



***Understanding Society* Longitudinal Teaching Dataset  
USER GUIDE**

**Version 2, October 2021**



## Table of Contents

1. Introduction .....	1
2. Overview of understanding society .....	1
3. What happens during the interviews?.....	2
3.1 Enumeration and household grid .....	2
3.2 Household interview .....	2
3.3 Young children (10-15 years) interviews .....	3
3.4 Adult (16+ years) interviews .....	3
3.5 Interview mode .....	3
3.6 Incentives and consent .....	3
3.7 Fieldwork period .....	3
4. What is in this datafile?.....	4
4.1 What information is included in this datafile? .....	4
4.2 Who is included in this datafile? .....	4
4.3 What does the datafile look like? .....	4
4.4 What variables does this datafile include? .....	5
4.5 Missing values.....	5
5. To know more about the survey .....	6
6. File formats .....	6
7. Citation.....	7
Appendix A: Variable list and description.....	8
Appendix B: Ethnic group question showcard.....	13
Endnotes .....	14

## 1. Introduction

This data file combines the responses from interviews of adult (16+ year old) respondents of the General Population and Ethnic Minority Boost samples from Wave 1 to Wave 9 of Understanding Society into one long format file. Only those who responded to all 9 annual interviews are included. This is a person-wave level file, with nine rows of observations for each respondent. The dataset also includes some basic household level information collected in the household questionnaire such as household income. For the list of variables included in this data file and further details about the data please refer to this User Guide included in the documentation pack. Also included in the pack is a short guide for inspecting, managing and analysing longitudinal data using Stata together with the corresponding Stata syntax and output log files.

## 2. Overview of understanding society

*Understanding Society*: the UK Household Longitudinal Study (UKHLS), is a survey that started in 2009 with a nationally representative sample of individuals living in approx. 26,000 private households in the UK (referred to as the General Population Sample, GPS), and a sample of individuals living in approx. 4,000 private households where at least one household member had an ethnic minority background (referred to as the Ethnic Minority Boost Sample, EMBS).<sup>1</sup>

*Understanding Society* is a longitudinal survey, which means information is collected from the same people every year to measure changes over time and across the lifecourse. The first time the sample was interviewed is referred to as Wave 1, and the next interview, approximately a year later, as Wave 2 and so on. But after a year some individuals may have moved within or outside the UK, household members may have moved out and new members moved in (partnerships form and dissolve, new relatives move in and so on). Then the question is who is eligible to be interviewed each year? Do these new individuals qualify to be interviewed? If someone moves from England to Scotland, are they still eligible to be interviewed? To answer these questions we need to keep in mind that the population of interest is the UK residential population in 2009-10 and their future generations. So, all individuals in the responding households in Wave 1 form the core sample and are followed wherever they move as long as it is within the UK.<sup>2</sup>

However, not everyone who is eligible for an interview participates in the interview. This could be for many different reasons. Sometimes interviewers are not able to contact eligible sample members as they were away each time the interviewer visited. In some cases it is not possible to locate sample members who have moved and did not inform the survey of their new address. Even when sample members are contacted they may not participate because they are too busy, or ill. This non-response and attrition over time is generally not random, that is, certain types of individuals are more likely to not participate in interviews. As a result estimates based using this sample may be biased. The weights provided with the data are designed to adjust for this (see discussion later).

### 3. What happens during the interviews?

Data is collected by interviewing adults (16+ year olds) and young persons or children (10-15 year olds) where they are asked questions about their lives and by interviewing a responsible adult to ask questions about the household. Some information is also collected by interviews, and some by the fieldwork agency. This section gives a quick overview of these different sources of data.

#### 3.1 Enumeration and household grid

When an interviewer arrives at the sampled address, they are instructed to select a maximum of 3 dwelling units<sup>3</sup> and within each dwelling unit a maximum of 3 households<sup>4</sup> to be part of the sample. In most cases, there is just one dwelling unit at an address and within it one household. After this the interviewer asks an adult in the household to tell them who else is in the household and what their name, dates of birth or estimated age, sex, employment status and marital status are. This is recorded in the enumeration and household grids.

#### 3.2 Household interview

After this interviewer requests an adult to complete the household questionnaire which includes questions about tenancy, mortgage, household expenditures etc. Interviewers are instructed to try and find the person who is most likely to know this information. This is generally considered to be the person who owns or rents the accommodation. So, interviewers are instructed to identify such a person and to interview them. If there is more than one such person then the eldest is chosen.

### **3.3 Young children (10-15 years) interviews**

The interviewer requests the parents or guardians of 10-15 year old children in these households if they would consent to their children completing a short self-completion questionnaire, which as the name suggests is completed by these children on their own.

### **3.4 Adult (16+ years) interviews**

The interviewer also asks everyone who is 16 years of age or older if they would like to participate in an individual interview and if they agree then the interviewer asks them questions in the adult individual questionnaire. Some questions that are deemed to be too sensitive for an individual to be able to comfortably answer in front of the interviewer, are included in a separate **self-completion questionnaire** which the survey participant answers by themselves.

### **3.5 Interview mode**

Until Wave 7, most interviews were conducted as described above, that is with interviewers visiting the homes of sample members (except for around 500 households who were interviewed by telephone). After that an increasing proportion were invited to complete the interviews by themselves online.<sup>5</sup>

### **3.6 Incentives and consent**

Informed consent to participate in the survey is asked of sample members and the interviews take place only with sample members who give consent and they are made aware that they can stop the interview at any point. High Street vouchers of small amounts (generally £5 for 10-15 year olds, £10 for 16+ and higher for some non-responding household members) are given to sample members who participate in an interview as a token of appreciation for participating in the survey.

### **3.7 Fieldwork period**

The fieldwork or interview period stretches for approximately 24 months. Sometimes when interviewers visit the homes of sample members they may not be at home as they are away on holiday, or at work, or on a business trip. In such situations the interviewer attempts to visit again later to interview them. So, a few interviews are conducted beyond the 24 month fieldwork period. For example, in Wave 1, the interviews were conducted between January 2009 and March 2011.

## 4. What is in this datafile?

### 4.1 What information is included in this datafile?

This datafile includes the responses collected during the adult interviews for the first nine waves spanning 2009-2018. The datafile also includes some information collected about the households (via the household questionnaire and household grid) of these respondents. See tables in Appendix A for a complete list of variables included.

### 4.2 Who is included in this datafile?

The datafile includes 16+ year old core sample members who responded continuously to the adult interviews from the Wave 1 until they dropped out of the survey<sup>6</sup>. Figure 4.1 is a diagrammatic representation of the different response patterns observed in this datafile.

Figure 4.1: Response patterns

Waves →	1	2	3	4	5	6	7	8	9
	R								
	R	R							
	R	R	R						
	R	R	R	R					
	R	R	R	R	R				
	R	R	R	R	R	R			
	R	R	R	R	R	R	R		
	R	R	R	R	R	R	R	R	
	R	R	R	R	R	R	R	R	R

### 4.3 What does the datafile look like?

There are two ways to organise a longitudinal dataset: a long format and a wide format. See Tables 4.1 and 4.2

Table 4.1: Example of a longitudinal data in wide format

Person ID	Age at Wave 1	Age at Wave 2	Age at Wave 3	Country of birth	Sex
1	20	21	22	Bangladesh	Woman
2	53	54		Jamaica	Man
3	59	.		England	Woman

Table 4.2: Example of a longitudinal data in long format

Person ID	Wave	Age	Country of birth	Sex
1	1	20	Bangladesh	Woman
1	2	21	Bangladesh	Woman
1	3	22	Bangladesh	Woman
2	1	53	Jamaica	Man
2	2	54	Jamaica	Man
3	1	59	England	Woman

A wide format file looks like a cross-sectional dataset, where one row is for one unit of observation and the variables represent measures from each wave. In Table 4.1 we see that the data consists of responses from three persons across three waves. While person 1 responds in all three waves, person 2 drops out in the 3<sup>rd</sup> wave and person 3 after the 2<sup>nd</sup> wave. There are four variables in this file: unique person identifier, variables that measure age each wave, country of birth and sex<sup>7</sup>. As country of birth and sex do not change there are only two variables and not six representing these at each wave.

The same data is represented in long format in Table 4.2. Here each row represents a person and the wave in which the person was interviewed. So, Person with ID=1 appears three times as this person was interviewed in all three waves while the other two persons appear once or twice. Here there is only one variable that measures age of each person at each wave. This means that now we need another variable to identify the wave from which that data was collected. In the long format you will not be able to identify missed interviews by looking at missed variable values as in a wide format file, but by checking if a person has as many rows of observations as interview waves.

When there are holes or gaps in the data, i.e., the same set of observations do not appear in every wave, then that is an unbalanced panel, otherwise it is a balanced panel. So, this datafile is an unbalanced panel.

### 4.4 What variables does this datafile include?

This data file includes the following sets of variables: Identifiers and interview information, socio-demographic characteristics, education and labour market information, residential information, income, health and wellbeing, political behaviour and opinions, environmental attitudes and behaviours. See Tables A1-A9 for a detailed list of variables and their descriptions.

### 4.5 Missing values

There are some variables for which the values are missing even though the person was interviewed. Why? There are different reasons. It could be because:

- The question was not asked in every wave. For example, the environmental module was asked in Waves 1 and 4.

- The question was not asked of some people as it was not relevant for them. For example, earnings and pay questions were not asked of those who said they were not in paid employment last week.
- The question was asked of the sample member, but they either didn't know or refused to answer the question.
- In a few cases, it could be missing due to some type of technical error.

## 5. To know more about the survey

To know more about the survey, you can visit the [Study website](#). Here is a list of useful links to this online documentation.

[About the study](#)

[User Guide](#)

[Search for variables](#)

[Questionnaires](#)

[Training videos, webinar recordings](#)

[Help and support](#)

## 6. File formats

The data is provided in Stata, SPSS and tab delimited formats. If you want to read the data in R, first download the dataset in Stata format. Then you need to install and load haven package (you need to install it only once but load it every time you start a new session). You can do this running the following commands:

```
install.packages("haven")  
library(haven)
```

Then use the following command to read the Stata data file into R. Replace `mydata` in the command below with the filepath to the folder where you have stored `longitudinal_td.dta` file.

```
longitudinal_td <- read_dta("mydata/longitudinal_td.dta")
```



## 7. Citation

The bibliographic citation for this user guide is:

Institute for Social and Economic Research (2020) *Understanding Society Longitudinal Teaching Dataset User Guide*. Version 2, October 2021. Colchester: University of Essex.

The bibliographic citation for the *Understanding Society* Longitudinal Teaching Dataset data is:

University of Essex. Institute for Social and Economic Research. (2020). *Understanding Society: Longitudinal Teaching Dataset, Waves 1-9, 2009-2018*. [data collection]. 1<sup>st</sup> Edition. UK Data Service. SN: 8715, <http://doi.org/10.5255/UKDA-SN-8715-1>

## Appendix A: Variable list and description

<b>Table A1: Identifiers and interview information</b>	
<b>Variable name</b>	<b>Description</b>
pidp	Individual identifier (unique within and across waves)
wave	Interview wave
hidp	Household identifier (unique within a wave, not across waves)
buno_dv	Benefit unit within a household
intdaty_dv	Interview year
intdatm_dv	Interview month
intdatd_dv	Interview day
indmode	Interview mode

<b>Table A2: Sampling and weight variables</b>	
<b>Variable name</b>	<b>Description</b>
hhorig	Sample origin
psu	Primary sampling unit
strata	Sampling strata
sampst	Sample status
scflag_dv	Whether completed self-completion part of the adult questionnaire
indinus_lw_n	Longitudinal weight to be used when analysing responses to individual adult interview excluding self-completion questions from Waves 1 to n
indscus_lw_n	Longitudinal weight to be used when analysing responses to individual adult interview including self-completion questions (or only responses to self-completion questions) from Waves 1 to n

<b>Table A3: Residential information</b>	
<b>Variable name</b>	<b>Description</b>
mvever <sup>a</sup>	Lived at this address whole life ( <i>Waves: 1</i> )
mvmnth <sup>a</sup>	Month moved to current address, if not living at current address ( <i>Waves: 1</i> )
mveyr <sup>a</sup>	Year moved to current address, if not living at current address ( <i>Waves: 1</i> )
distmov_dv	Distance participant moved since last wave in Km ( <i>Waves: 3-9</i> )
addrmov_dv	Participant changes address postcode since last wave ( <i>Waves: 3-9</i> )
lkmove	Prefers to move from current address
xpmove	Expects to move from current address
gor_dv	Government office region
urban_dv	Whether participant lives in an urban or a rural area

<b>Table A4: Socio-demographic &amp; household characteristics</b>	
<b>Variable name</b>	<b>Description</b>
age_dv	Age in years at time of interview
doby_dv	Year of birth
sex_dv	Sex
ethn_dv	Ethnic group (complete labels are shown in Appendix B)
cob_dv	Country of birth
bornuk_dv	Whether born in UK
yr2uk4	Year arrived into UK
hysize_dv	Household size
hhtype_dv	Type of household based on household composition
tenure_dv	Whether the accommodation is owned or rented (private, social housing)
hhtype_dv	Composition of household, LFS-version
mstat_dv	De facto marital status
livesp_dv	Lives with spouse in household
cohab_dv	Lives with cohabitee in household
nchild_dv	Number of own children in household
depchl_dv	Whether dependent child - official definition
ndepchl_dv	Number of own dependent children in household
hiqual_dv	Highest educational qualification

<b>Table A5: Health and wellbeing</b>	
<b>Variable name</b>	<b>Description</b>
sf1_dv <sup>a</sup>	General health
bmi_dv	Body Mass Index ( <i>Waves: 1</i> )
sf12pcs_dv <sup>a</sup>	SF-12 Physical Component Summary
sf12mcs_dv <sup>a</sup>	SF-12 Mental Component Summary
scghq1_dv <sup>a</sup>	Subjective wellbeing (GHQ): Likert
scghq2_dv <sup>a</sup>	Subjective wellbeing (GHQ): Caseness
swemwbs_dv <sup>a</sup>	Short Warwick-Edinburgh Mental Well-being Scale ( <i>Waves: 1 4 7</i> )
sclfsato <sup>a</sup>	Satisfaction with life overall

<sup>a</sup>Asked in self-completion questionnaire. But General Health question was asked by interviewers and not included in the self-completion questionnaire in Wave 1, in all other waves it was included in the self-completion questionnaire.

<b>Table A6: Labour market and job related information</b>	
<b>Variable name</b>	<b>Description</b>
jbstat	Current labour force status
jbhas_dv	Whether did any paid work last week, and was this paid or self-employment
jbsoc00_cc	Current job: SOC 2000, condensed
jbsic07_cc	Current job: SIC 2007, condensed
jbnsec8_dv	Current job: Eight Class NS-SEC
jbmngr	Current job: Has managerial duties
jbsize	Current job: No. employed at workplace
jbterm_dv	Current job: Type of job contract
jbsect_dv	Current job: Type of organisation working for
jbhrs	Current job: no. of hours normally worked per week
jbot	Current job: no. of overtime hours in normal week
jbft_dv	Current job: Full or part-time employee
jbotpd	Current job: No. of hours worked as paid overtime
jbpl	Current job: Work location
jbttwt	Current job: Minutes spent travelling to work
workdis	Current job: Distance from work ( <i>Waves: 1 2 4 6 8</i> )
worktrav	Current job: Mode of transport for journey to work
jbsat	Job satisfaction (1-7)
j2has	Has a second job?
j2semp	Second job: employee or self-employed
j2soc00_cc	Second job: SOC 2010, condensed
j2nsec8_dv	Second job: NSSEC 8 classes
j2hrs	Second job: no. of hours worked per month
jsboss	Self-employed: hires employees
jssize	Self-employed: number of employees
jshrs	Self-employed: hours normally worked per week
jstypeb	Self-employed: nature of employment
jsaccs	Self-employed: draws up profit/loss accounts
jspart	Self-employed: own account or partnership
jspl	Self-employed: work location
jsttwt	Self-employed: commuting time provided
jsttwtb	Self-employed: commuting time
jsworkdis	Self-employed: commuting distance ( <i>Waves: 1</i> )
jsworktrav	Self-employed: mode of transport to work

<b>Table A6: Labour market and job related information (continued)</b>	
<b>Variable name</b>	<b>Description</b>
jbhad	Ever had paid employment ( <i>Waves: 1</i> )
jlsemp	Last job: Employee or self-employed? ( <i>Waves: 1</i> )
jlendy	Last job: year left job ( <i>Waves: 1</i> )
jlendm	Last job: month left job ( <i>Waves: 1</i> )
jlsoc00_cc	Last job: SOC 2010, condensed ( <i>Waves: 1</i> )
jlsic07_cc	Last job: SIC 2007, condensed ( <i>Waves: 1</i> )
jlnssec8_dv	Last job: Eight Class NS-SEC ( <i>Waves: 1</i> )
jlmngr	Last job: Has managerial duties ( <i>Waves: 1</i> )
jlboss	Last job: hired employees ( <i>Waves: 1</i> )
jlsize	Last job: number of people employed at workplace ( <i>Waves: 1</i> )
paygu_dv	Usual gross monthly pay of current job
paygu_if	Whether paygu_dv that was imputed
paynu_dv	Usual net monthly pay of current job
paynu_if	Whether paynu_dv that was imputed
j2pay_dv	Gross monthly pay of second job
j2paynet_dv	Net monthly pay of second job
j2pay_if	Whether j2pay_dv that was imputed
seearngrs_dv	Gross monthly self-employment earnings
seearnnet_dv	Net monthly self-employment earnings
seearngrs_if	Whether seearngrs_dv that was imputed

<b>Table A7: Income</b>	
<b>Variable name</b>	<b>Description</b>
fimngrs_dv	Gross monthly personal income
fimnnet_dv	Net monthly personal income, no deductions
fimngrs_if	Share if fimngrs_dv that was imputed
fihhmngrs_dv	Gross monthly household income (before housing costs)
fihhmnet1_dv	Net monthly household income (before housing costs), no deductions
fihhmngrs_if	Share of fihhmngrs_dv that was imputed
ieqmoecd_dv	Modified OECD equivalence scale

<b>Table A8: Political behaviour and opinions</b>	
<b>Variable name</b>	<b>Description</b>
vote1	supports a particular political party ( <i>Waves: All except 8</i> )
vote2	closer to one political party than others ( <i>Waves: All except 8</i> )
vote3	Party would vote for tomorrow ( <i>Waves: All except 8</i> )
vote4	Which political party closest to ( <i>Waves: All except 8</i> )
vote5	strength of support for stated party ( <i>Waves: All except 8</i> )
vote6	level of interest in politics ( <i>Waves: All except 8</i> )
vote7	voted in last general election ( <i>Waves: 2 7 8 9</i> )
vote8	Party voted for in last general election ( <i>Waves: 2 7 8 9</i> )
votenorm	Voting as a social norm ( <i>Waves: 2 3 6 9</i> )
voteintent	voting intention ( <i>Waves: 2 3 6 9</i> )
grpbfts	Group benefit from voting ( <i>Waves: 2 3 6 9</i> )
perbfts	Personal benefit in <i>voting</i> ( <i>Waves: 2 3 6 9</i> )

<b>Table A9: Environmental attitudes and behaviours (<i>Waves 1 &amp; 4</i>)</b>	
<b>Variable name</b>	<b>Description</b>
envhabit1	How often leave your TV on standby for the night
envhabit2	How often switch off lights in rooms that aren't being used
envhabit3	How often keep the tap running while you brush your teeth
envhabit4	How often put more clothes on when you feel cold rather than putting the heating on or turning it up
envhabit5	How often decide not to buy something because you feel it has too much packaging
envhabit6	How often buy recycled paper products such as toilet paper or tissue
envhabit7	How often take your own shopping bag when shopping
envhabit8	How often use public transport (e.g. bus, train) rather than travel by car
envhabit9	How often walk or cycle for short journeys less than 2 or 3 miles
envhabit10	How often car share with others who need to make a similar journey
envhabit11	How often take fewer flights when possible

## Appendix B: Ethnic group question showcard

### White

1. British/English/Scottish/Welsh/Northern Irish
2. Irish
3. Gypsy or Irish Traveller
4. Any other white background

### Mixed

5. White and Black Caribbean
6. White and Black African
7. White and Asian
8. Any other mixed background

### Asian or Asian British

9. Indian
10. Pakistani
11. Bangladeshi
12. Chinese
13. Any other Asian background

### Black / African / Caribbean / Black British

14. Caribbean
15. African
16. Any other Black background

### Other ethnic group

17. Arab
97. Any other ethnic group

## Endnotes

---

<sup>1</sup> Subsequently, two additional samples were added. In 2010, the long running British Household Panel Survey (BHPS) sample was added. To know more about the study design and the different samples see the section on [Study Design](#) in the User Guide.

<sup>2</sup> Children born to women core sample members also become part of the core sample. But partners, spouses, and other household members who join these households after the first wave do not become part of the core sample and are interviewed only as long as they are living with the core sample member as they provide information about the household context (e.g., to measure household income, household composition and so on). For more information on sample status see the section on [Interview & Fieldwork Process](#) in the User Guide.

<sup>3</sup> A dwelling unit is a living space with its own front door this can be either a street door or a door within a house or block of flats.

<sup>4</sup> A household is defined as when one person or a group of people have the accommodation as their only or main residence and share at least one meal a day or share the living accommodation.

<sup>5</sup> Since March 2020, due to the Covid19 pandemic, all face-to-face interviewing has been suspended and everyone was invited to participate online.

<sup>6</sup> The datafile excludes responses collected during proxy interviews (when someone is not able to respond they may nominate someone else to complete a shorter questionnaire of factual questions on their behalf) and data collected from the BHPS and IEMB samples which were added later (see Endnote 1). It also excludes core sample members who turn 16 after the initial waves and then become eligible for adult interviews.

<sup>7</sup> Here sex is considered to be the biological sex at birth as reported by the respondent and not gender identity.