

## Understanding Society User Support - Support #715

### Wave 1 to 6 Data Release - Differences compared to Wave 1 to 5 data release

02/06/2017 02:34 PM - Ben Clark

<b>Status:</b>	Closed	<b>Start date:</b>	02/06/2017
<b>Priority:</b>	Normal	<b>% Done:</b>	60%
<b>Assignee:</b>	Ben Clark		
<b>Category:</b>			
<b>Description</b>			
Good afternoon,			
I have recently downloaded the Wave 1 to 6 data release having previously been working with the wave 1 to 5 release.			
There appear to be a small number of data differences in the wave 1 to 6 data files compared against the wave 1 to 5 release. In particular, I am using the special licence LSOA codes to link to other spatial data sets. In doing so I have noticed that there are different numbers of missing values in the 'country' variable (in most waves) comparing the wave 1 to 5 release to the wave 1 to 6 release.			
What leads to these differences and should I be concerned about this? It looks like my results are likely to be very slightly different if I run my analysis on the wave 1 to 6 release. This means I will need to maintain two sets of data files in order to reproduce my earlier results consistently.			
Many thanks, Ben			

#### History

##### #1 - 02/07/2017 10:29 AM - Ben Clark

Good afternoon,

I have recently downloaded the Wave 1 to 6 data release having previously been working with the wave 1 to 5 release.

There appear to be a small number of data differences in the wave 1 to 6 data files compared against the wave 1 to 5 release. In particular, I am using the special licence LSOA codes to link to other spatial data sets. In doing so I have noticed that there are different numbers of missing values in the 'country' variable (in most waves) comparing the wave 1 to 5 release to the wave 1 to 6 release.

What leads to these differences and should I be concerned about this? It looks like my results are likely to be very slightly different if I run my analysis on the wave 1 to 6 release. This means I will need to maintain two sets of data files in order to reproduce my earlier results consistently.

There also seem to be 799 records in the wave 6 indresp file that do not have a matched record in the hhresp file (based on a merge on hidp). Is this correct of have I made a mistake in my merge command?

Many thanks in anticipation,  
Ben

##### #2 - 02/08/2017 01:12 PM - Victoria Nolan

- Status changed from New to In Progress

- Assignee set to Victoria Nolan

- % Done changed from 0 to 10

- Private changed from Yes to No

Dear Ben,

Many thanks for your enquiry. The team is looking into it and we will get back to you shortly.

Best wishes, Victoria.

On behalf of the Understanding Society Data User Support Team

##### #3 - 02/22/2017 10:59 AM - Victoria Nolan

- Status changed from In Progress to Feedback

- Assignee changed from Victoria Nolan to Ben Clark

- % Done changed from 10 to 60

Dear Ben,

I'm sorry for the delay in getting back to you, while we have been looking into this.

Re: Missing values in the country variable being slightly different for the various waves in the wave 5 release and wave 6 release. This is possible as we take the opportunity to make corrections in previous waves when we release a new wave of data. In addition, it is not clear from your comments at what point the numbers you are using were generated – this is significant as you mention LSOA. The mechanism used for generating LSOA01 and LSOA11 (and all of the special licence type geography fields) has changed from wave 5 deposit to wave 6 deposit and this may have had an impact. Could you please provide a bit more information about the LSOA numbers?

Re: 799 records in the wave 6 indresp file that do not have a matched record in the hhresp file:  
The figure you quote is correct. This appears to be an issue with the fieldwork protocol, which we are discussing with the fieldwork agency to prevent in future. These are individuals that completed a questionnaire for the survey but for which no household level questionnaire was completed. It was decided, given the numbers involved, to include these individual's responses in INDRESP. Depending on your specific analysis requirements these individual's records can be kept or dropped as appropriate. Please note that if you are interested in household level variables derived from the household grid, for example household size, number of children in the household etc, that you can derive them yourself by using INDALL

We hope this helps, best wishes, Victoria

On behalf of the Understanding Society Data User Support Team

**#4 - 03/16/2017 11:07 AM - Victoria Nolan**

Dear Ben,

Could I please just check whether you still need our help with this issue? If you could provide some more details about the LSOA, we can look into it.

Many thanks, Victoria.

**#5 - 03/28/2017 01:44 PM - Victoria Nolan**

- Status changed from Feedback to Closed