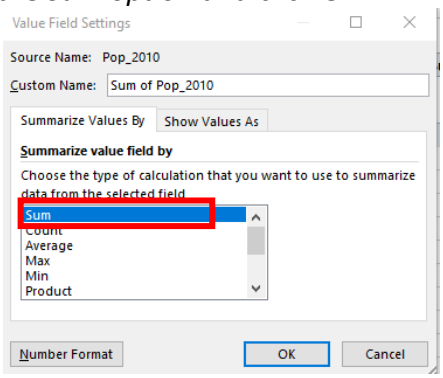
3. In the panel:
    a. Check the *Mun_Code* box in the top window. This will make this field appear in the *Row area* at the bottom left.
    b. *Right-click* on *Pop_2010* in the top window, hold and drag to the *Values* box in the bottom right area.



4. Check that the parameter for the value field is "Sum" as is the case in the example above. If not, click on the small triangle to the right of the field and select the 

Value Field Settings option. 

. In the window that opens, select the *Sum option* and click OK



5. Click in the spreadsheet to see the resulting pivot table

M	N
Row Labels	**Sum of Pop_2010**
PH13994	6544
PH13997	7714
Grand Total	**14258**

3.14 Check for data consistency between records

One of the things to avoid in a table is inconsistencies between records concerning data elements based on a classification table (example: types of health facilities, ownership, names of administrative units…).

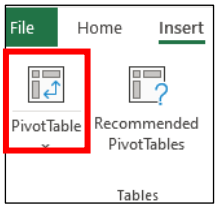
Checking consistency between these records can be done using the *PivotTable tool*.

In this example we will check if the health facility types (HF_Type column) are consistent between all the records in the table reported in the "Verify consistency" spreadsheet of the MS Excel example file mentioned in the introduction:

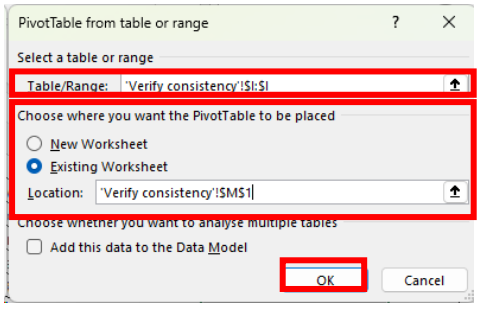
1. Select the column I that contains the health facility types (HF_Type).

I
HF_Type
Provincial Hospital
Hospital
Barangay Health Station
Rural Health Unit
RHU
Barangay Health Station
Barangay Health Station
Barangay Health Station
Barangay Health Station
Rural Health Unit
BHS
BHS
Barangay Health Station
Barangay Health Station
BHS
Rural Health Unit
Barangay Health Station
BHS
Barangay Health Station
Barangay Health Station

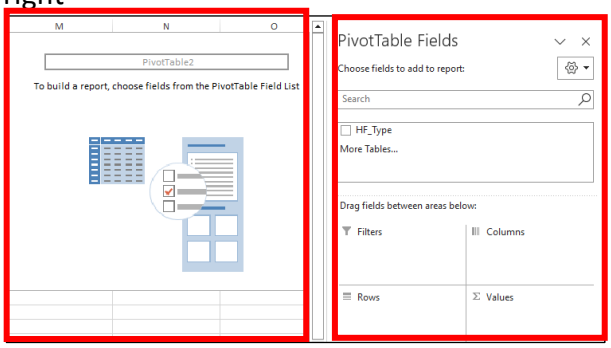
2. Under the *Insert* menu, click on the *Pivot Table icon* in the *Tables section*



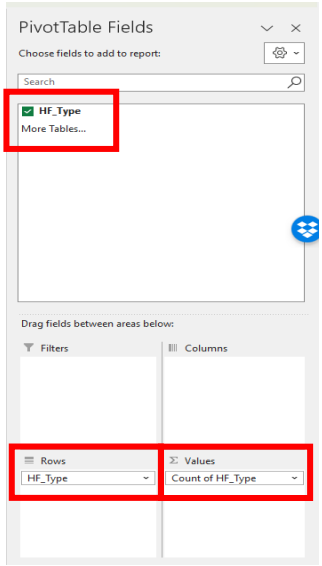
3. In the window that opens
    a. Table/Range: keep the information as such (column you selected)
    b. Choose where you want to place the pivot table: choose the *Existing Worksheet option*
    c. Location: click in the field then on cell M1 in the spreadsheet



4. Click OK. This will bring up the pivot table in the worksheet and the panel for this tool on the right



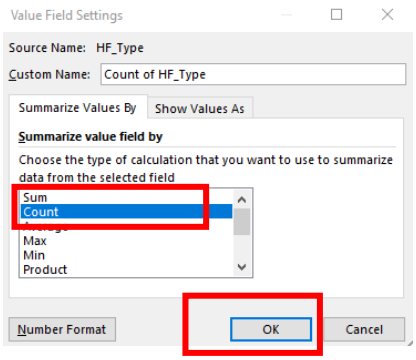
5. In the panel:
    a. Check the *HF_Type box* in the top window. This will make this field appear in the *Row area* at the bottom left.
    b. *Right click* on *HF_Type* In the top window, hold click and drag to the *Values box* in the bottom right area.



6. Check that the parameter for the value field is “Count” as is the case in the example above. If not, click on the small triangle to the right of the field and select the 

Value Field Settings option. 

. In the window that opens, select the *Number option* and click OK



7. Click in the spreadsheet to see the resulting pivot table

M	N
Row Labels	**Count of HF_Type**
Barangay Health Station	10
BHS	4
Hospital	1
Provincial Hospital	1
RHU	1
Rural Health Unit	3
Grand Total	**20**

Several inconsistencies can be noted:

- The Barangay Health Station type is captured in its abbreviation form (BHS) for 4 of the records.
- The same is true for the Rural Health Unit type which presents a record for which the information is captured by an abbreviation (RHU).

The information for these records should be corrected based on the corresponding classification table to allow consistent use.

References

[1] Ebener S. (2016): Guidance for the management and use of geospatial data and technologies in health. Part 1 - Introduction to the data-information-knowledge-decision continuum and the geospatial data management chain. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part1.pdf [Accessed 21 February 2024]

[2] Ebener S. (2016): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.1 Documenting the process and defining the data needs. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_1.pdf [Accessed 21 February 2024]

[3] Ebener S. (2016): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.2 Defining the terminology, data specifications, and the ground reference. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_2.pdf [Accessed 21 February 2024]

[4] Pantanilla I., Ebener S., Mercado C.E., Maude R. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.3 Compiling existing data and identifying gaps. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_3.pdf [Accessed 21 February 2024]

[5] Ebener S. (2021): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.4 Creating geospatial data - 2.4.1 Extracting vector format geospatial data from basemaps. Health GeoLab document: https://healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_4_1.pdf [Accessed 21 February 2024]

[6] Ebener S., Maude R.J., Gault P. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.4 Creating geospatial data - 2.4.2 Collecting data in the field. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_4_2.pdf [Accessed 21 February 2024]

[7] Ebener S., Pantanilla I., Mercado C.E., Maude R. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.5 Cleaning, validating, and documenting the data - 2.5.1 Documenting the data using a metadata profile. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_5_1.pdf [Accessed 21 February 2024]

[8] Pantanilla I., Ebener S., Maude R. (2018): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data. 2.6.1 Creating good thematic maps using desktop GIS software. Health GeoLab document: https://www.healthgeolab.net/DOCUMENTS/Guide_HGL_Part2_6_1.pdf [Accessed 21 February 2024]

[9] Ebener S. (under preparation): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data - 2.6.2 Using thematic maps for decision making. Health GeoLab document.

[10] Ebener S. (under preparation): Guidance for the management and use of geospatial data and technologies in health. Part 2 - Implementing the geospatial data management cycle: 2.6 Distributing, using, and updating the data - 2.6.3 Developing and implementing the appropriate data policy. Health GeoLab document.

Annex 1 – MS Excel functions/tools used in this guide

Function/tools	brief description	Web site
CONCATENATE	Combines text from multiple ranges and/or strings	http://tinyurl.com/3h7nv6ud
FILTER	Allows you to filter a range of data based on criteria you define	http://tinyurl.com/mr2skb8v
LEFT / RIGHT	Returns the last character(s) of a text string, based on the number of characters specified from either side	http://tinyurl.com/5n7s67nt / http://tinyurl.com/2w3v2e44
MIN	Returns the smallest numeric value in a range of values.	http://tinyurl.com/yjjen9yb
MAX	Returns the largest numeric value in a range of values.	http://tinyurl.com/4sfwhs8k
LEN	Returns the number of characters contained in a text string.	http://tinyurl.com/2s39kbn6
COUNTIF	Allows you to count the number of cells that meet a criterion	http://tinyurl.com/3jmadh2b
VLOOKUP	Find items in a table or range by row	http://tinyurl.com/5azjk28k
XLOOKUP	Find items in a table or range by row and this in any direction	https://tinyurl.com/mr3a9a9t
SORT	Sorts the content of a range or table	http://tinyurl.com/4prf92h7
IF	Allows you to make logical comparisons between a value and the expected result	http://tinyurl.com/3ccswt5u
Conditional formatting	Highlight cells with color using rules you set based on the values in those cells.	http://tinyurl.com/3ucxubyc
Pivot table	Tool capable of calculating, synthesizing and analyzing data in tabular form	http://tinyurl.com/y3rdu96k
Text-to-Column wizard	Retrieves and redistributes text in several cells, text from one or more cells	http://tinyurl.com/5n8jmtma