2011 Census DataPacks

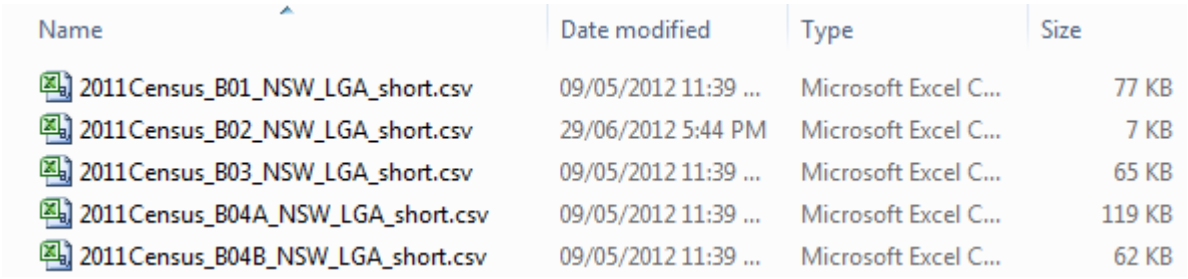
Adding geographic labels to DataPack files

The following guide will walk you through the process of attaching geographic labels to DataPack .csv files using Microsoft Excel.

Step 1: Obtain relevant files

At this stage, we're going to assume that you've obtained the various DataPack files that you're after by way of downloading them from our website or from purchasing one of the DataPack DVD's.

As an example, Basic Community Profile DataPacks, separated into Local Government Areas (LGA's) for NSW with short descriptors should look something like this in your directory:



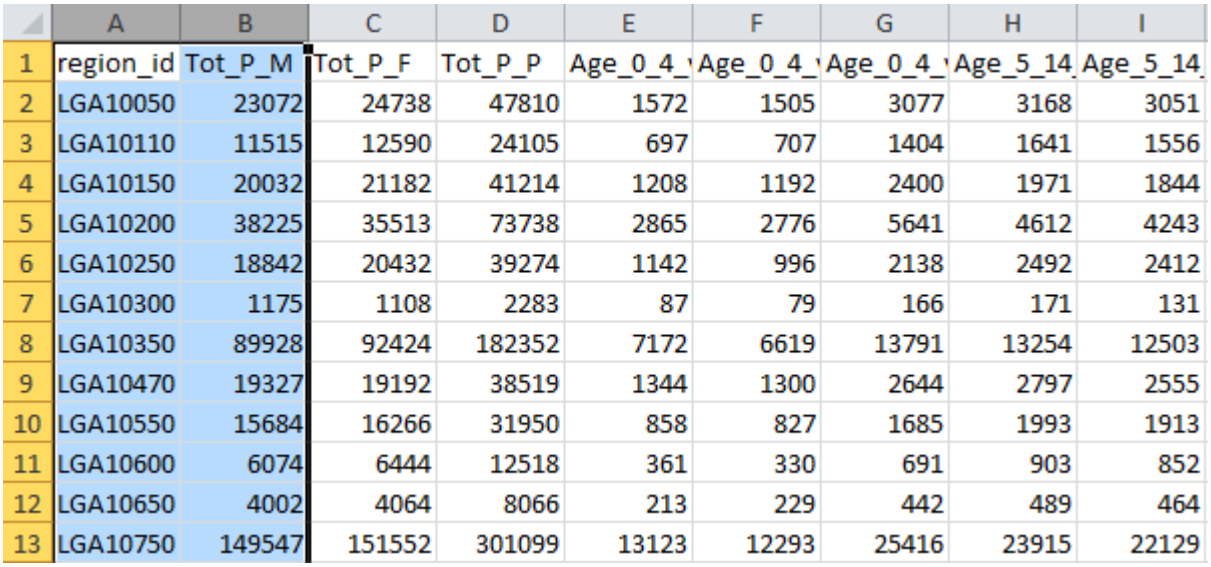
Name	Date modified	Type	Size
2011Census_B01_NSW_LGA_short.csv	09/05/2012 11:39 ...	Microsoft Excel C...	77 KB
2011Census_B02_NSW_LGA_short.csv	29/06/2012 5:44 PM	Microsoft Excel C...	7 KB
2011Census_B03_NSW_LGA_short.csv	09/05/2012 11:39 ...	Microsoft Excel C...	65 KB
2011Census_B04A_NSW_LGA_short.csv	09/05/2012 11:39 ...	Microsoft Excel C...	119 KB
2011Census_B04B_NSW_LGA_short.csv	09/05/2012 11:39 ...	Microsoft Excel C...	62 KB

You're also going to need the geographical description file (2011Census_geog_desc.xlsx) in order to extract the relevant labels/names and attach them to the DataPack files.

Step 2: Making room for geographic labels in the DataPack file

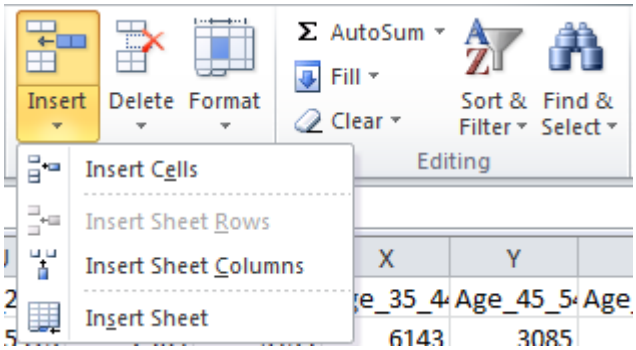
In this example, we'll be opening 2011Census_B01_NSW_LGA_short.csv in Microsoft Excel and inserting two blank columns.

Simply highlight the entire column A and column B as so:



	A	B	C	D	E	F	G	H	I
1	region_id	Tot_P_M	Tot_P_F	Tot_P_P	Age_0_4_	Age_0_4_	Age_0_4_	Age_5_14	Age_5_14
2	LGA10050	23072	24738	47810	1572	1505	3077	3168	3051
3	LGA10110	11515	12590	24105	697	707	1404	1641	1556
4	LGA10150	20032	21182	41214	1208	1192	2400	1971	1844
5	LGA10200	38225	35513	73738	2865	2776	5641	4612	4243
6	LGA10250	18842	20432	39274	1142	996	2138	2492	2412
7	LGA10300	1175	1108	2283	87	79	166	171	131
8	LGA10350	89928	92424	182352	7172	6619	13791	13254	12503
9	LGA10470	19327	19192	38519	1344	1300	2644	2797	2555
10	LGA10550	15684	16266	31950	858	827	1685	1993	1913
11	LGA10600	6074	6444	12518	361	330	691	903	852
12	LGA10650	4002	4064	8066	213	229	442	489	464
13	LGA10750	149547	151552	301099	13123	12293	25416	23915	22129

And then select “Insert Cells” from the MS Excel Ribbon Menu:



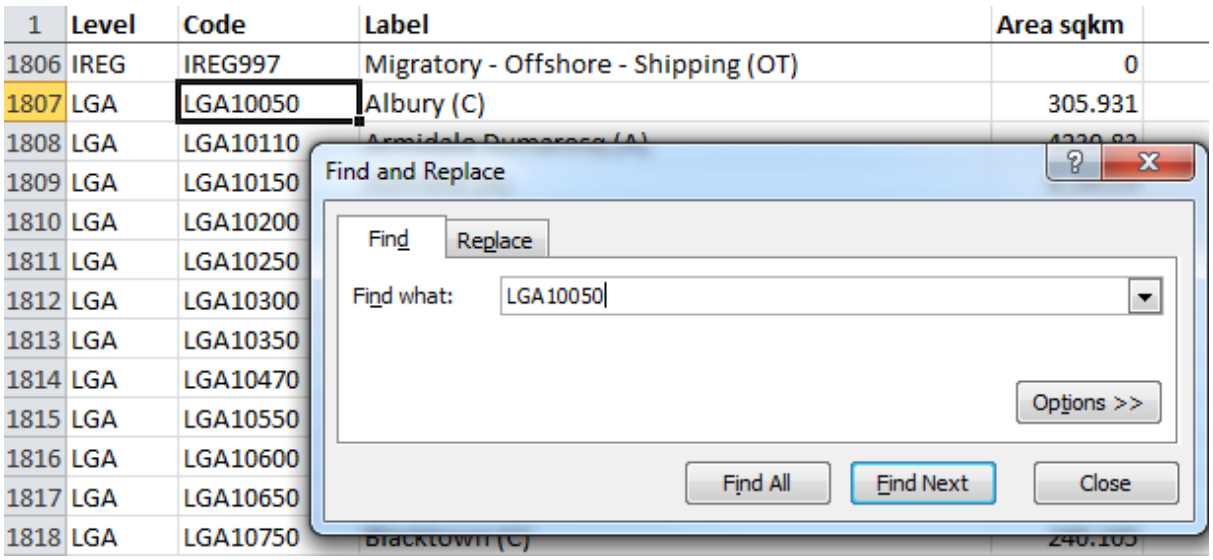
Alternatively, you can use the keyboard shortcut <ctrl>+<+>.

Your DataPack file should now appear as follows:

	A	B	C	D	E	F	G	H	I
1			region_id	Tot_P_M	Tot_P_F	Tot_P_P	Age_0_4_	Age_0_4_	Age_0_4_
2			LGA10050	23072	24738	47810	1572	1505	3077
3			LGA10110	11515	12590	24105	697	707	1404
4			LGA10150	20032	21182	41214	1208	1192	2400
5			LGA10200	38225	35513	73738	2865	2776	5641
6			LGA10250	18842	20432	39274	1142	996	2138
7			LGA10300	1175	1108	2283	87	79	166
8			LGA10350	89928	92424	182352	7172	6619	13791
9			LGA10470	19327	19192	38519	1344	1300	2644
10			LGA10550	15684	16266	31950	858	827	1685

Step 3: Extracting geographic labels

Open the file named 2011Census_geog_desc.xlsx and search for the relevant region ID’s relevant to the DataPack file that you’re working with. In this case, we want to locate the NSW LGA codes (utilise the Excel Find function by pressing <ctrl>+<f>):



In this example, highlight the Code and Label columns for LGA codes LGA10050 through to LGA19499 and press <ctrl>+<c> to copy the data to Microsoft's clipboard:

	A	B	C	D
1	**Level**	**Code**	**Label**	**Area sqkm**
1806	IREG	IREG997	Migratory - Offshore - Shipping (OT)	0
1807	LGA	LGA10050	Albury (C)	305.931
1808	LGA	LGA10110	Armidale Dumaresq (A)	4230.83
1809	LGA	LGA10150	Ashfield (A)	8.28119
1810	LGA	LGA10200	Auburn (C)	32.4769
1811	LGA	LGA10250	Ballina (A)	484.716
1812	LGA	LGA10300	Balranald (A)	21693.12
1813	LGA	LGA10350	Bankstown (C)	76.8001
1814	LGA	LGA10470	Bathurst Regional (A)	3816.43
1815	LGA	LGA10550	Bega Valley (A)	6278.96
1816	LGA	LGA10600	Bellingen (A)	1600.38
1817	LGA	LGA10650	Berrigan (A)	2065.93
1818	LGA	LGA10750	Blacktown (C)	240.105
1819	LGA	LGA10800	Bland (A)	8557.71
1820	LGA	LGA10850	Blayney (A)	1524.88
1821	LGA	LGA10900	Blue Mountains (C)	1431.15

Step 4: Pasting the geographic labels into the DataPack files

Now simply return to the DataPack file that we manipulated in Step 2, click on cell A2 and press <ctrl>+<v> to paste the labels that we extracted/copied in Step 3.

For headings, you can type "Code" and "Label" into the top of Column A and Column B:

	A	B	C	D	E	F	G
1	Code	Label	region_id	Tot_P_M	Tot_P_F	Tot_P_P	Age_0_4_
2	LGA10050	Albury (C)	LGA10050	23072	24738	47810	1572
3	LGA10110	Armidale Dumaresq (A)	LGA10110	11515	12590	24105	697
4	LGA10150	Ashfield (A)	LGA10150	20032	21182	41214	1208
5	LGA10200	Auburn (C)	LGA10200	38225	35513	73738	2865
6	LGA10250	Ballina (A)	LGA10250	18842	20432	39274	1142
7	LGA10300	Balranald (A)	LGA10300	1175	1108	2283	87
8	LGA10350	Bankstown (C)	LGA10350	89928	92424	182352	7172
9	LGA10470	Bathurst Regional (A)	LGA10470	19327	19192	38519	1344

Step 5: Saving

Now that you've finished altering the DataPack file to include geographic labels, simply save the .csv file in a location of your choice and repeat the process for as many DataPack files that you'd like to alter.

Appendix 1: Checking that the Code and region id columns match

The “Code” and “region_id” columns should match exactly, though to be sure that they do you can utilise the following MS Excel check:

- Insert a blank Column A the same way we did in Step 2 and include the heading “Check”:

	A	B	C	D	E	F	G
1	Check	Code	Label	region_id	Tot_P_M	Tot_P_F	Tot_P_P
2		LGA10050	Albury (C)	LGA10050	23072	24738	47810
3		LGA10110	Armidale Dumaresq (A)	LGA10110	11515	12590	24105
4		LGA10150	Ashfield (A)	LGA10150	20032	21182	41214
5		LGA10200	Auburn (C)	LGA10200	38225	35513	73738
6		LGA10250	Ballina (A)	LGA10250	18842	20432	39274
7		LGA10300	Balranald (A)	LGA10300	1175	1108	2283

- Now highlight the column (A2:A155 in this example), type the formula =EXACT(B2,D2) and press <ctrl>+<Enter> to copy the formula down for all of the cells.

	A	B	C	D	E	F
1	Check	Code	Label	region_id	Tot_P_M	Tot_P_F
2	=EXACT(B2,D2)	LGA10050	Albury (C)	LGA10050	23072	24738
3		LGA10110	Armidale Dumaresq (A)	LGA10110	11515	12590
4		LGA10150	Ashfield (A)	LGA10150	20032	21182
5		LGA10200	Auburn (C)	LGA10200	38225	35513
6		LGA10250	Ballina (A)	LGA10250	18842	20432
7		LGA10300	Balranald (A)	LGA10300	1175	1108

- If the cells match then the entire column should be populated by “TRUE”.
- Please note that if you do perform this check the formula functionality will not be retained when saved as a .csv file.