Tutorial Instructions

Preparing Data for Dashboard Visualisations

OCHA Field Information Services (FIS)

Version 1.0 | July 2020

Getting Started

This tutorial is designed to accompany the training webinar *Data Preparation for Dashboards*.

In this tutorial we will transform example data to prepare a dashboard showing people reached in response to an emergency.

We will firstly use Power Query to prepare data on people reached into an appropriate format for inputting to a dashboard. We will then establish a relationship with a separate table of geographical information so we can display information on a map.

What you need before undertaking this tutorial:

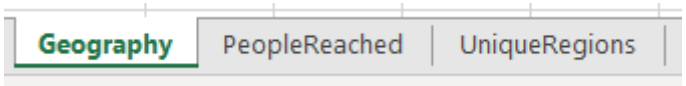
1. Power BI desktop installed, and logged into your organisation (OCHA) account within Power BI desktop
2. Have previously watched the webinar 'Data Preparation for Dashboards'
3. Have a copy of the following files:
    a. Data Preparation for Dashboards - Tutorial.pbix
    b. PeopleReachedAndGeography.xlsx

Tutorial Format

This tutorial is written as a series of steps on the left hand side of the page. Each step has accompanying screenshots on the right hand side to help make the instructions clearer.

Feedback

All feedback to fis-ocha@un.org.

Step	Images
Go to the "TutorialMaterials" folder that was provided to your prior to the tutorial and unzip it. **Open** the Excel spreadsheet "*PeopleReachedAndGeography.xlsx*". It has three worksheets 1) *Geography* 2) *PeopleReached* 3) *UniqueRegions*	
Review the *PeopleReached* worksheet. Based on your understanding of required data structures for dashboards explained in the webinar, identify some required changes.	

	A	B	C	D	E	F	G	H	I	J	K	L
1	STATE	REGION	DISTRICT	District PCODE	Number of people affected	Number of people Displaced	Deaths	Population	Type	June People Reached by district Nutrition	June People Reached by district WASH	June People reached by district Health
2	Fallhill	Brighttow	Coldwater	AA1101	0		0	453434		0	0	0
3	Fallhill	Brighttow	Ironview	AA1102	0		0	99157		0	0	0
4	Fallhill	Brighttow	Wayville	AA1103	0		0	99157		0	0	0
5	Fallhill	Brighttow	Wellfield	AA1104	0		0	72825		0	0	0
6	Fallhill	Newden	Barrowedge	AA1201	600		3	1044086		0	0	0
7	Fallhill	Newden	Greysnow	AA1202	0		0	179997		0	0	0
8	Fallhill	Newden	Wildebourne	AA1203	0		0	97441		0	0	0
9	Fallhill	Woodbutt	Glassfield	AA1301	720		2	498249	Tsunami	0	0	0
10	Fallhill	Woodbutt	Raygate	AA1303	0		0	87386		0	0	0
11	Fallhill	Woodbutt	Springacre	AA1304	0		0	71387		0	0	0
12	Silversage	Eridell	Newbrook	AA1601	91	91	0	467020	Tsunami	0	0	0
13	Silversage	Eridell	Wellfox	AA1602	8000	18	0	15023	Tsunami	0	0	0
14	Silversage	Eridell	Stonewater	AA1603	0		0	51766		0	0	0
15	Silversage	Eridell	Belmere	AA1604	0		0	62264		0	0	0
16	Silversage	Eridell	Highwolf	AA1605	5	5	0	52515		0	0	0
17	Silversage	Eridell	Goldden	AA1606	100000	47000	8	64346	Tsunami	20000	95400	0
18	Silversage	Aldlight	Goldhill	AA1701	0		0	209904		0	0	0
19	Silversage	Aldlight	Highfield	AA1702	0		0	59092		0	0	0
20	Silversage	Aldlight	Moorpond	AA1703	0		0	68593		0	0	0
21	Silversage	Aldlight	Northfalcon	AA1801	53646		0	26906	Tsunami	0	0	0
22	Marshton	Wellspell	Blueby	AA1801	6780		0	377473	Tsunami	0	3720	0
23	Marshton	Wellspel	Deerbarrow	AA1802	5		0	72580		0	0	0
24	Marshton	Wellspell	Coastton	AA1803	50000		2	69851	Tsunami	1500	2000	500

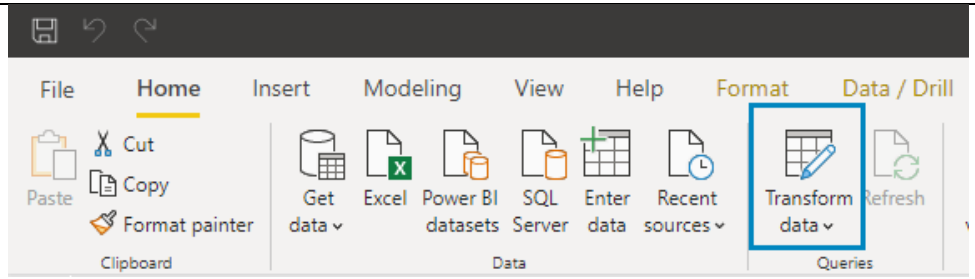
Review the *UniqueRegions* worksheet. As outlined in the webinar, for visualisations it is important that things that are the same are spelt the same. This list was created using the remove duplicates tool in Excel. Reviewing this you can see that later in PowerBI you will need to make the following changes: • Wellspell -> Wellspel • Marblegate -> Marble Gate • Lakemarshs -> Lakemarsh	
Hold Point 0	

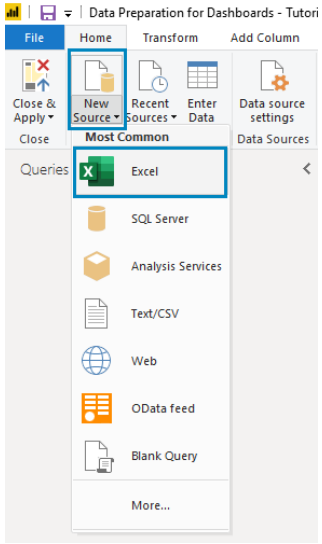
	A
1	Brighttown
2	Newden
3	Woodbutter
4	Eridell
5	Aldlight
6	Wellspell
7	Wellspel
8	Marbleburn
9	Eastloch
10	Corwyn
11	Marblegate
12	Marble Gate
13	Stonewind
14	Havenbush
15	Shoremoor
16	Lakemarshs
17	Lakemarsh
18	Grassedge
19	Northfort
20	Bymeadow
21	Northmeadow
22	East Hills

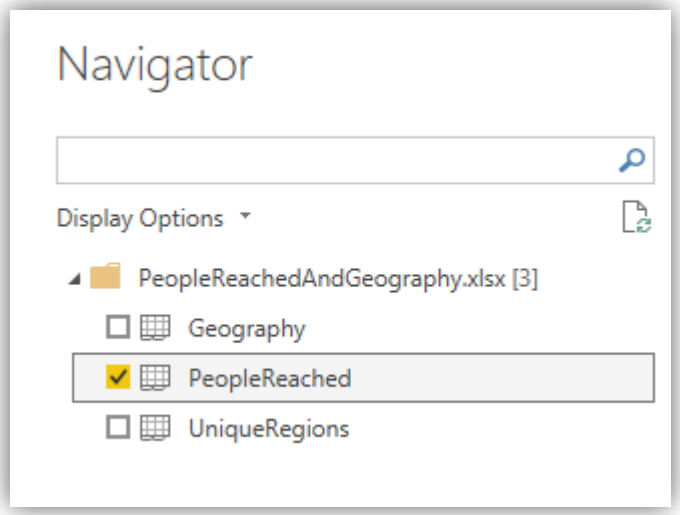
Load the data into Power BI as follows:

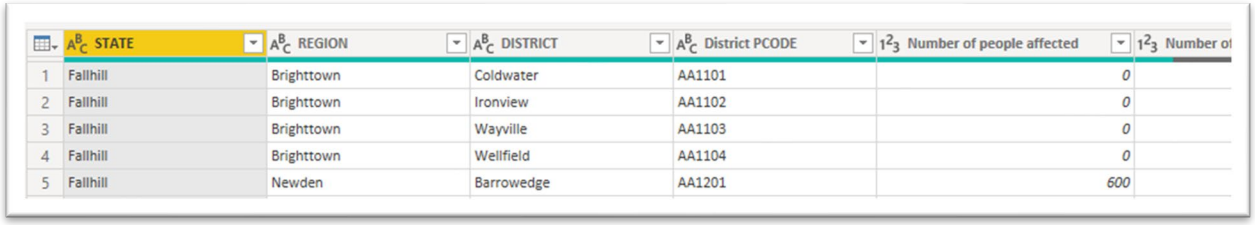
- **Open** Power BI desktop
- **Open** the "Data Preparation for Dashboards - Tutorial.pbix" PowerBI (pbix) report that was provided in the zipped folder
- **Click** "Transform data" in the home ribbon at the top. In some older versions of PowerBI it was called "Edit Queries"
- Now you are in the transform data window select "New Source" and 'Excel"
- Select the spreadsheet "PeopleReachedAndGeography.xlsx' and select the worksheet PeopleReached"

Note there are other ways of bringing in the data in the PowerBI Desktop interface.









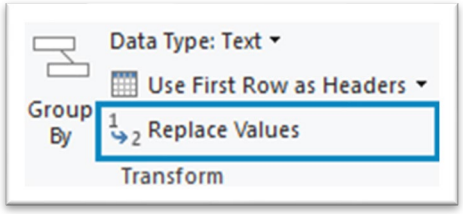
	STATE	REGION	DISTRICT	District PCODE	Number of people affected	Number of
1	Fallhill	Brighttown	Coldwater	AA1101	0	
2	Fallhill	Brighttown	Ironview	AA1102	0	
3	Fallhill	Brighttown	Wayville	AA1103	0	
4	Fallhill	Brighttown	Wellfield	AA1104	0	
5	Fallhill	Newden	Barrowedge	AA1201	600	

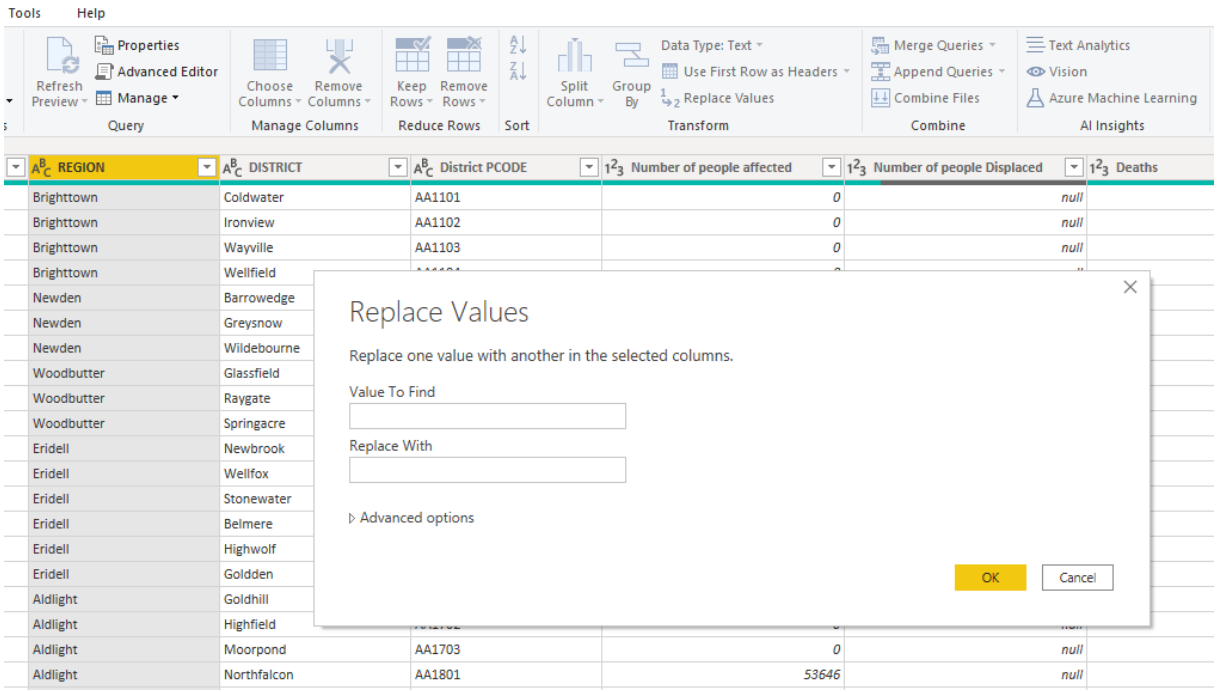
To ensure the quality and accuracy of our visualisations it is important that we have data that is spelt consistently.

Earlier we identified several issues with the *UniqueRegions* worksheet.

We will fix these using Power Query to make the values consistent:

- Recall the spelling changes you need to make based on the earlier step.
- In the PowerBI transform data window click "Replace Values' under the home ribbon after selecting the "REGION' column.
- Perform the replacements:
  Wellspell -> Wellspel
  Marblegate -> Marble Gate
  Lakemarshs -> Lakemarsh
- Review your data

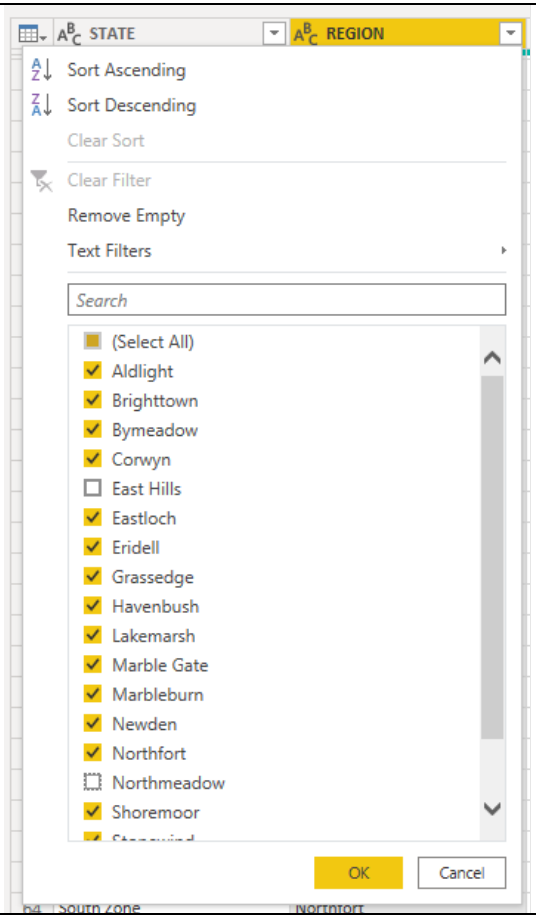




You have been instructed that the data from the regions "Northmeadow" and ‘East Hills" ***cannot be released publicly***.

- At the top of the region column click the filter icon and remove these two regions so that their data is not included in the visualization.

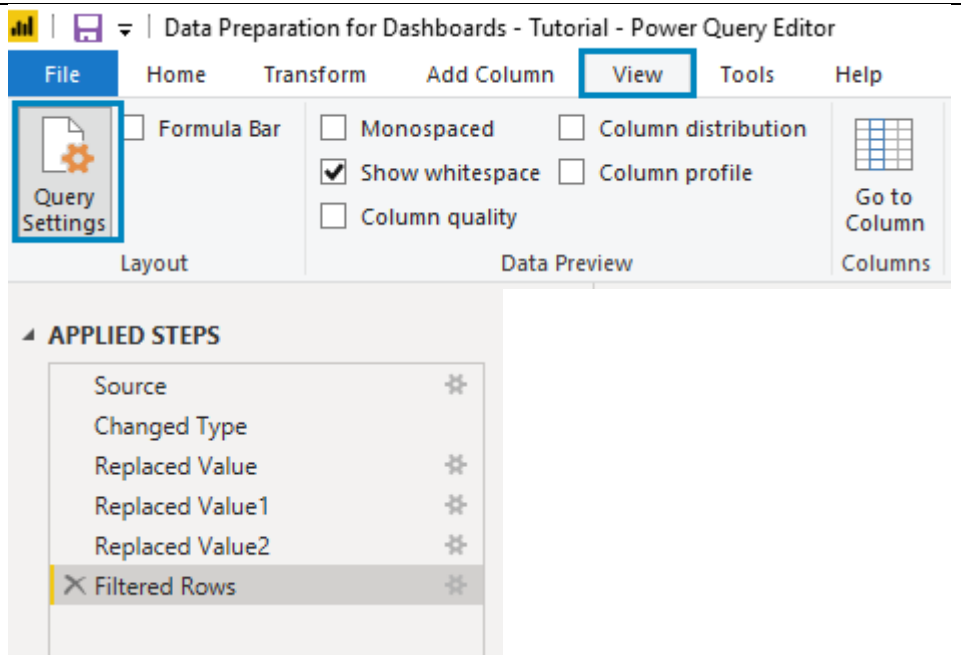
- Click "OK’.



Now that you have completed several data processing steps, look at your applied steps window.

It should look like the adjacent screenshot. Note that you can delete steps by clicking on the cross, or edit by clicking on the gear. You can rename steps.

If your applied steps window does not appear by default. Click view on the ribbon and click "Query Settings".

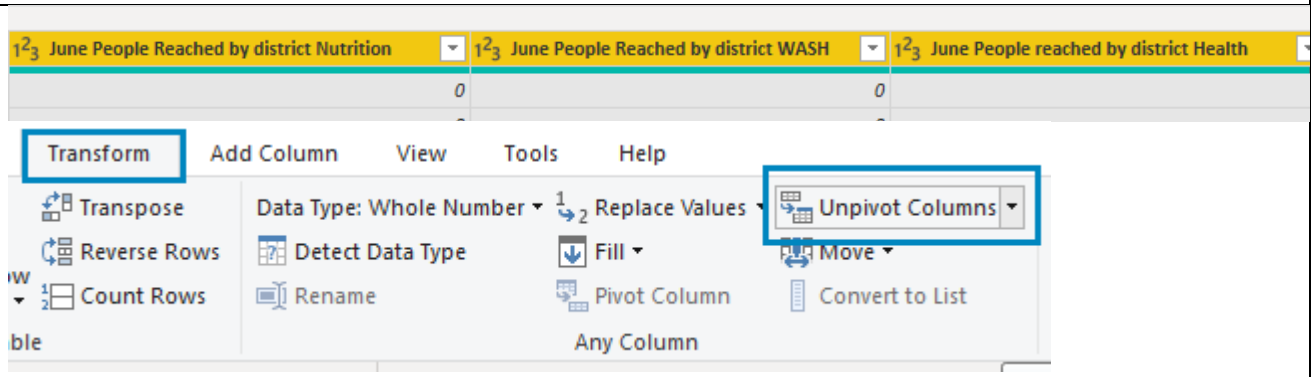


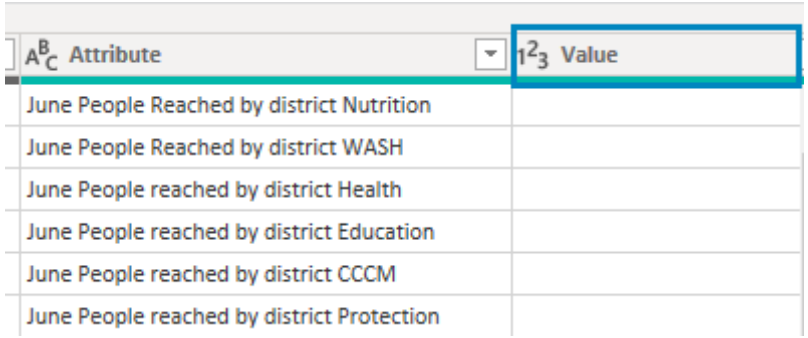
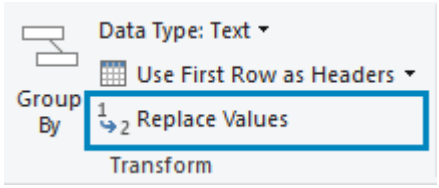
Hold Point 1

We are now going to unpivot the data so that the people reached goes down the page in rows rather than across the page in columns, as this is a preferred data structure for dashboards.

Note that the people reached by sector has columns for both June and July.

- Using the shift key select all of the fields beginning with “June [or July] People Reached____”



• Under the transform ribbon select the function “Unpivot Columns” Now the data has been unpivoted you can change the column name. • Change the column name “*value*” to “People Reached”. **Double-click** on the column name to change it.	
Hold Point 2	
So our data structure now is that we have the people reached in one column, and in another column we have two pieces of data in one column. These are the month and the sector. A dashboard can’t read two data pieces from one text string so we need to separate them. We are going to use the replace function to make it easier for a split function to then be used afterwards. We are going to use replace to get our data into the format [*Month,Sector]* then we just need to split by comma to get two useful columns. But one complication is that in some columns “reached” is capitalised but in another it is not. So, first perform a series of find and replace on the column as follows:	

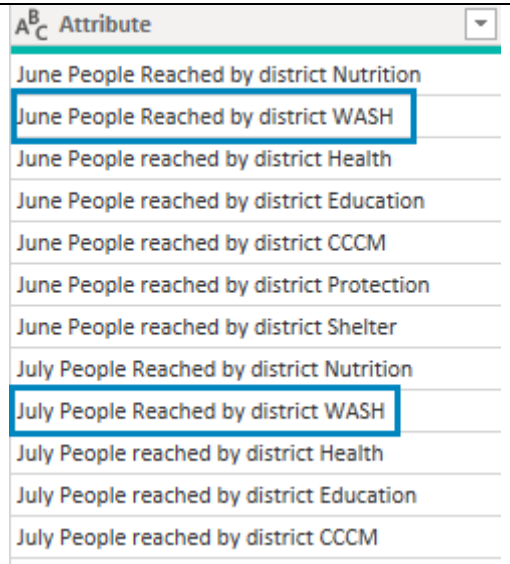
Replace: " People Reached by district "
With: ",". Note there is a space both before "People" and after "district"

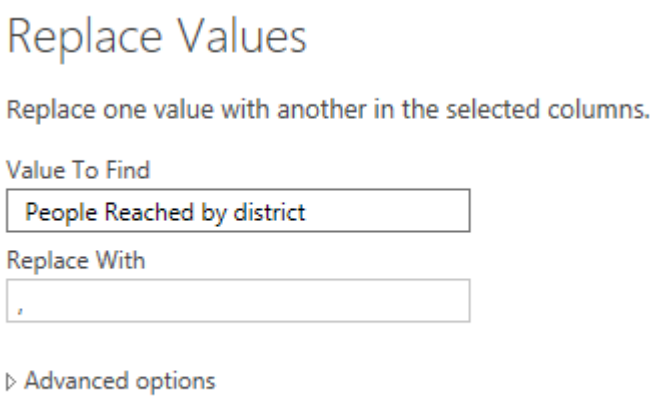
Click "OK".

Do the same again with lower case *reached*:
Replace: " People reached by district "
With: ",".

Click "OK".

Now we can use the split column function to split the column into two columns with a value in each.

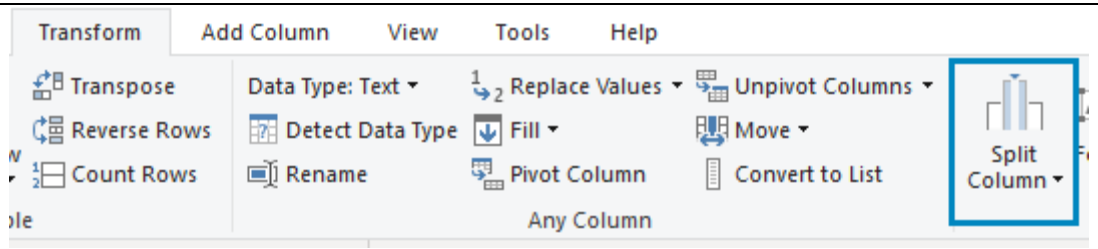


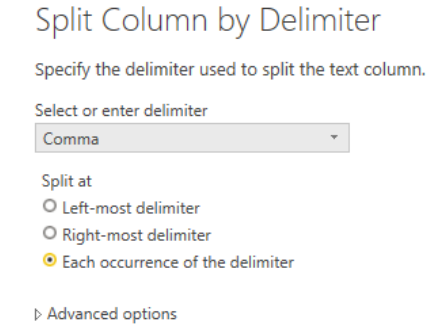


To prepare for this we will split the values into two columns as follows:

- In the transform ribbon, click split column and by Delimeter
- Select "Comma'
- It doesn't matter in this case what option you choose after "Split at"
- Press "OK'

Now the columns of data are separated, we can rename them "Month" and "Sector"





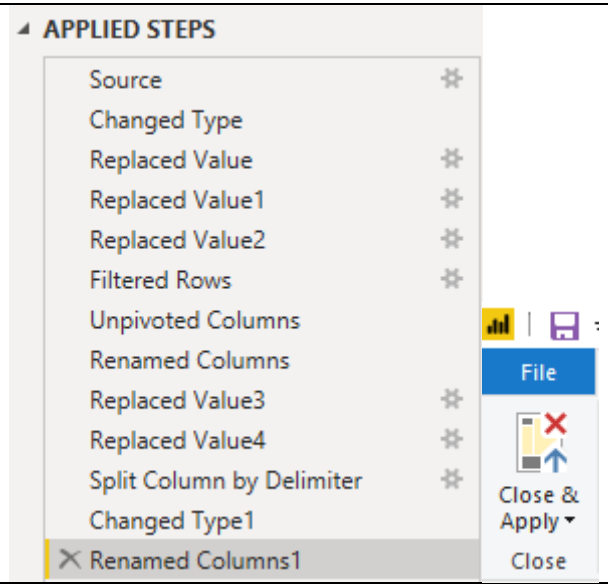
Attribute.1	Attribute.2
June	Nutrition
June	WASH
June	Health
June	Education
June	CCCM
June	Protection
June	Shelter
July	Nutrition
July	WASH
July	Health
July	Education
July	CCCM
July	Protection
July	Shelter

Month	Sector
June	Nutrition
June	WASH
June	Health
June	Education
June	CCCM
June	Protection
June	Shelter
July	Nutrition
July	WASH
July	Health
July	Education
July	CCCM
July	Protection
July	Shelter

Review your applied steps to see all the changes you have made. You can come back to these steps at any time and adjust them or make further edits. Be careful though, because editing earlier steps can break later steps, causing error messages

We are now finished with our work in power query so under the home ribbon

Click "Close"" & apply".



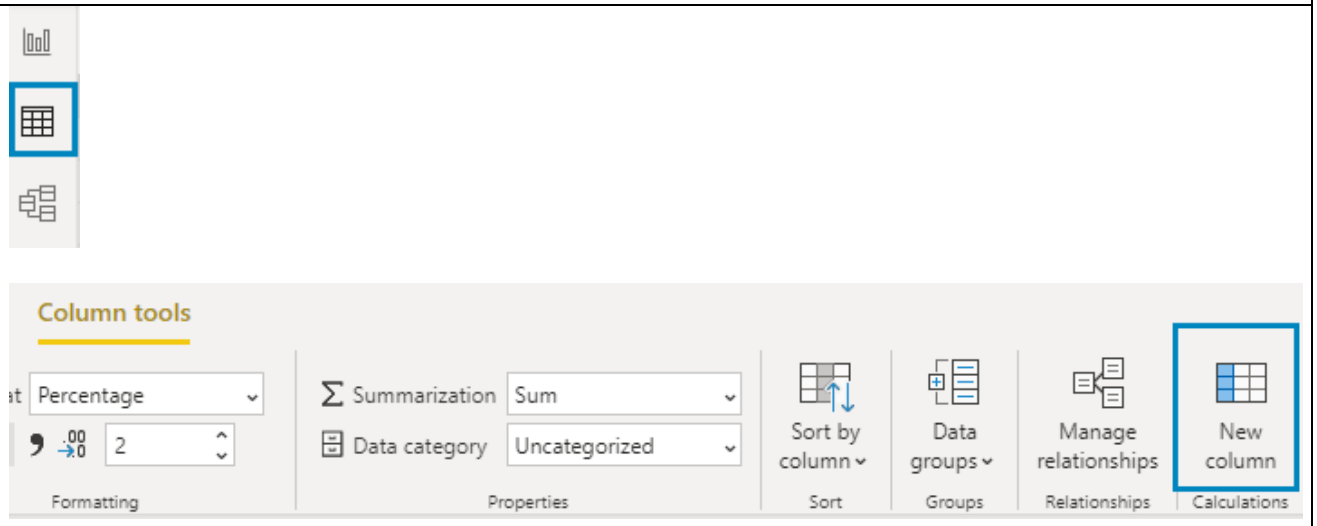
Hold Point 3

Your transformed data is now ready to use in PowerBI.

We could go straight to the report tab to start on visualisations, but we are now going to add an additional calculated column via DAX coding in the data tab.

We are doing this to calculate the percentage of people reached as a proportion of the affected population for each district.

- Select the **Data** tab on the left side of the screen.



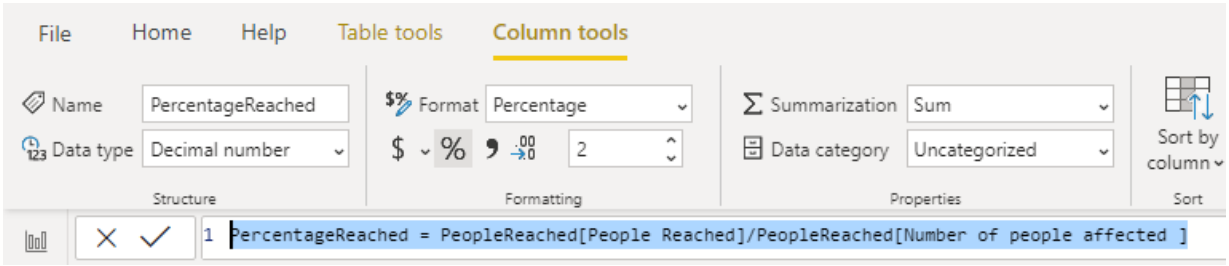
- Click on a column in the table and a new ribbon 'column tools' will appear. Click column tools and **new column.** In previous versions this button was elsewhere on the ribbon.

- In the bar that now appears at the top **copy and paste** the text to the right

- This training is not aiming to be a comprehensive tool for DAX coding; rather this just demonstrates potential capabilities

- With the new column selected select the format as "Percentage" in the column tools ribbon.

```
PercentageReached = PeopleReached[People Reached]/PeopleReached[Number of people affected ]
```



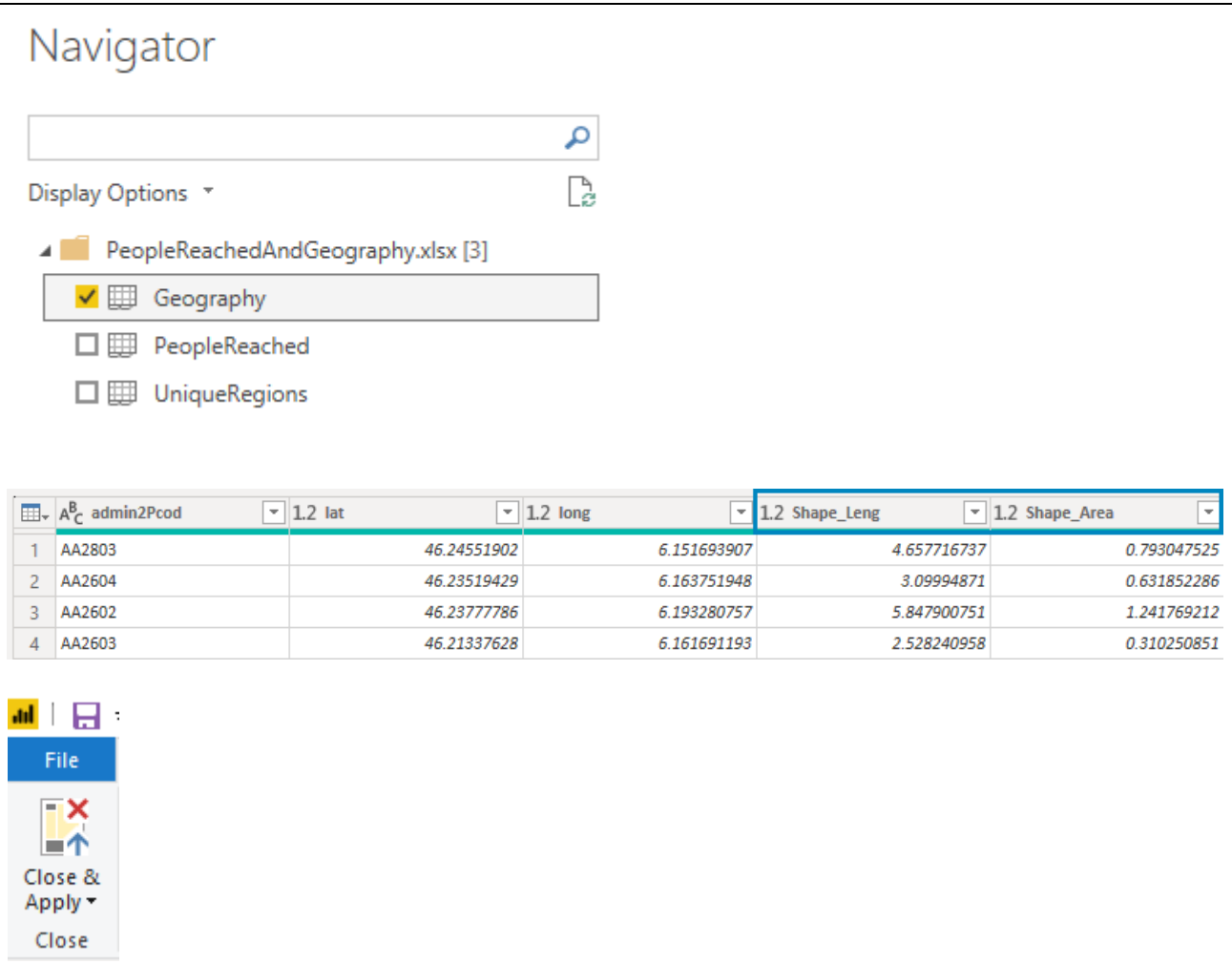
Month	Sector	PercentageReached
June	Shelter	3.53%
July	Nutrition	40.00%
July	WASH	27.00%
June	WASH	54.87%
June	Nutrition	3.00%
June	WASH	4.00%

Hold Point 4

The next step is to bring in some geography data so we know the locations of all the districts so we can show them on the map.

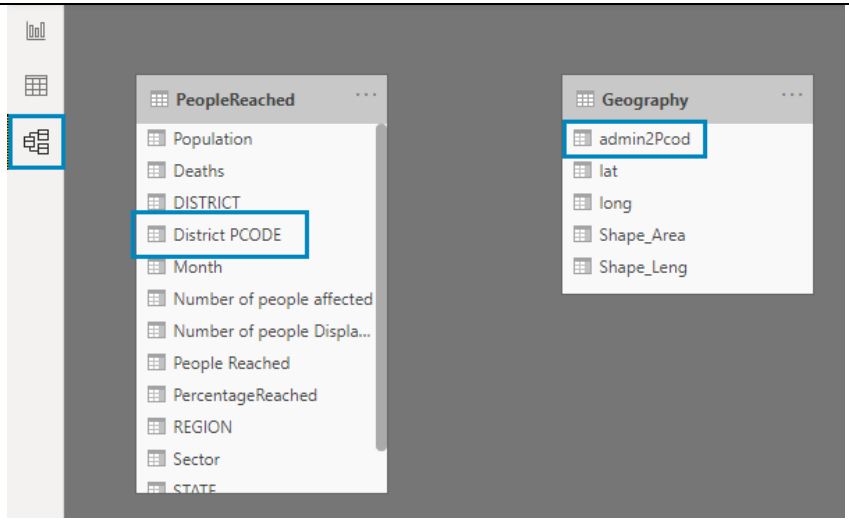
We have P-codes in both datasets so we are going to do a relationship based on this to join the geographic data with the people reached data. We will do this as follows:

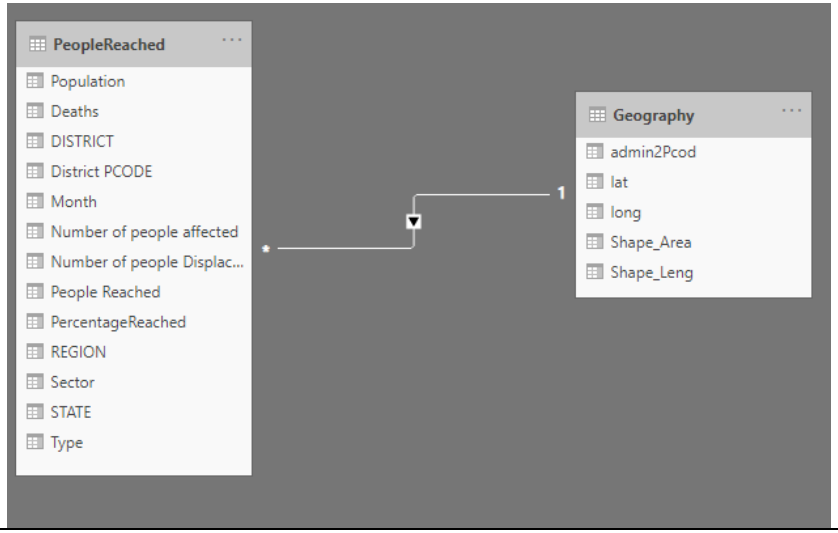
- Click on ‘Transform data’, ‘New Source’, and select the same spreadsheet as before but this time select the “Geography” worksheet.
- We can see two columns of no use to us and left over from GIS processing "Shape_Leng" and "Shape_Area". Select them and choose “Remove Columns”. This only removes the data from the PowerBI model, not changing the original data
- This is the only transformation we need on this so select “Close & Apply”



Now we can see that the P-codes in both data sets match each other so we need to create a relationship between the two as follows:

- Click on the left pane to select the model tab.
- The field “admin2Pcod” in the geography table corresponds to “District PCODE”. Therefore, **drag** from the relevant field on one table and **drop** of the corresponding field of the other table.
- You can now see that a 1 to many relationship is created based on the "1” and the "*".
- Hover over the connection line to make sure it is connecting the right fields.
- You can right-click on the relationship and view its properties. We do not need to change them for this exercise.



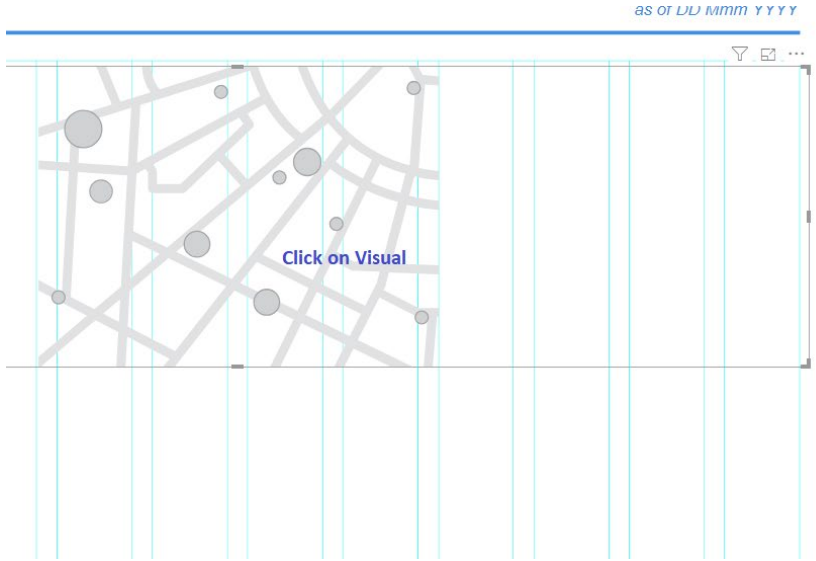


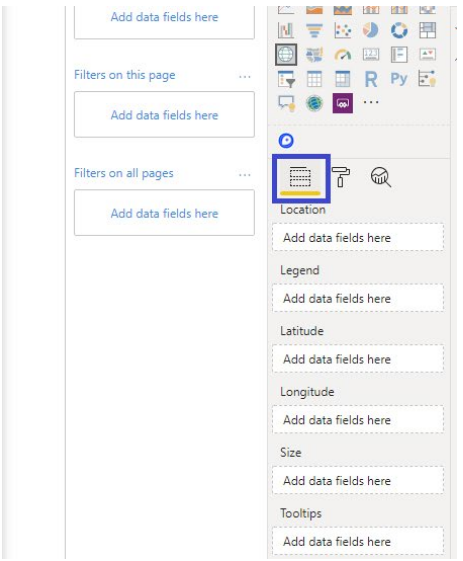
Hold Point 5

Now we have our data prepared and related we are going to do an example visualization to demonstrate what we could now do with our data.

This training is not intended to be about the final visualization process but we will do an example to demonstrate the benefit of our preparation:

- Click on the report tab on the left pane
- The template file you were provided has one map visualisation already in place but with no data fields selected
- Click on the provided visualization and you will see the field input pane on the right hand side change to receive inputs for this visual.
- Click on the “Fields” button as shown to the right to confirm that you want to input fields
- You can now drag the relevant fields from the fields pane into the input pane.



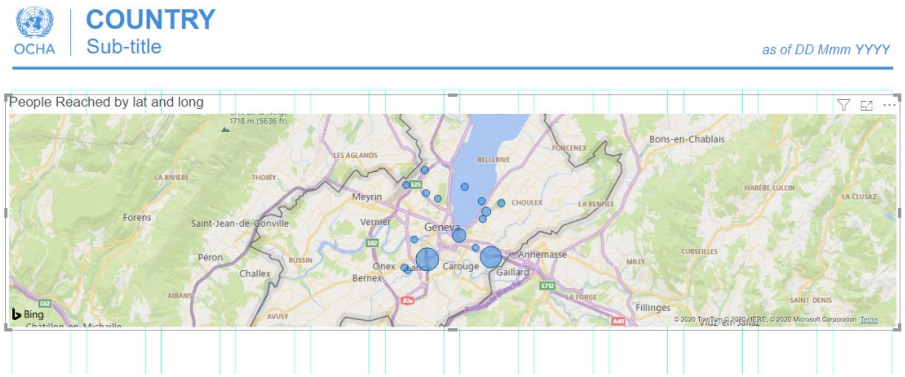


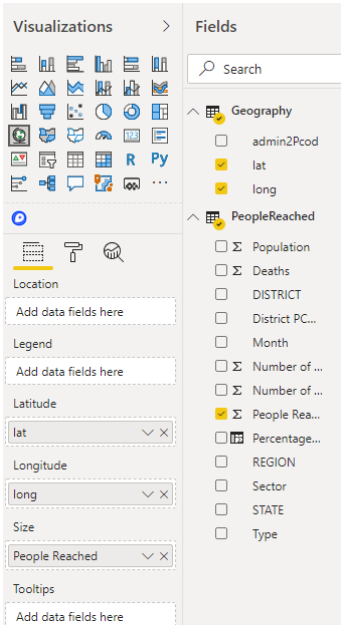
- The lat and long fields can be dragged from the geography table. After dragging the lat and long in, click on the small down arrow next to the title “lat”. Check that it is not summarizing the data such as showing average, min or max. We just want to map based on the value, not some statistic of it.

- Now you will see dots appearing on the map throughout Geneva.

- But we also want the size of the dots to be based on the number of people reached field. So we drag it into the size box

- If we had no relationship between the tables we would get an error here because PowerBI would not know how to relate the location name with the lat and long

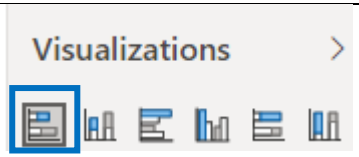


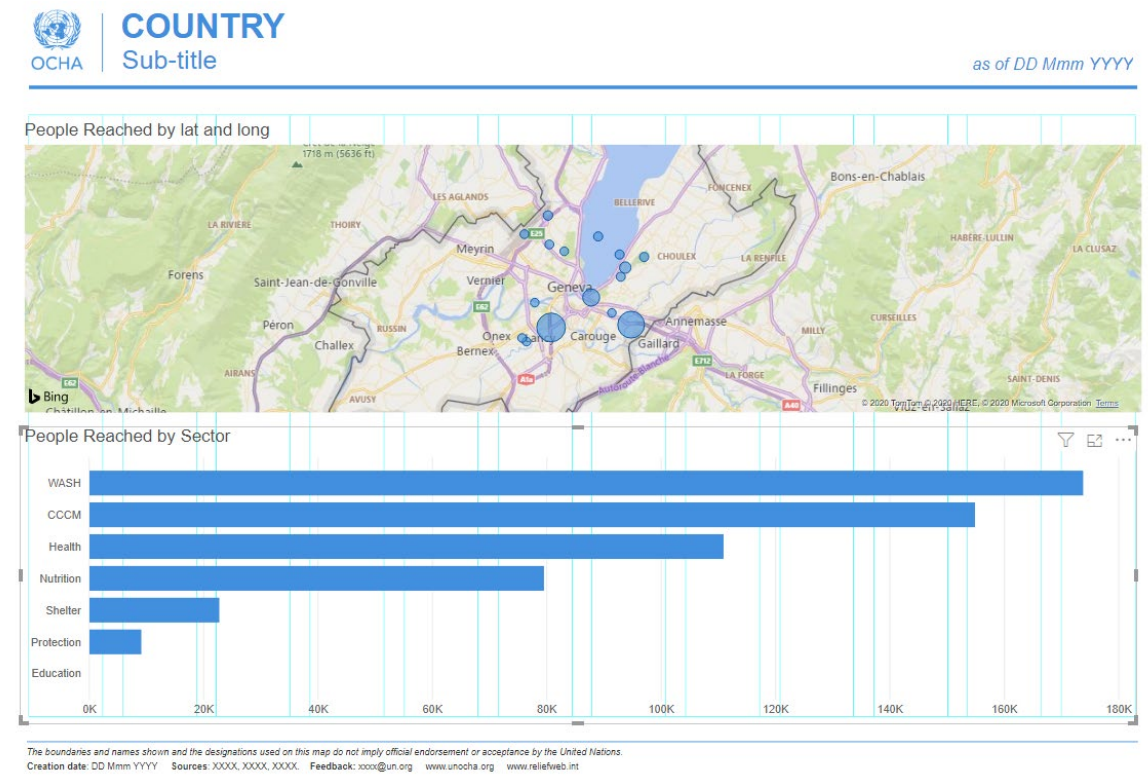


Bonus Step if You have time

Now we are going to make a bar graph showing the people reached by sector

- Deselect the map visualization by clicking elsewhere on the dashboard background
- Click on the stacked bar chart visualization
- Drag the Sector field into the axis box
- Drag the people reached to the values box





Exercise End