



# Using Neighbourhood Data in Understanding Society

presented by Gundi Knies and Min Zhang  
Institute for Social and Economic Research

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research.



# Aims of the course



Understanding Society facilitates exciting and innovative longitudinal research about the UK's society, including about how people shape places and places shape people.

- Learn directly from the experts about the study
- Gain skills in thinking about which questions can be addressed & how
- Learn about the data & navigate the online resources
- Learn how to do stuff in Stata
- Expand your network of researchers from a wide range of disciplinary backgrounds

# Outline



Time	Topic	Where?
10.00-10.30	Registration	SSCR (Foyer)
10.30-11.30	Understanding Society study in a nutshell	Essex Lab
11.30-12.30	Use of neighbourhood data in empirical research (group exercise)	
12.30-13.15	Lunch break	4.08
13.15-14.00	Example 1: Using information collected in the survey	
14.00-15.00	Example 2: Using linked geographical data	
15.00-15.15	Coffee break	
15.15-16.15	Example 3: Exploiting change over time	
16.15-16.30	Close	

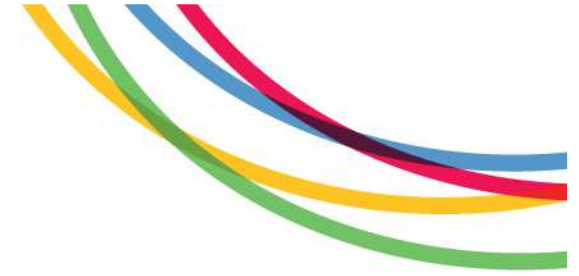
# What is *Understanding Society*?



- High quality longitudinal data aimed to help us understand the short/medium/long term effects of social and economic change
- The Study collects both *objective* and *subjective* indicators of different aspects of life
- It offers opportunities for research across multiple disciplines (e.g. sociology, economics, geography, psychology and health sciences)

# Key strengths:

## Longitudinal and long-term



Repeated cross-sectional data	Longitudinal data
<ul style="list-style-type: none"><li>• Data from different samples collected at different points in time</li><li>• Snapshots of the population, observation of the change on societal level</li></ul>	<ul style="list-style-type: none"><li>• Data from the same sample collected at different points in time</li><li>• Possible to observe change at the individual level, easier to draw causal inferences</li></ul>



# Key strengths:

- Household focus



Data collected from every adult and child aged 10+

- Statistical precision



Large sample sizes -> analysis possible at regional and country level

- Ethnic minority and immigrant boost samples

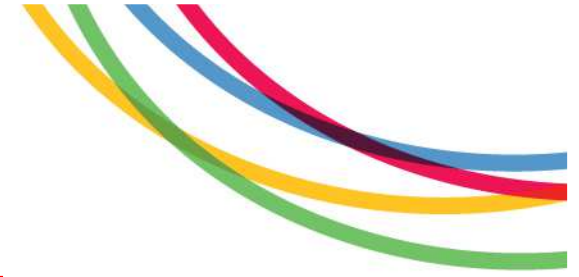
Enables meaningful analysis of the make-up of UK society



- Multi-topic



# Content: Measuring the full richness of lives and pressing societal issues



## Key Topics: significant research domains

- Education
- Employment
- Family and household
- Health, health behaviours, wellbeing
- Income, housing, wealth, expenditure & deprivation
- Ethnicity

# Content: “Supporting” topics



Contextual or explanatory factors for research on key topics **and** potential research foci in their own right:

- Neighbourhood characteristics
- Preferences, expectations across topic domains
- Social networks, support, reciprocity
- Transport
- Time Use
- Personality traits, identity, beliefs



# Content: Additional topics



Important research topics, which benefit from the design of Understanding Society, but more limited in their contribution:

- Environmental behaviours
- Political behaviour
- Leisure activities

# Target population and samples



## The BHPS (British Household Panel Survey) element of Understanding Society:

- 1991 - began with a representative probability sample of the residential population living in private households in Britain
- 1997 - Scotland and Wales boost samples added
- 2001 - Northern Ireland sample added -> BHPS became representative of the residential population living in private households in the United Kingdom.

# Target population and samples



General Population (GP) sample 2009/10: 26,000  
UK households or around 100,000 individuals

- A much LARGER sample was drawn to increase statistical power
- The large sample spread across the UK allows geographical analysis and analysis of small populations

# Target population and samples



PLUS Ethnic Minority Boost (EMB) sample: 4,000 households with at least one ethnic minority individual

- To allow ethnicity related longitudinal analysis
- Target minority groups: Indian, Pakistani, Bangladeshi, Caribbean, and African

Refreshed in 2015 this time additionally boosting immigrants to the UK (Immigrant and Ethnic Minority (IEMB) sample)

- Designed to provide around 2,000 adult immigrant respondents and around 2,500 from the target ethnic minority groups

# Development of the adult sample over time

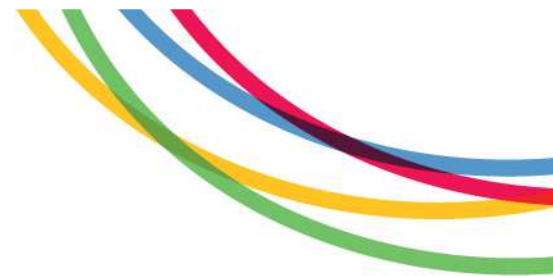
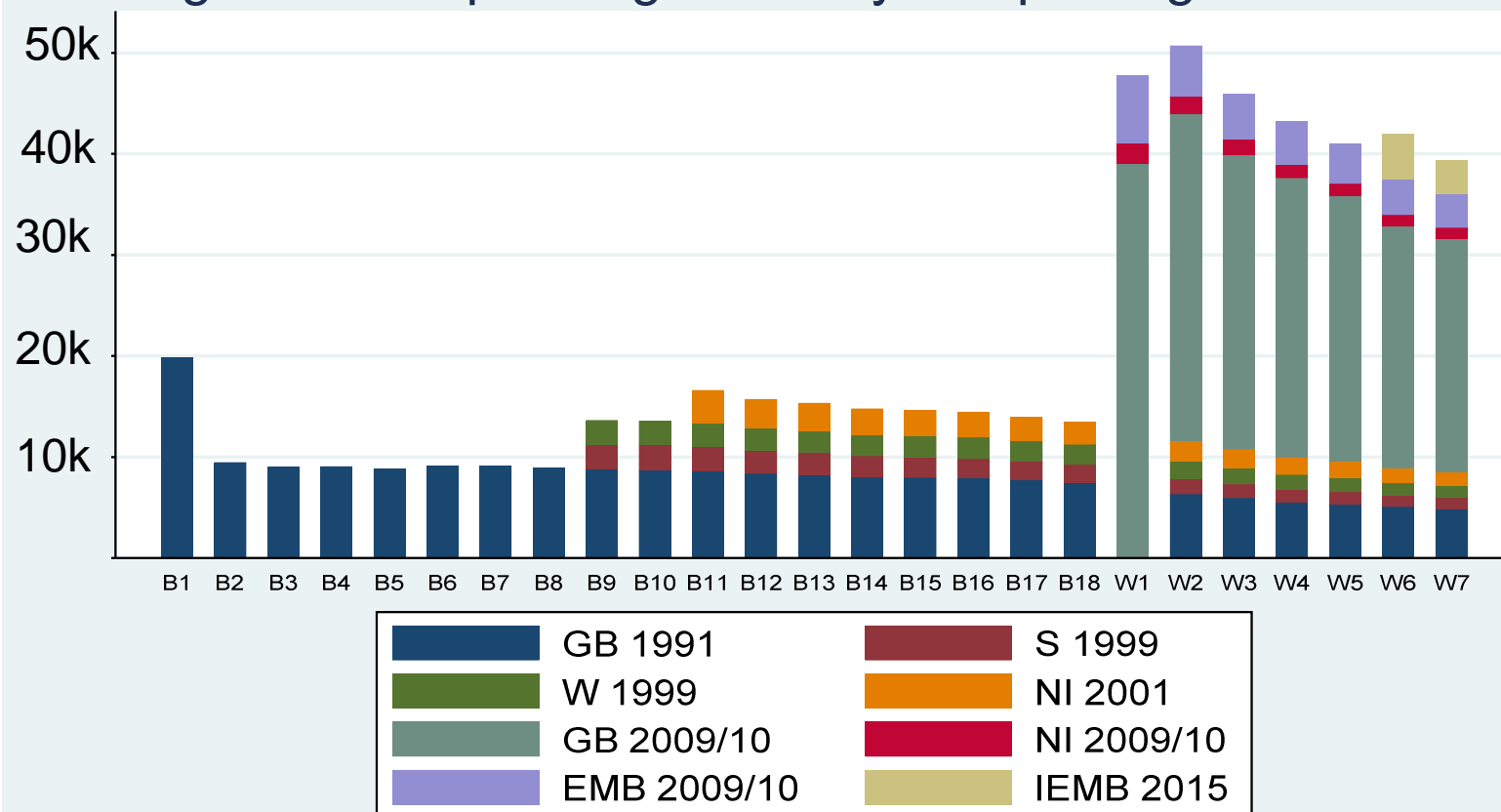


Figure 2: Responding adults by sample origin and wave



Source: Understanding Society, 2018, 7.1

# Development of the youth sample over time

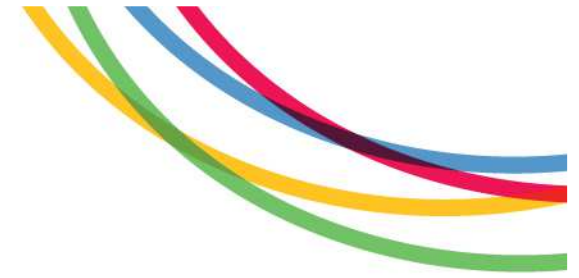
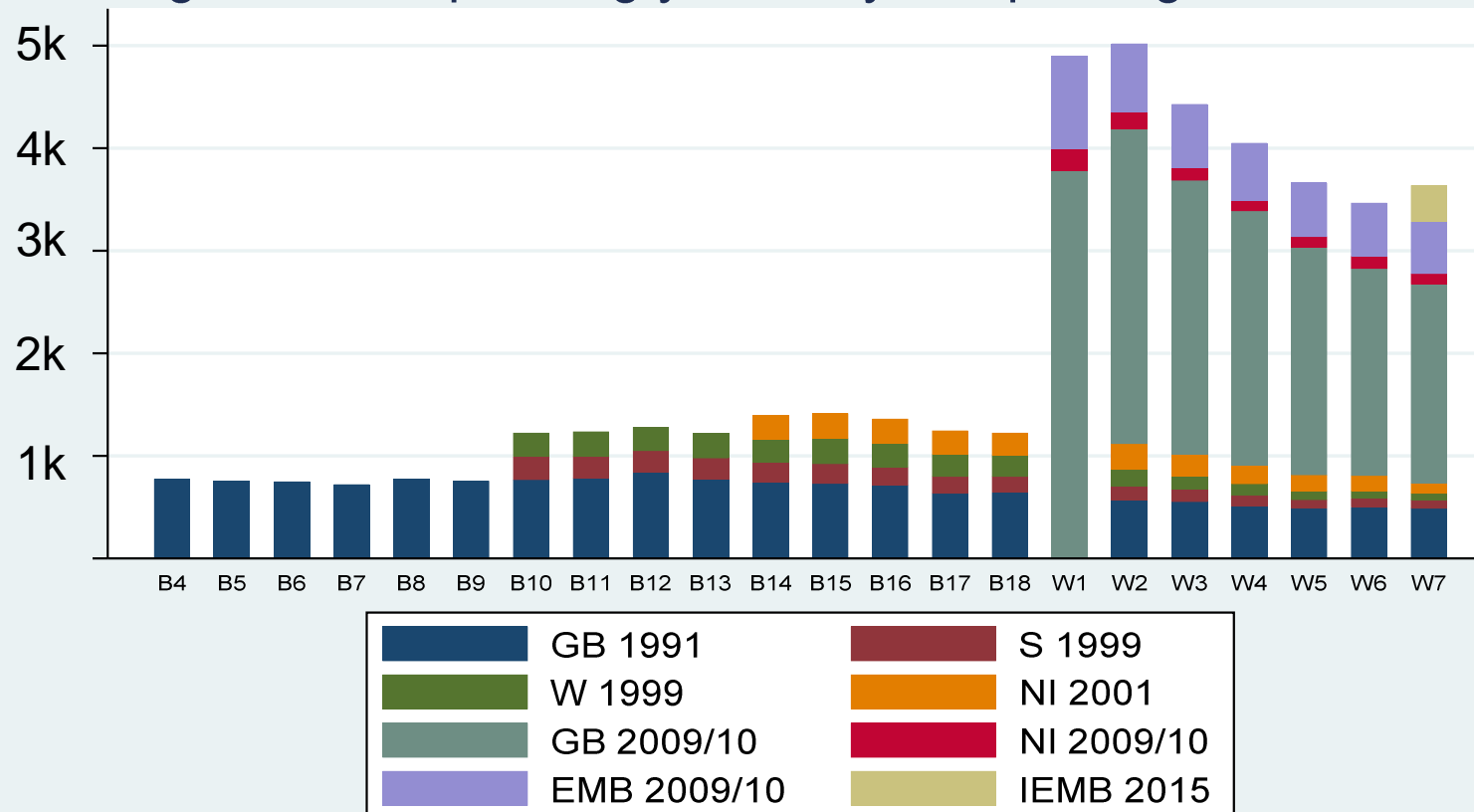
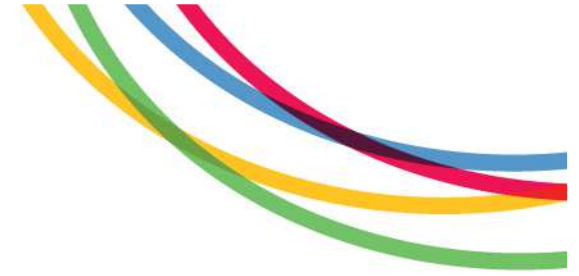


Figure 3: Responding youths by sample origin and wave



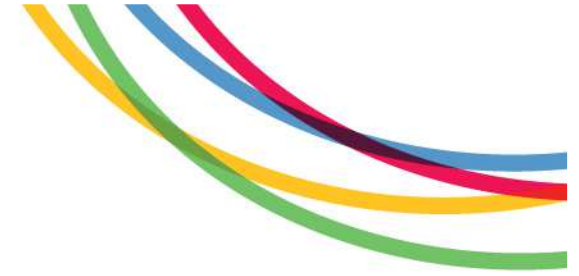
Source: Understanding Society, 2018, 7.1

# Data collection



- Prospective survey with retrospective elements
- Indefinite life
- Most of the data is collected using **face-to-face** interviews with adults (aged 16+), but there are a large number of additional survey instruments

# Data collection



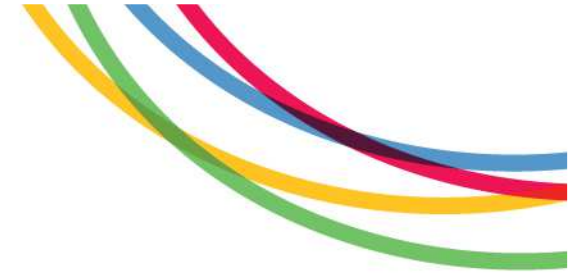
- Since 1994, children aged 11-15 years are asked to complete a short **self-completion youth questionnaire**
- Since 2009/10 children aged 10 additionally became eligible for the youth questionnaire.

## In addition to data directly collected in the survey

- Linked to spatial context data (e.g. Census)
- Linked to administrative data (e.g. NPD)
- Collection of biological specimens during a health assessment



# Interview structure



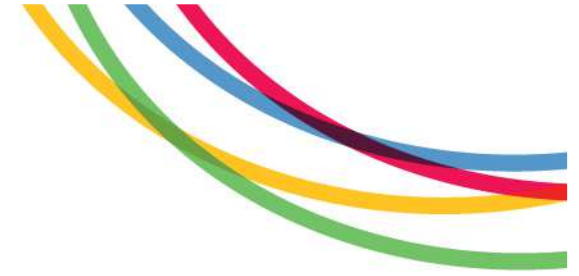
- Address record file

Interviewer collects basic information about the address and neighbourhood upon arrival

- Enumeration grid

If contact successfully made, interviewer collects basic information about each member of the household and within-household relationships (can be completed by any member of the household).

# Interview structure



## Household interview module

- Includes questions about housing, housing costs, and deprivation, among others (typically completed by the head of household)

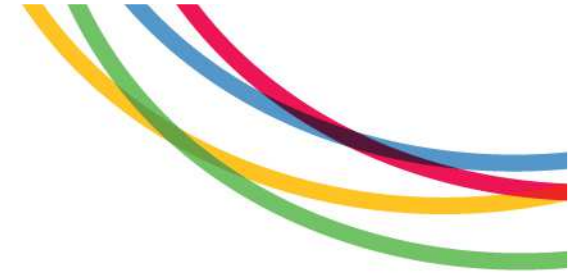
## Individual adult interview

- The main part of the survey completed by every adult (16+) living in the household

## Youth self-completion interview

- Short questionnaire handed out to all eligible children in the household

# Long term content



- Questions are repeated, which is what allows us to look at change over time.
- Not every question appears in every year – Rotating Modules, Event triggered and Age triggered questions
- Can get an idea of which themes are in each wave by using the long term content plan, or the online documentation list of questionnaire modules.

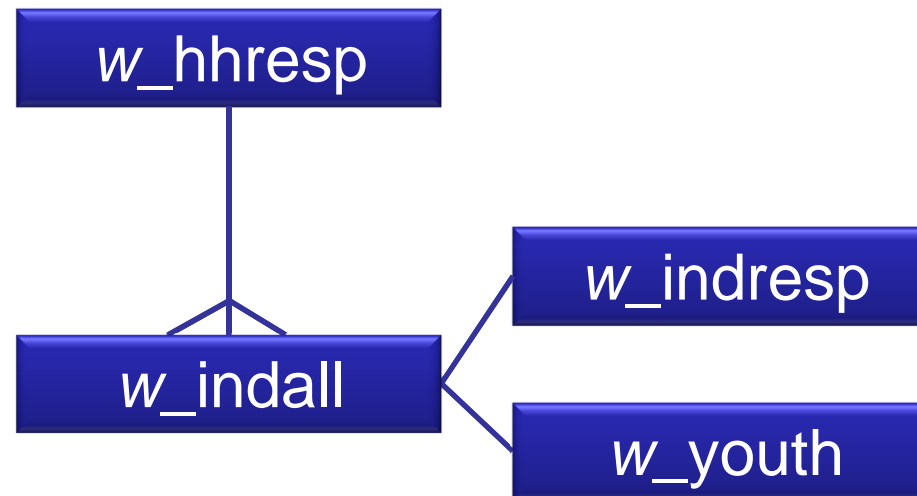
<https://www.understandingsociety.ac.uk/documentation/mainstage/long-term-content-plan>

# In the Understanding Society data from the UK Data Service...

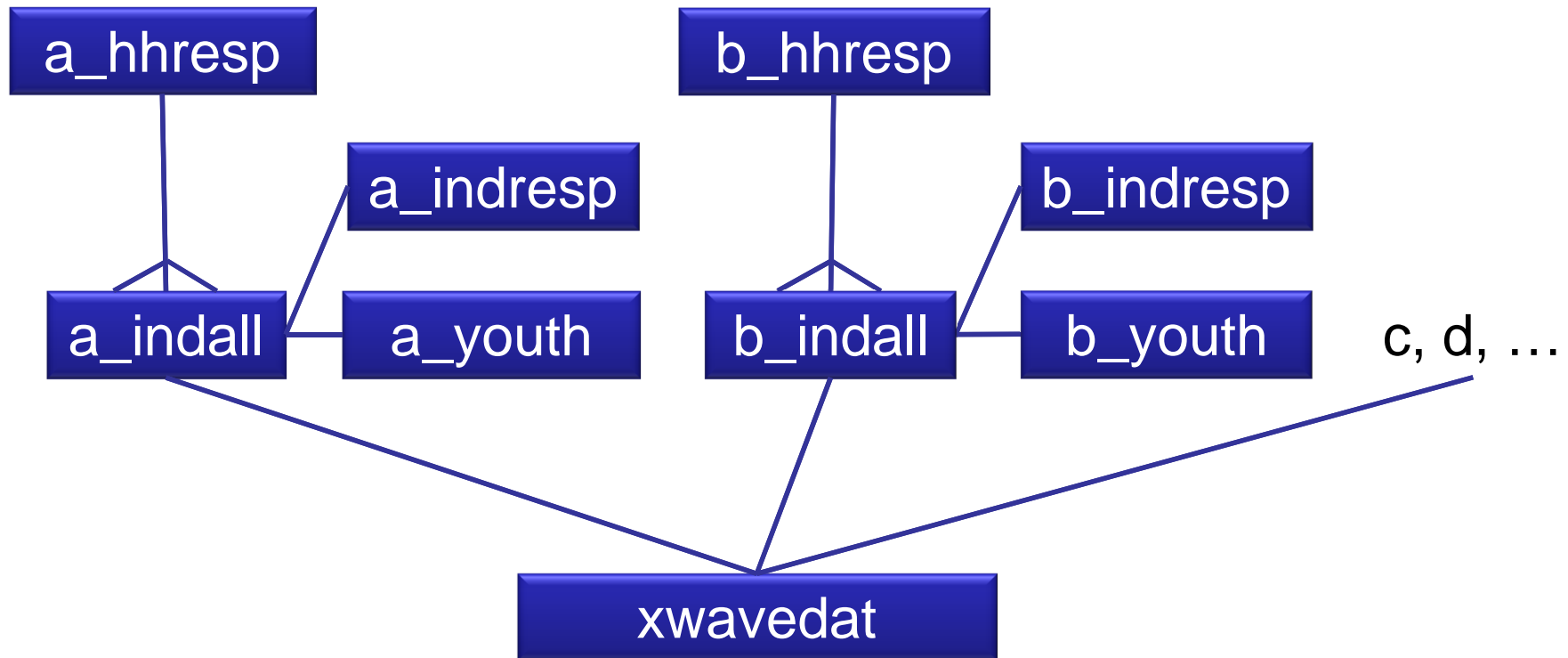
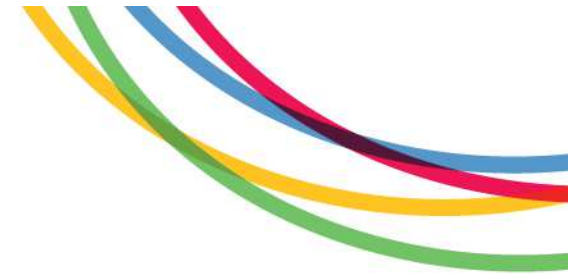


- Data for different waves are presented in separate files
- File names begin with a prefix designating the wave of data collection (“a\_” for Wave 1, “b\_” for Wave 2 etc., and “w\_” for waves in general)
- Data collected in the BHPS has a “b” in from of the wave prefix, i.e. “bw\_”
- A small number of files do not have wave prefixes as they store information across all waves

# Key data file structure within a wave



# Key data file structure



# There also is a bit of structure to variables...

- Naming convention of data files also applies to variables, **SAME ROOT NAME WITH WAVE PREFIX, w\_**  
Age and sex in Wave 1: a\_dvage, a\_sex  
Age and sex in Wave 2: b\_dvage, b\_sex  
Age and sex in Wave 3: c\_dvage, c\_sex
- Most derived variables are identified by suffix **\_dv**  
Usual gross monthly pay: a\_paygu\_dv, b\_paygu\_dv, ...  
HINT: Derived variables are typically placed at the bottom of the data file.

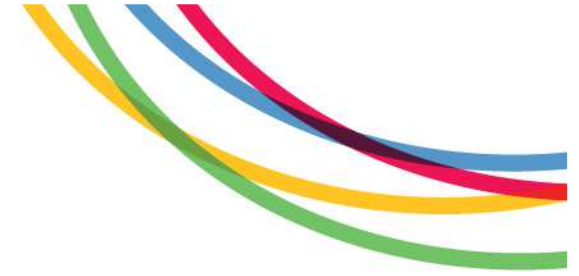
# How do we identify individuals and households?



- pidp → unique person identifier to link across waves and across files within the same wave; **constant over time**, no wave prefix for this variable
- w\_hidp → household identifier to identify members of the same households within one wave
- w\_pno → person number within the household in that wave



# Identifying individuals and households across waves

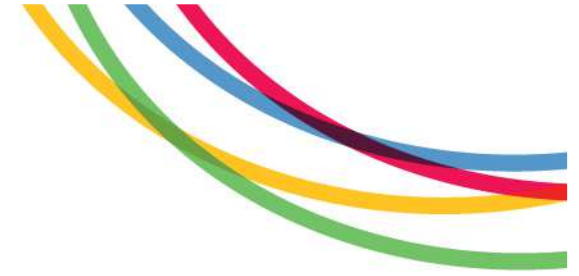


a_hidp	a_pno	pidp	a_sex	a_dvage
100056	1	1005789	Female	48
100056	2	2598635	Male	47
100056	3	1232659	Female	17

- Individuals within the same wave: pidp OR w\_hidp and w\_pno
- Individuals across waves: pidp
- Households within the same wave: w\_hidp

b_hidp	b_pno	pidp	b_sex	b_dvage
122345	1	1232659	Female	18
122345	2	1005789	Female	49
597856	1	2598635	Male	48

# Missing values



- In Understanding Society data, when the answers to questions are missing negative values are assigned to these cases. Each negative value shows the reason why the response to the question was missing.
- These are the most common, but there are others.

Missing Value	Applied to cases where the reason for the data being missing is
-9	Missing or wild
-8	Respondent was not eligible for that question and so it was not asked
-7	Respondent is a proxy respondent and the question is not in the proxy questionnaire
-2	Respondent refused to answer the question
-1	Respondent said he/she did not know the answer

# Understanding Society Website



DATA & DOCUMENTATION ▾ RESEARCH ▾ PARTICIPANTS ▾ ABOUT US ▾

## THE UK HOUSEHOLD LONGITUDINAL STUDY

Capturing life in the UK in the 21st century


Understanding Society is the largest longitudinal household panel study of its kind and provides vital evidence on life changes and stability

ABOUT THE STUDY

USING THE DATA



# Computing FAQ



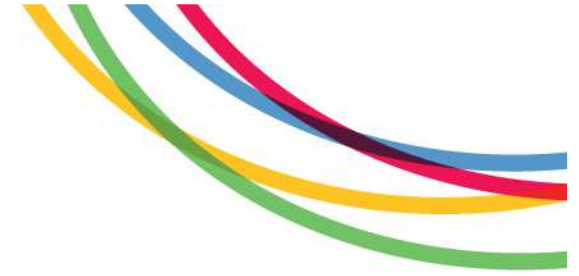
DO NOT LOSE THE LOGIN CARD &  
RETURN IT TO US AT THE END OF  
THE COURSE

Log out and log back in as shown below

## How do I login?

1. Login for the first time using **signup** both as login and password
2. You will be asked for your login and password, now use the ones printed on the card you received at registration
3. Next when prompted, write in your **first and last name** in the boxes specified and read and **accept the terms and conditions** of computer usage.
4. Now **your temporary account has been activated** and will remain activated until the end of the workshop. This means you will not have to repeat the activation process every time you login – **JUST USE THE LOGIN AND PASSWORD ON THE CARD.**
5. This is your temporary University of Essex account. Please use this even if you have a permanent University of Essex account.

# Computing FAQ



## **Where is the course material?**

<\\iserlin1.essex.ac.uk\\ConferenceData\\NbrDataInUnderstandingSociety\\>

You are allowed to copy these course materials and take with you

## **Where is the data stored?**

<\\iserlin1.essex.ac.uk\\ConferenceData\\ukhls\\>

<\\iserlin1.essex.ac.uk\\ConferenceData\\ukhls\\SL>

You are not allowed to copy any data files to take with you but you have permission to use these for this course.

## **Where can you save your files?**

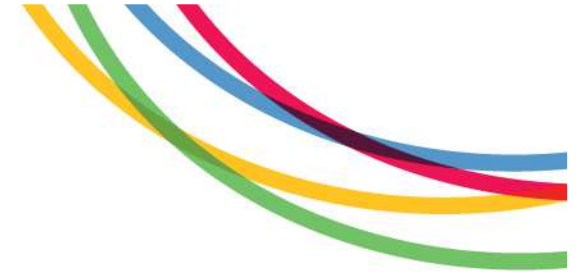
[M:\\](#)

We will also ask you to create the folder:

[M:\\mySLdata](#)

Do not copy the data files you save there. They are fake data.

# Computing FAQ



## **What to do if you cannot see Stata icon on your desktop?**

Start > Programmes > Applications > Stata 14

Then you can create a shortcut on your desktop

## **What to do if the computer restarts automatically?**

The computer will go back to the generic EssexLab login. So, you will have to log out and log in with your login.

# Understanding Society Website



DATA & DOCUMENTATION ▾ RESEARCH ▾ PARTICIPANTS ▾ ABOUT US ▾

## THE UK HOUSEHOLD LONGITUDINAL STUDY

Capturing life in the UK in the 21st century

Understanding Society is the largest longitudinal household panel study of its kind and provides vital evidence on life changes and stability

ABOUT THE STUDY

USING THE DATA





# Online Documentation



DATA & DOCUMENTATION ▾

RESEARCH ▾

PARTICIPANTS ▾

ABOUT US ▾

## DATA & DOCUMENTATION

User guides, fieldwork documents, questionnaires, technical reports.

Home > Data and documentation

Data collected from the survey's thousands of participants is securely stored by the [UK Data Service](#), from where researchers can access it online. All of the associated documentation is available here.

Main survey	Health and biomarkers	Innovation panel
OVERVIEW	OVERVIEW	OVERVIEW
USER GUIDE	USER GUIDE	USER GUIDE
DATASET DOCUMENTATION	DATASET DOCUMENTATION	DATASET DOCUMENTATION
QUESTIONNAIRES	QUESTIONNAIRES	QUESTIONNAIRES
TECHNICAL REPORTS		TECHNICAL REPORTS
FIELDWORK DOCUMENTS	FIELDWORK DOCUMENTS	FIELDWORK DOCUMENTS
LONG TERM CONTENT PLAN		
QUALITY PROFILE		
<a href="#">SEE MORE</a>	<a href="#">SEE MORE</a>	<a href="#">SEE MORE</a>

<https://www.understandingsociety.ac.uk/documentation>



# Questionnaires



DATA & DOCUMENTATION ▾

RESEARCH ▾

PARTICIPANTS ▾

ABOUT US ▾

## DATA & DOCUMENTATION

User guides, fieldwork documents, questionnaires, technical reports.

Home > Data and documentation

Data collected from the survey's thousands of participants is securely stored by the [UK Data Service](#), from where researchers can access it online. All of the associated documentation is available here.

Main survey	Health and biomarkers	Innovation panel
OVERVIEW	OVERVIEW	OVERVIEW
USER GUIDE	USER GUIDE	USER GUIDE
DATASET DOCUMENTATION	DATASET DOCUMENTATION	DATASET DOCUMENTATION
<b>QUESTIONNAIRES</b>	QUESTIONNAIRES	QUESTIONNAIRES
TECHNICAL REPORTS	FIELDWORK DOCUMENTS	TECHNICAL REPORTS
FIELDWORK DOCUMENTS		FIELDWORK DOCUMENTS
LONG TERM CONTENT PLAN		
QUALITY PROFILE		
SEE MORE	SEE MORE	SEE MORE

# Questionnaires (pdf files)



DATA & DOCUMENTATION ▾ RESEARCH ▾ PARTICIPANTS ▾ ABOUT US ▾

## MAIN SURVEY

Home > Data and documentation > Main survey > Questionnaires

## Questionnaires

Youth and Adult

We use different types of questionnaire depending on who is answering the question. Every section of the questionnaire, and each question, is answered voluntarily.

- Each adult participant, aged 16 and above, answers a questionnaire asking them about different aspects of their life.
- For households where there are children under the age of 10, parents or carers answer a set of questions about the children in their care.
- There is also a household information questionnaire which one adult in the house completes on behalf of the whole household.
- Young people aged 10-15 complete a short questionnaire looking at different parts of their life, after their parent or carer has given permission for them to complete it.

WAVE 10 (2018-2020)

WAVE 9 (2017-2019)

WAVE 8 (2016-2018)

WAVE 7 (2015-2017)

WAVE 6 (2014-2016)

WAVE 5 (2013-2015)

WAVE 4 (2012-2014)

WAVE 3 (2011-2013)

WAVE 2 (2010-2012)

WAVE 1 (2009-2011)

### Wave 1

- [Mainstage - questionnaire \(CAPI\)](#)
- [Mainstage - adult self-completion questionnaire](#)
- [Mainstage - youth self-completion questionnaire](#)

OVERVIEW

SURVEY TIMELINE

USER GUIDES

DATASET DOCUMENTATION

INDEX TERMS

QUESTIONNAIRES

TECHNICAL REPORTS

FIELDWORK DOCUMENTS

LONG TERM CONTENT PLAN

QUALITY PROFILE

# Questionnaires (pdf files)

[DATA & DOCUMENTATION](#)[RESEARCH](#)[PARTICIPANTS](#)[ABOUT US](#)

## MAIN SURVEY

[Home](#) > [Data and documentation](#) > [Main survey](#) > [Questionnaires](#)

## Questionnaires

Youth and Adult

We use different types of questionnaire depending on who is answering the question. Every section of the questionnaire, and each question, is answered voluntarily.

- Each adult participant, aged 16 and above, answers a questionnaire asking them about different aspects of their life.
- For households where there are children under the age of 10, parents or carers answer a set of questions about the children in their care.
- There is also a household information questionnaire which one adult in the house completes on behalf of the whole household.
- Young people aged 10-15 complete a short questionnaire looking at different parts of their life, after their parent or carer has given permission for them to complete it.

**WAVE 10** (2018-2020)

**WAVE 9** (2017-2019)

**WAVE 8** (2016-2018)

**WAVE 7** (2015-2017)

**WAVE 6** (2014-2016)

**WAVE 5** (2013-2015)

**WAVE 4** (2012-2014)

**WAVE 3** (2011-2013)

**WAVE 2** (2010-2012)

**WAVE 1** (2009-2011)

### Wave 1

[Mainstage - questionnaire \(CAPI\)](#)

[Mainstage - adult self-completion questionnaire](#)

[Mainstage - youth self-completion questionnaire](#)

Future waves

OVERVIEW

SURVEY TIMELINE

USER GUIDES

DATASET DOCUMENTATION

INDEX TERMS

QUESTIONNAIRES

TECHNICAL REPORTS

FIELDWORK DOCUMENTS

LONG TERM CONTENT PLAN

QUALITY PROFILE

# Dataset Documentation

[DATA & DOCUMENTATION](#) ▾[RESEARCH](#) ▾[PARTICIPANTS](#) ▾[ABOUT US](#) ▾

## MAIN SURVEY

[Home](#) > [Data and documentation](#) > [Main survey](#) > [Questionnaires](#)

## Questionnaires

Youth and Adult

We use different types of questionnaire depending on who is answering the question. Every section of the questionnaire, and each question, is answered voluntarily.

- Each adult participant, aged 16 and above, answers a questionnaire asking them about different aspects of their life.
- For households where there are children under the age of 10, parents or carers answer a set of questions about the children in their care.
- There is also a household information questionnaire which one adult in the house completes on behalf of the whole household.
- Young people aged 10-15 complete a short questionnaire looking at different parts of their life, after their parent or carer has given permission for them to complete it.

**WAVE 10** (2018-2020)

**WAVE 9** (2017-2019)

**WAVE 8** (2016-2018)

**WAVE 7** (2015-2017)

**WAVE 6** (2014-2016)

**WAVE 5** (2013-2015)

**WAVE 4** (2012-2014)

**WAVE 3** (2011-2013)

**WAVE 2** (2010-2012)

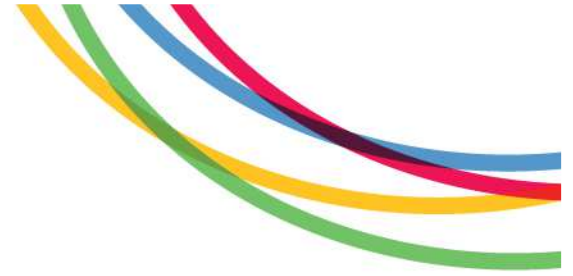
**WAVE 1** (2009-2011)

### Wave 1

- [Mainstage - questionnaire \(CAPI\)](#)
- [Mainstage - adult self-completion questionnaire](#)
- [Mainstage - youth self-completion questionnaire](#)

[OVERVIEW](#)[SURVEY TIMELINE](#)[USER GUIDES](#)[DATASET DOCUMENTATION](#)[INDEX TERMS](#)[QUESTIONNAIRES](#)[TECHNICAL REPORTS](#)[FIELDWORK DOCUMENTS](#)[LONG TERM CONTENT PLAN](#)[QUALITY PROFILE](#)

# Dataset Documentation



Provides a search function to search for variables, data files, questions or question modules.

- Find out about the context in which information was collected, which variables are associated with derived variables, whether a module was asked in this wave, which other survey(s) carried this question, and response frequencies ..

# Dataset documentation



DATA & DOCUMENTATION ▾ RESEARCH ▾ PARTICIPANTS ▾ ABOUT US ▾

## MAIN SURVEY

Home > Documentation > Main survey > Dataset documentation

## Dataset documentation

Understanding Society collects information from everyone aged 10 and over in 40,000 UK households. For more background information read our [About Understanding Society Guide](#).

**DATAFILES** QUESTIONNAIRES

Click on a tab to see files for specific waves

ALL WAVES (1991-2017)

XWAVE (1991-2017)

WAVE 7 (2015-2017)

WAVE 6 (2014-2016)

WAVE 5 (2013-2015)

WAVE 4 (2012-2014)

WAVE 3 (2011-2013)

WAVE 2 (2010-2012)

WAVE 1 (2009-2011)

WAVE B18 (2008)

WAVE B17 (2007)

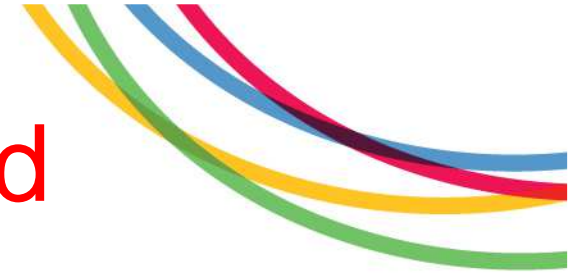
### All Waves (284 Datafiles)

Datafile	Description	UKHLS	BHPS
egoalt	Kin and other relationships between pairs of individuals in the household	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhresp	Substantive data from responding households	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhsamp_bh	Sample and Household level data for issued households		B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
income	Income and payment information	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13

Stem-name of data file

Brief description of content

# Data file names correspond quite well to instruments ...

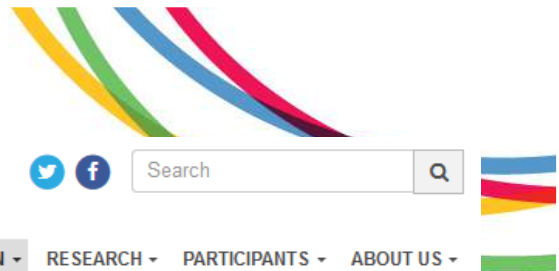


- Data collected from different sources e.g. the household interview, the adult interview, the youth interview, are stored in separate files
- Key data files for analysing data from responding households and individuals

Filename	Description
bw_hhresp w_hhresp	Substantive data from responding households
bw_indresp w_indresp	Substantive data from responding adults (16+) including self-completion questions, proxies and telephone interviews
bw_youth w_youth	Substantive data from youth questionnaire (UKHLS: age 10-15, all waves; BHPS: age 11-15, Waves 4-19 only)
xwavedat	Includes stable characteristics of all individuals in the household, including responses reported when first entering the study



# Understanding Society Files



DATA & DOCUMENTATION ▾

RESEARCH ▾

PARTICIPANTS ▾

ABOUT US ▾

## MAIN SURVEY

Home > Documentation > Main survey > Dataset documentation

## Dataset documentation

Understanding Society collects information from everyone aged 10 and over in 40,000 UK households. For more background information read our [About Understanding Society Guide](#).

Mainstage Variable ▾

SEARCH

DATAFILES

QUESTIONNAIRES

ALL WAVES (1991-2017)

XWAVE (1991-2017)

WAVE 7 (2015-2017)

WAVE 6 (2014-2016)

WAVE 5 (2013-2015)

WAVE 4 (2012-2014)

WAVE 3 (2011-2013)

WAVE 2 (2010-2012)

WAVE 1 (2009-2011)

WAVE B18 (2008)

WAVE B17 (2007)

WAVE B16 (2006)

### All Waves (284 Datafiles)

Datafile	Description	UKHLS	BHPS
egoalt	Kin and other relationships between pairs of individuals in the household	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhresp	Substantive data from responding households	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhsamp_bh	Sample and Household level data for issued households		B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
income	Income and payment information	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13,

OVERVIEW

SURVEY TIMELINE

USER GUIDES

DATASET DOCUMENTATION

QUESTIONNAIRES

TECHNICAL REPORTS

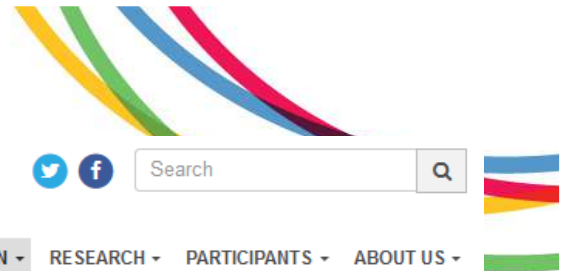
FIELDWORK DOCUMENTS

LONG TERM CONTENT PLAN

QUALITY PROFILE



# BHPS Files (*starting with “B”*)



DATA & DOCUMENTATION ▾

RESEARCH ▾

PARTICIPANTS ▾

ABOUT US ▾

## MAIN SURVEY

Home > Documentation > Main survey > Dataset documentation

## Dataset documentation

Understanding Society collects information from everyone aged 10 and over in 40,000 UK households. For more background information read our [About Understanding Society Guide](#).

Mainstage Variable ▾

SEARCH

DATAFILES

QUESTIONNAIRES

ALL WAVES (1991-2017)

**XWAVE (1991-2017)**

WAVE 7 (2015-2017)

WAVE 6 (2014-2016)

WAVE 5 (2013-2015)

WAVE 4 (2012-2014)

WAVE 3 (2011-2013)

WAVE 2 (2010-2012)

WAVE 1 (2009-2011)

**WAVE B18 (2008)**

**WAVE B17 (2007)**

WAVE B16 (2006)

### All Waves (284 Datafiles)

Datafile	Description	UKHLS	BHPS
egoalt	Kin and other relationships between pairs of individuals in the household	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhresp	Substantive data from responding households	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhsamp_bh	Sample and Household level data for issued households		B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
income	Income and payment information	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13,

OVERVIEW

SURVEY TIMELINE

USER GUIDES

**DATASET DOCUMENTATION**

QUESTIONNAIRES

TECHNICAL REPORTS

FIELDWORK DOCUMENTS

LONG TERM CONTENT PLAN

QUALITY PROFILE

# Dataset documentation



DATA & DOCUMENTATION ▾ RESEARCH ▾ PARTICIPANTS ▾ ABOUT US ▾

## MAIN SURVEY

Home > Documentation > Main survey > Dataset documentation

## Dataset documentation

Understanding Society collects information from everyone aged 10 and over in 40,000 UK households. For more background information read our [About Understanding Society Guide](#).

search Mainstage Variable SEARCH

DATAFILES QUESTIONNAIRES

ALL WAVES (1991-2017)

XWAVE (1991-2017)

WAVE 7 (2015-2017)

WAVE 6 (2014-2016)

WAVE 5 (2013-2015)

WAVE 4 (2012-2014)

WAVE 3 (2011-2013)

WAVE 2 (2010-2012)

WAVE 1 (2009-2011)

WAVE B18 (2008)

WAVE B17 (2007)

### All Waves (284 Datafiles)

Datafile	Description	UKHLS	BHPS
egoalt	Kin and other relationships between pairs of individuals in the household	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhresp	Substantive data from responding households	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
hhsamp_bh	Sample and Household level data for issued households		B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, B16, B17, B18
income	Income and payment information	1, 2, 3, 4, 5, 6, 7	B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13,



Click on a # to see a specific UKHLS file for a specific wave



Click on a # to see a specific BHPS file for a specific wave



# If you click on indresp for Wave 1 you will see



Understanding Society  
THE UK HOUSEHOLD LONGITUDINAL STUDY

Twitter Facebook Search

DATA & DOCUMENTATION RESEARCH PARTICIPANTS ABOUT US

MAIN SURVEY

Home > Documentation > Main survey > Dataset documentation > Wave 1 > Datafile: a\_indresp

## a\_indresp

Substantive data for responding adults (16+), incl. proxies

Variables (1417)

Search in these variables

Name	Description
<a href="#">a_hidp</a>	household identifier (public release)
<a href="#">a_pno</a>	person number in household grid
<a href="#">pidp</a>	cross-wave person identifier (public release)
<a href="#">a_sex</a>	sex
<a href="#">a_dvage</a>	age for whole sample, from birth or ageif
<a href="#">a_istrtdatd</a>	date interview with this respondent was started - date of day
<a href="#">a_istrtdatm</a>	date interview with this respondent was started - month
<a href="#">a_istrtdaty</a>	date interview with this respondent was started - year
<a href="#">a_mvnever</a>	lived at address whole life
<a href="#">a_mvmonth</a>	month moved to current address
<a href="#">a_mvyr</a>	year moved to current address
<a href="#">a_lkmove</a>	prefers to move house

- OVERVIEW
- SURVEY TIMELINE
- USER GUIDES
- DATASET DOCUMENTATION**
- QUESTIONNAIRES
- TECHNICAL REPORTS
- FIELDWORK DOCUMENTS
- LONG TERM CONTENT PLAN
- QUALITY PROFILE

# If you click on a variable you will see

Variable name & label in data file

Name of datafile found in

Variable name in q'naire and description

Click on this for more information

a\_payruk  
father lived in uk

Datafile: [a\\_indresp](#)

Related to 1 Questions

[familybackground\\_w1.payruk](#)  
Father lived in UK

Universe if ( PACOB>4 ) (Father not born in the UK)  
Text Has your father ever lived in the UK?

Associated variables

• [a\\_payruk \(a\\_indresp\)](#) father lived in uk

Who was asked the question?

Exact question wording

Click on this to save to a basket

## Frequencies

Value label	Value	Absolute frequency	Relative frequency	
inapplicable	-8	35346	69.31%	<div><div></div></div>
proxy respondent	-7	3262	6.4%	<div><div></div></div>
refused	-2	2	0.0%	<div><div></div></div>
don't know	-1	24	0.05%	<div><div></div></div>
father lived in uk	1	6143	12.05%	<div><div></div></div>
father never lived in uk	2	6217	12.19%	<div><div></div></div>
Total		50994	100%	

Descriptive statistics for a continuous variable, or a discrete one with many categories

## Wave Occurrences

[1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#)

Waves asked in

# Online Documentation



DATA & DOCUMENTATION ▾

RESEARCH ▾

PARTICIPANTS ▾

ABOUT US ▾

## DATA & DOCUMENTATION

User guides, fieldwork documents, questionnaires, technical reports.

Home > Data and documentation

Data collected from the survey's thousands of participants is securely stored by the [UK Data Service](#), from where researchers can access it online. All of the associated documentation is available here.

Main survey	Health and biomarkers	Innovation panel
OVERVIEW	OVERVIEW	OVERVIEW
USER GUIDE	USER GUIDE	USER GUIDE
DATASET DOCUMENTATION	DATASET DOCUMENTATION	DATASET DOCUMENTATION
QUESTIONNAIRES	QUESTIONNAIRES	QUESTIONNAIRES
TECHNICAL REPORTS	FIELDWORK DOCUMENTS	TECHNICAL REPORTS
FIELDWORK DOCUMENTS		FIELDWORK DOCUMENTS
LONG TERM CONTENT PLAN		
QUALITY PROFILE		
SEE MORE	SEE MORE	SEE MORE

<https://www.understandingsociety.ac.uk/documentation>



## Part 2 -- Hands on

4 sessions, each with brief presentation followed by an exercise

# Potential for Neighbourhood research in Understanding Society



- Too many examples and applications. Some parameters:
  - Dependent or independent variable?
  - Cross-sectional or longitudinal research question?
  - Linked admin data? Commercial data? Predictions from within the survey?
    - What is the scale of the neighbourhood, is this important?
  - Are neighbourhood dynamics considered?
  - Is selection into neighbourhoods considered?

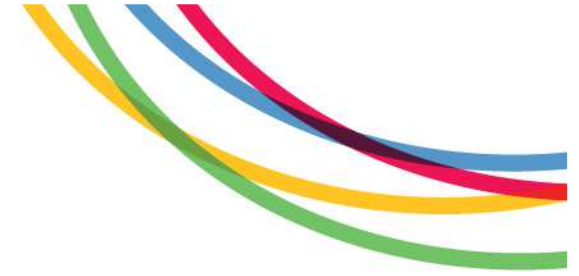
# Exercise (until 12:30)



- Starting with McCulloch (2003), read the paper(s) focussing on the data and methods sections to answer the research summary sheet as well as:
  - How is this an example of neighbourhood research?
  - Does the research use neighbourhood data from the survey, or linked geographical data, or both?
  - What challenges did the authors mention (if any), and how were they addressed?
- If the paper used UKHLS/BHPS data, can you find the neighbourhood variables in the Online Documentation?



# Neighbourhood research

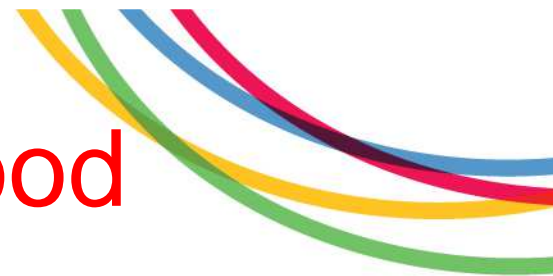


**McCulloch, A. (2003). "An examination of social capital and social disorganisation in neighbourhoods in the British household panel study." *Social Science & Medicine* 56(7): 1425-1438.**

## Additional options:

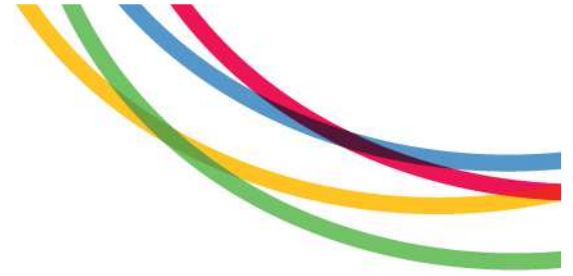
- Prior, L., D. Manley, et al. (2018). "Stressed out? An investigation of whether allostatic load mediates associations between neighbourhood deprivation and health." *Health & Place* **52**: 25-33.
- Knies, G., Nandi, A. & Platt, L. (2016). Life satisfaction, ethnicity and neighbourhoods: Is there an effect of neighbourhood ethnic composition on life satisfaction?
- Knies, G. (2017). Income effects on children's life satisfaction: Longitudinal evidence for England. ISER Working Paper Series 2017-02
- Knies, G. (2013), Neighbourhood social ties. *The British Journal of Sociology*, 64: 425-452. doi:10.1111/1468-4446.12026

# Worksheet 1: Neighbourhood data in the survey



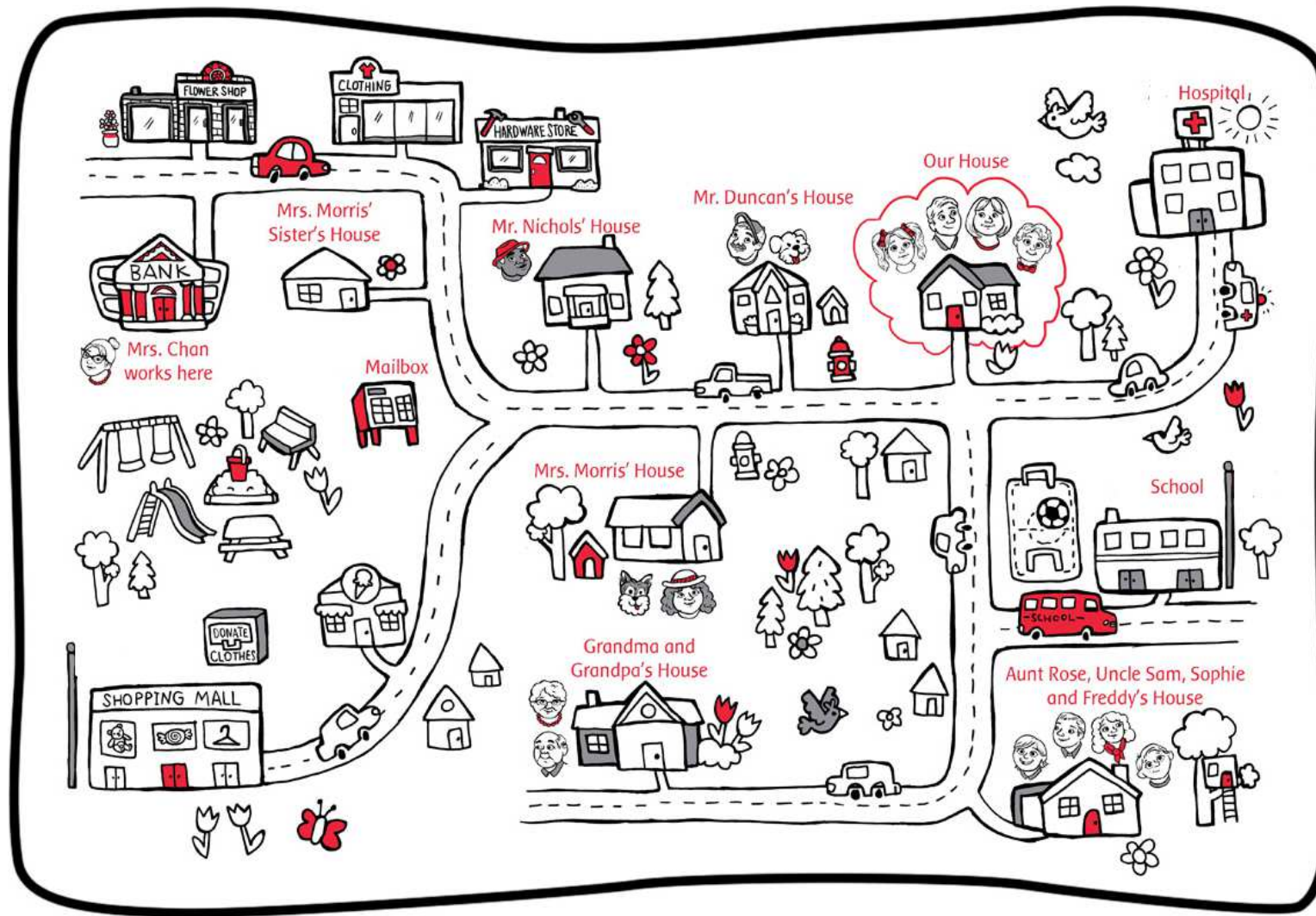
- Various sources within the survey data itself
  - Interviewer observations in the sampling file
  - Questions asked in the household questionnaire
  - Questions asked in the individual questionnaire
  - Questions asked in the self-completion component of the individual questionnaire
  - Questions asked in the youth self-completion questionnaire
- Plus linked geographical info (Worksheet 2)

# In this example you'll ...



- Examine how neighbourhood cohesion varies by individual, household and neighbourhood characteristics relying only on data from Wave 1
- Learn how to merge data files at different levels and with a different universe of cases
- Get to use some tried and trusted Stata commands that will make your work a little easier

# An example neighbourhood



# Merging files

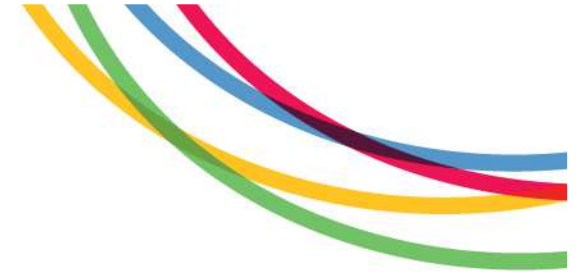
a_indresp			
a_hidp	a_pno	pidp	a_nbrsnci_dv
100	1	001	1.1
100	2	002	2.7
200	1	003	5
200	2	004	4.3

a_hhsamp	
a_hidp	a_vicini1
10	2
100	2
200	1

a_hhresp	
a_hidp	a_tenure_dv
100	1
200	2

Merged individual-level data file					
a_hidp	a_pno	pidp	a_nbrsnci_dv	a_tenure_dv	a_vicini1
100	1	001	1.1	1	1
100	2	002	2.7	1	2
200	1	003	5	2	0
200	2	004	4.3	2	0

# Merging two different level files

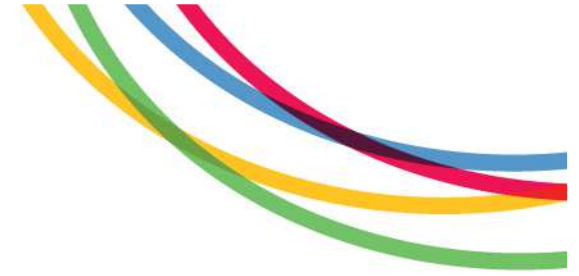


- What is the linking variable(s): **a\_hidp**
- The level of each dataset, that is, whether this variable uniquely defines each row

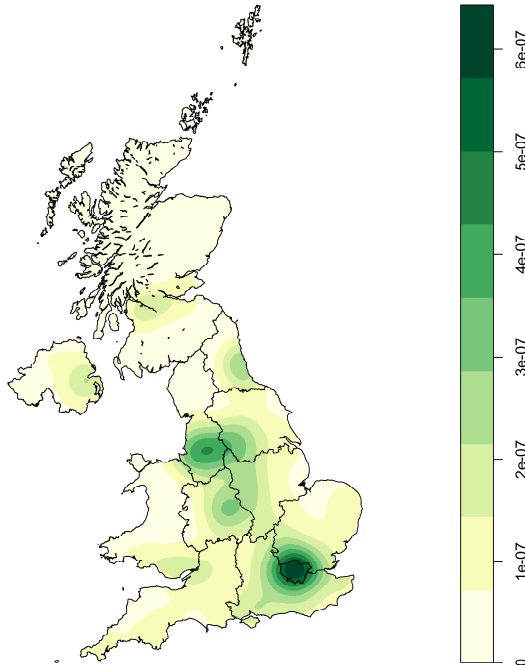
use a\_indresp, clear  
duplicates report a\_hidp  
use a\_hhresp, clear  
duplicates report a\_hidp

- You will find that a\_hidp uniquely determines each row of a\_hhresp but not a\_indresp. So, this will be a one-to-many OR many-to-one merge depending on the order of the files

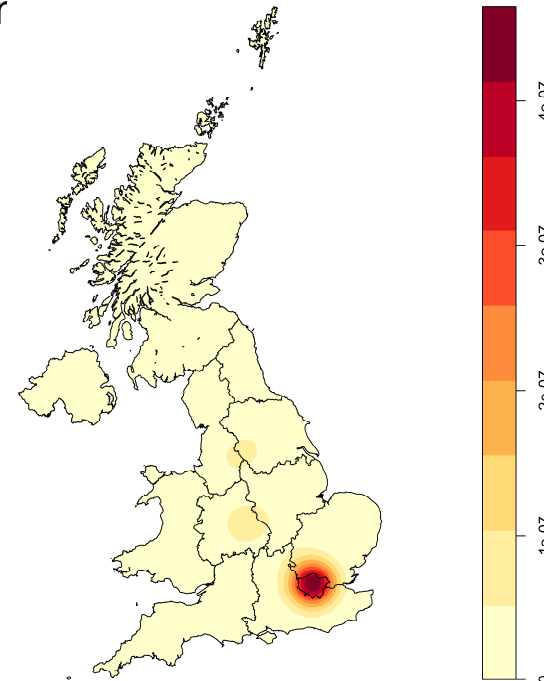
# Worksheet 2: Linking geographic data



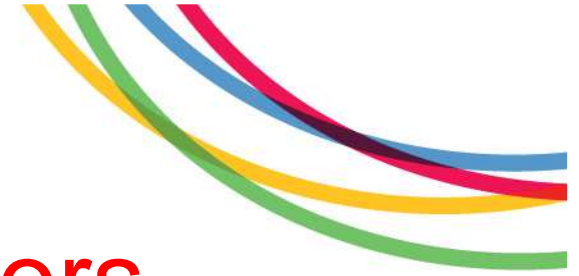
**General population** samples  
18 addresses randomly selected for  
interview from within 2,630 PSUs in 103  
strata, plus 2,395 addresses in Northern  
Ireland



Plus **Ethnic minority  
boost** sampled from 3,145  
postcode sectors with >5% EM  
cor



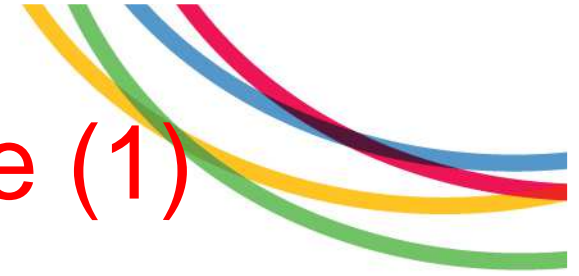
# From addresses to geographies and to indicators



- Start with sample households' exact addresses at the time of interview/fieldwork
- Match postcodes with the May release of the Year 2's ONS Postcode Directory containing:
  - Administrative, electoral, and health unit identifiers, and population-weighted centroids
  - Official neighbourhood classifications such as the Output Area Classification, and Rural-Urban indicators
  - For full list see ONSPD user guidance <http://geoportal.statistics.gov.uk/>



# Geographical data available (1)



Already included IDs/characteristics			
core	6614	EUL	Understanding Society: GOR_DV & URBAN_DV
core	6931	SL	Understanding Society: GOR_DV & URBAN_DV
linked	7454	SL	Census 2001 Rural-Urban Indicators
linked	7630	SL	Census 2011 Rural-Urban Indicators
linked	6674	SL	Census 2001 Output Area Classification
linked	7629	SL	Census 2011 Output Area Classification
linked	7453	SL	Acorn Types
linked	7533	SL	Geographical Accessibility (Waves 1-3)

Replace \$\$\$\$ by the study number to get to the data basket:

[https://discover.ukdataservice.ac.uk/catalogue/?sn=\\$\\$\\$\\$](https://discover.ukdataservice.ac.uk/catalogue/?sn=$$$$)

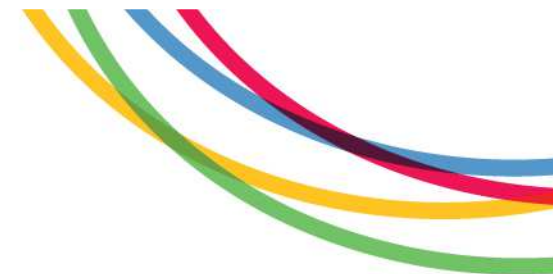
# Geographical data available (2)

## Merge your own geo-coded data using...

core	6676	SC	Understanding Society: postcode/Grid ref
link-id	7182	SL	School Codes (Wave 1)
link-id	6666	SL	Local Authority District
link-id	6668	SL	Westminster Parliamentary Constituencies
link-id	6671	SL	Local Education Authorities
link-id	6675	SL	Travel to Work Areas
link-id	6672	SL	Strategic Health Authorities
link-id	6673	SL	Primary Care Organisations
link-id	6669	SL	Census 2001 Area Statistics Wards
link-id	7245	SL	Census 2001 Middle Layer Super Output Areas
link-id	7249	SL	Census 2011 Middle Layer Super Output Areas
link-id	6670	SL	Census 2001 Lower Layer Super Output Areas
link-id	7248	SL	Census 2011 Lower Layer Super Output Areas

Test-run your external data merge on ONSPD

# Produce results for small geographies

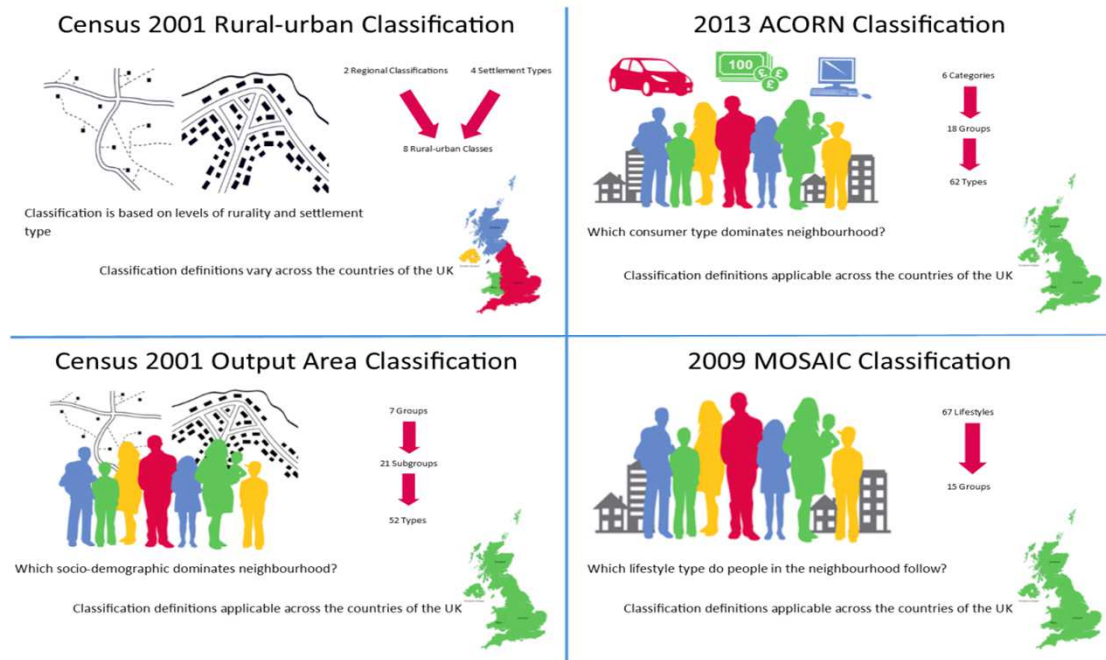


Geography	Number in sample	Number in UK	Sample members in unit (median and IQR)
OA 2001	26,446	223,019	2 (1;3)
LSOA 2001	13,437	34,379	4 (2;6)
MSOA 2001	4,783	7,195	11 (6;17)
TTWA	228	243	145 (64;321)
LAD	217	219	201 (112;330)
GOR	12	12	5,443 (3,829;6,784)

Base: Wave 1 Responding adults and children in their household

# Use small scale neighbourhood characteristics

For more information about these in the UKHLS, see:



## Exploring the Value of Understanding Society for Neighbourhood Effects Analyses

Gurdeep Kaur  
Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK  
g.kaur@essex.ac.uk

### Abstract

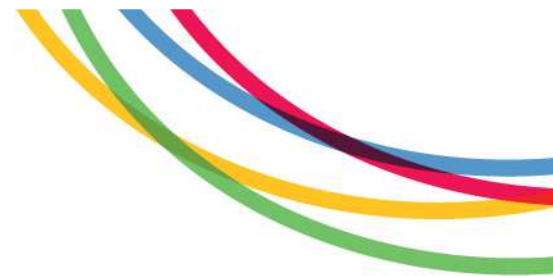
Related data set "Understanding Society: Waves 1-8, 2009-2015" with DOI <https://doi.org/10.5555/UKHLS-UK-001> in repository "UK Data Service".

Understanding Society is a large representative household panel study for the UK. The study follows the same 40,000 households over time, beginning in 2009 and providing a detailed picture of how people's lives are changing. One of the many innovative features of Understanding Society is that a great deal of information about neighbourhoods can be used alongside the individual and household-level information collected in the study, making it a useful study for neighbourhood effects analyses. In this paper we explore four Understanding Society data products, based on four different types of rural-urban neighbourhood classifications, to throw light on how much heterogeneity in neighbourhood contexts is captured in the first waves of Understanding Society, including change in neighbourhood contexts.

### Keywords

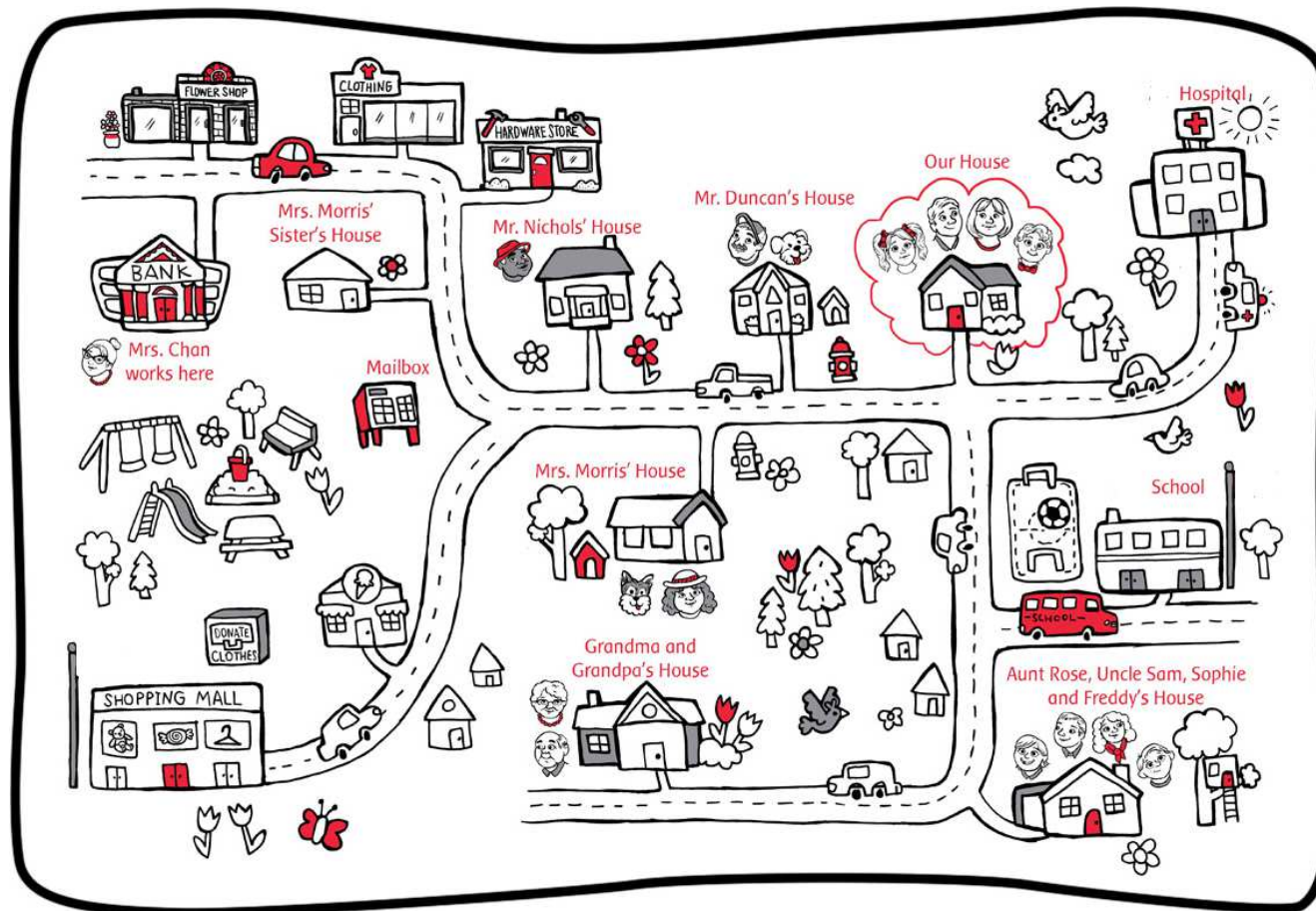
neighbourhoods – quantitative analysis – household panel study – record linkage – panel data analysis – geo-marketing

# In this example you'll ...



- Examine how neighbourhood cohesion varies by individual, household and neighbourhood characteristics relying on data from Wave 1, using FAKED Special Licence LSOA 2001 identifiers to add Townsend Area Deprivation Scores for England
- Apply your knowledge of how to merge data files at different levels and with a different universe of cases
- Get to use some trialled and trusted Stata commands that will make your work a little easier

# Returning to our example neighbourhood



	2001
Buildings	20
Schools	1
Hospitals	1
Shops	5
Bank	1
Houses	12
Leisure facilities	2
# of people	20
adults	15
children	5
pet owners	2



# Merging files

Merged individual-level data file					
a_hidp	a_pno	pidp	a_nbhrsci_dv	a_tenure_dv	a_vicini1
100	1	001	1.1	1	1
100	2	002	2.7	1	2
200	1	003	5	2	0
200	2	004	4.3	2	0

a_lsoa01	
a_hidp	a_lsoa01
10	E0099..
100	E0098..
200	W0097..

Townsend Score	
LSOA01	tscore
E0099..	-0.63
E0098..	3.97

Merged individual-level data file							
a_hidp	a_pno	pidp	a_nbhrsci_dv	a_tenure_dv	a_vicini1	a_lsoa01	a_tscore
100	1	001	1.1	1	1	E0098..	-0.63
100	2	002	2.7	1	2	E0098..	-0.63
200	1	003	5	2	0	W0097..	.
200	2	004	4.3	2	0	W0097..	.

# Worksheet 3: Thinking about individuals, households & neighbourhoods longitudinally: Modelling heterogeneity



- Very simple concept: people/households/neighbourhoods are different!

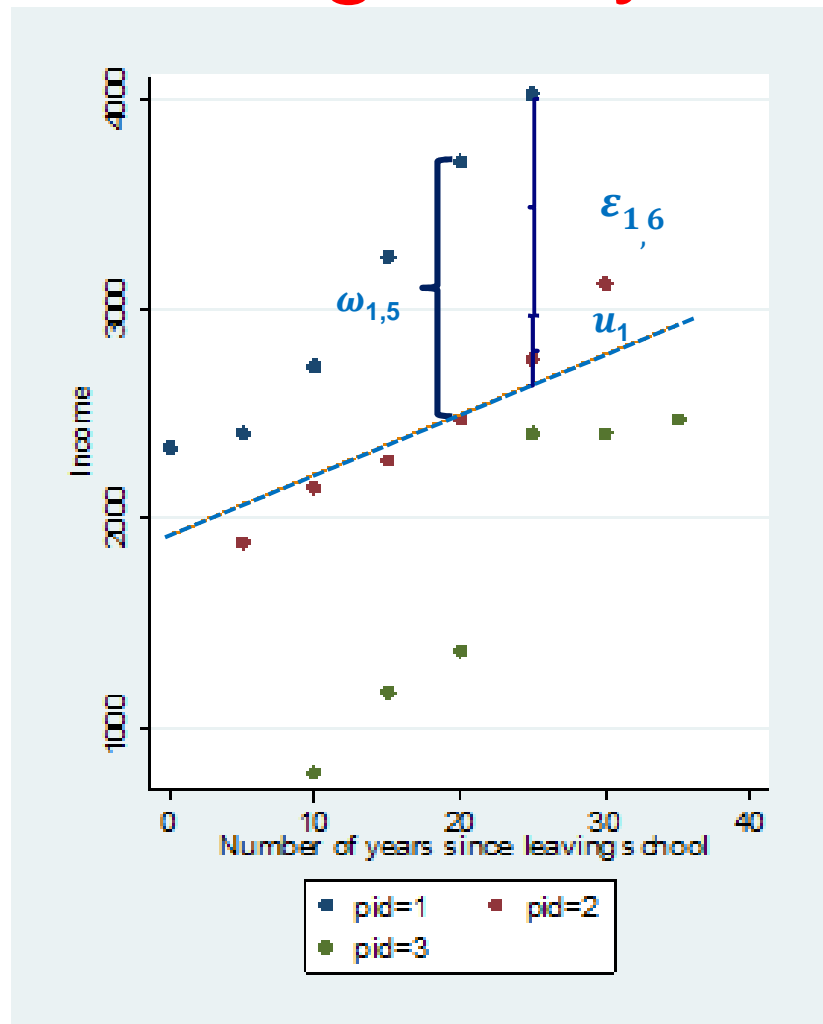
Observed heterogeneity: differences in education levels, or parental background, or anything else that we can measure and control for in regressions

Unobserved heterogeneity: anything which is fundamentally unmeasurable, or which is rather poorly measured, or which does not happen to be measured in the particular data set we are using.

- With panel data we can do something about **unobserved heterogeneity** as we can differentiate between person-level unobserved characteristics that are identical over time and those that vary over time!



# An illustration of unobserved heterogeneity



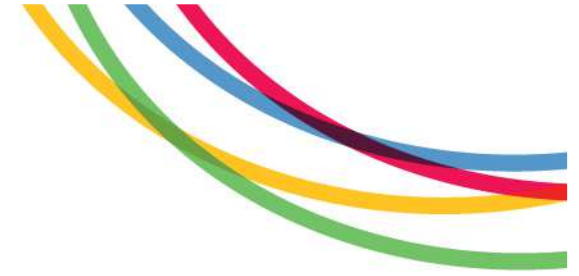
$$y_i = x_i\beta + \omega_i$$

Panel data allows you to break down the error term ( $\omega_i$ ) in two components:

- the unobservable characteristics of the person ( $u_i$ ),
- genuine “error” ( $\epsilon_{it}$ )

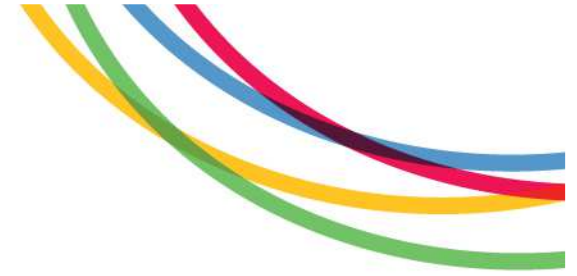
$$y_{it} = x_{it}\beta + u_i + \epsilon_{it}$$

# In this example you will...



- Add data for Wave 3 to the Wave 1 data from the previous Worksheet
- Append the data in a long format - Stata loves it for panel data analysis
- See how a number of different panel estimators are implemented and how they compare
- Get suggestions on how to add some more depth for neighbourhood research

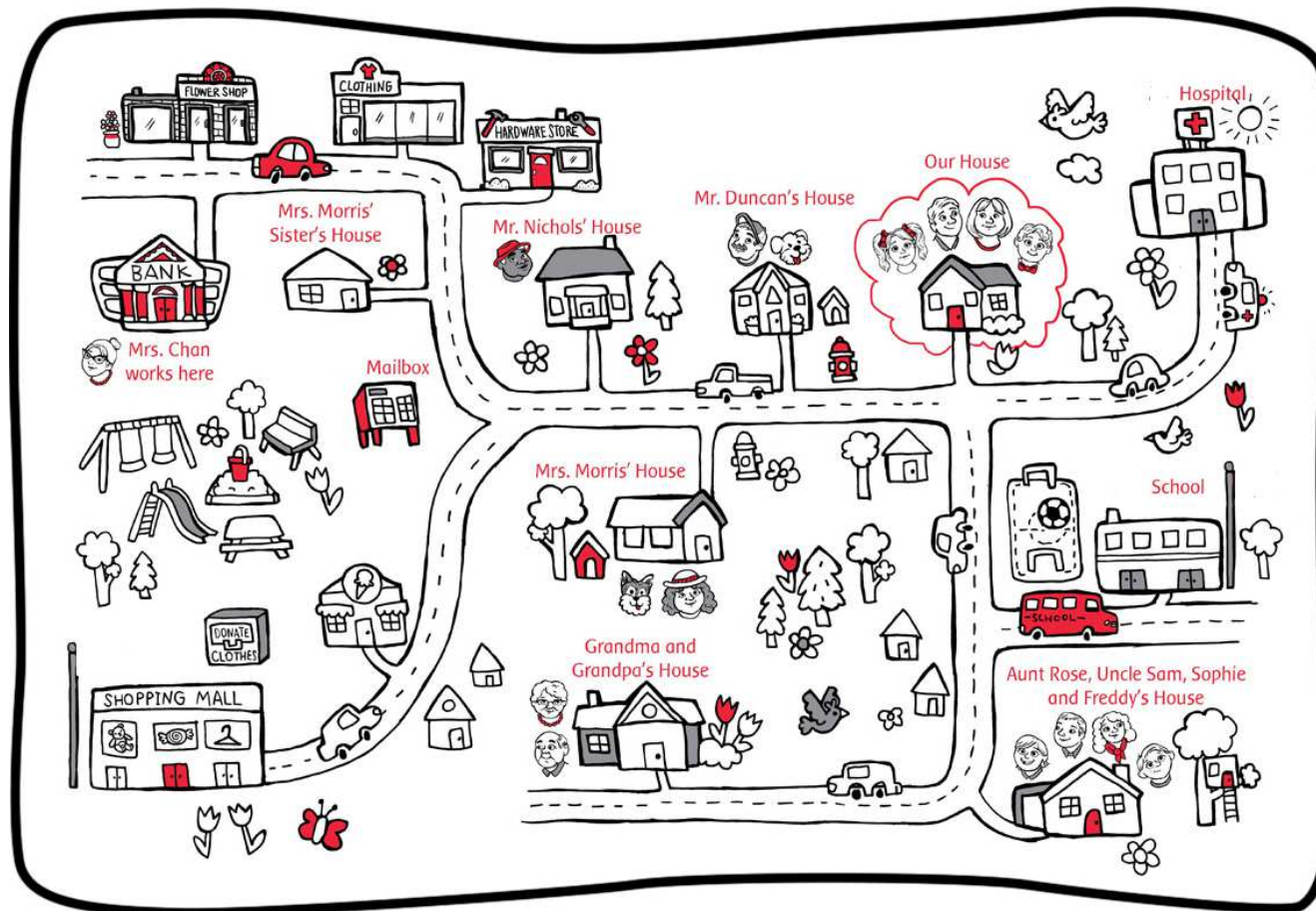
# Change in neighbourhood context



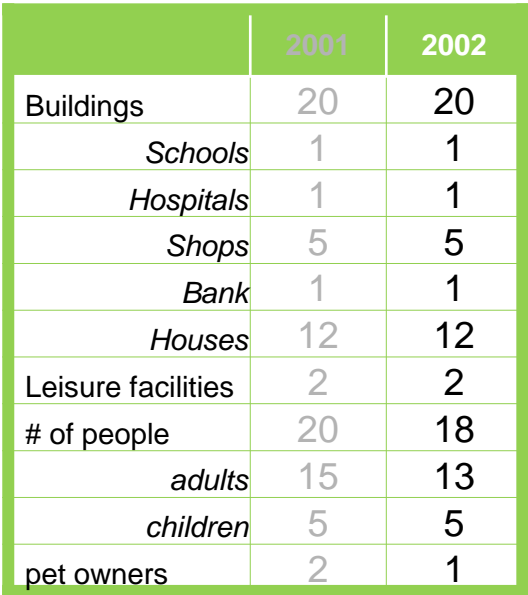
Change over time in the neighbourhood context can occur because:

- Individual moved to a different place (could be within the same micro-area though!)
- Neighbourhood context genuinely changed
- The neighbourhood boundary changed over time

# Individual, household and neighbourhood change in our example neighbourhood

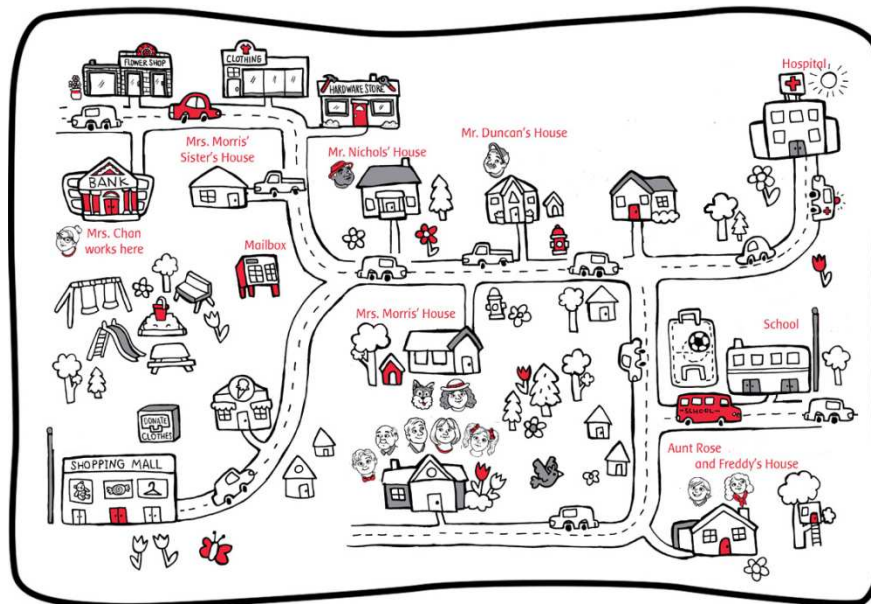


	2001
Buildings	20
Schools	1
Hospitals	1
Shops	5
Bank	1
Houses	12
Leisure facilities	2
# of people	20
adults	15
children	5
pet owners	2



- Grandma has passed away (Mr Duncan's dog, too)
- Uncle Sam's left Aunt Rose, but his daughter Sophie still hangs around to play in the treehouse)
- There are a lot more cars these days (and birds don't like the fumes much)

... after 3 years ...



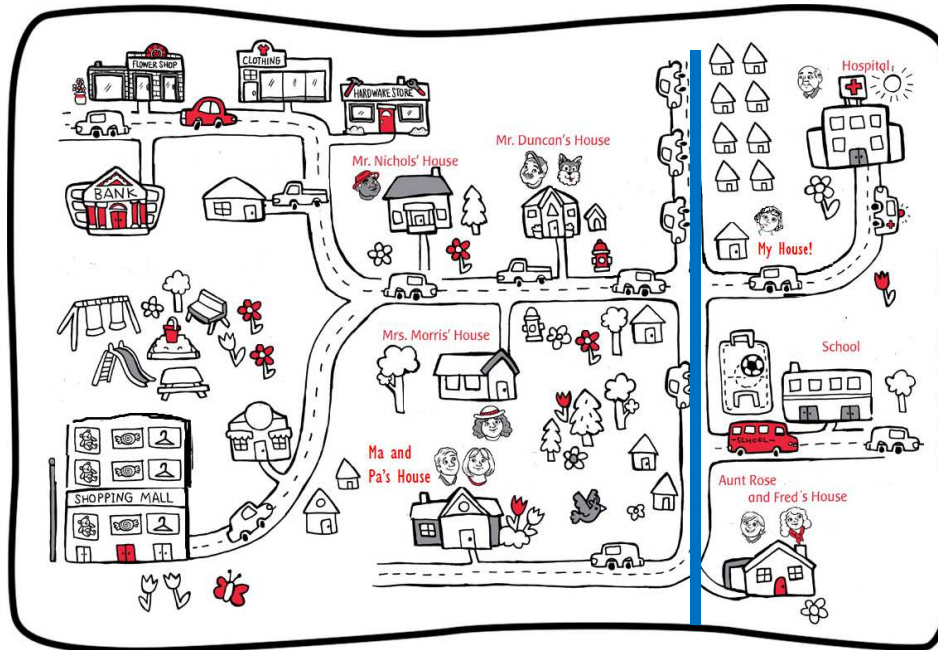
	2001	2002	2004
Buildings	20	20	20
Schools	1	1	1
Hospitals	1	1	1
Shops	5	5	4
Bank	1	1	1
Houses	12	12	12
Leisure facilities	2	2	2
# of people	20	18	18
adults	15	15	16
children	5	4	2
pet owners	2	1	1

### What's happened?

- We've moved in with grandpa as he was getting poorly
- We are now closer to Aunt Rose and Freddy (Sophie is not coming around anymore)
- We sold our house to a developer. The cash will come in handy for my brother's education.
- My brother has turned 18, he wants to go travel the world



# ... and after 10 years

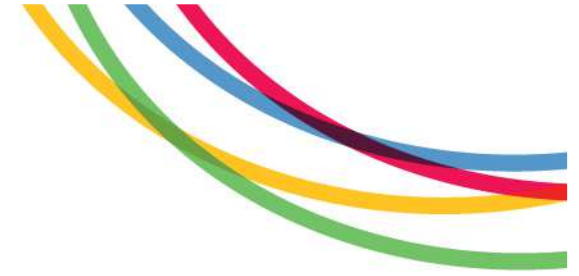


	2001	2002	2004	2011
Buildings	20	20	20	12
Schools	1	1	1	1
Hospitals	1	1	1	1
Shops	5	5	4	0
Bank	1	1	1	0
Houses	12	12	12	10
Leisure facilities	2	2	2	2
# of people	20	18	18	28
adults	15	15	16	19
children	5	4	2	9
pet owners	2	1	1	1

## What's happened?

- I have moved into the new development that they've built close to where our old house was
- The ice cream parlour has closed down, due to competition with the now a lot bigger shopping centre
- My parents live by themselves in my grandparent's house.
- Grandpa is in hospital (obviously I visit him as much as I can – I live close-by)
- My brother emigrated to Australia; he never returned from his gap year
- Freddy does not want to be called Freddy anymore and he's knocked down the treehouse.
- Mrs Morris does no longer have a dog (but Mr Duncan got a new one).
- Mrs Lee retired from working in the bank. Not sure what she does now.

# Important weblinks



Understanding Society Website

<https://www.understandingsociety.ac.uk/>

Data access routes

<https://www.understandingsociety.ac.uk/documentation/access-data>

Online Documentation

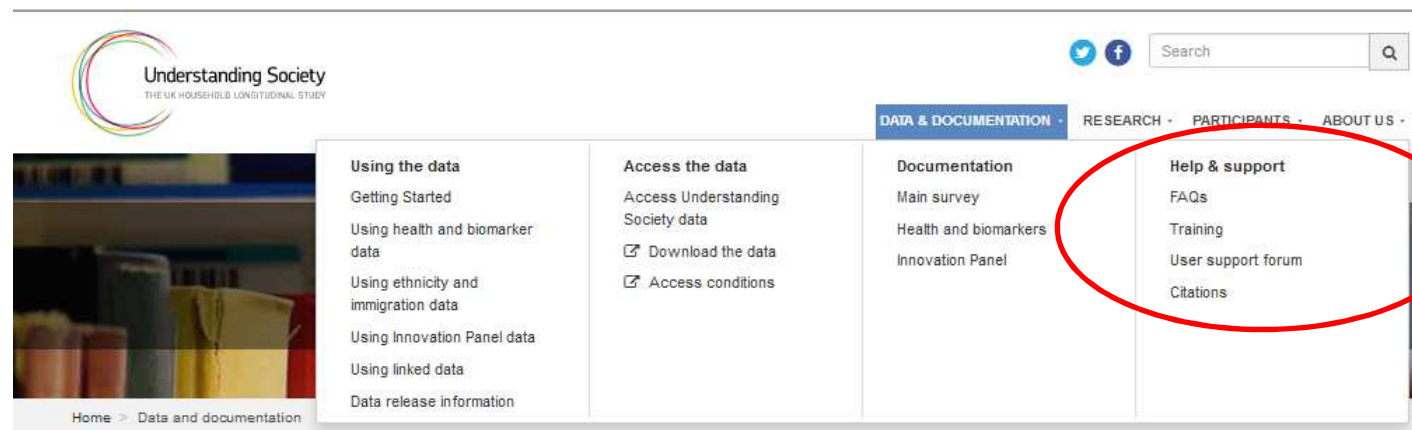
<https://www.understandingsociety.ac.uk/documentation>

Understanding Society on UKDS website

<https://discover.ukdataservice.ac.uk/series/?sn=2000053>



# Help & Support



Understanding Society  
THE UK HOUSEHOLD LONGITUDINAL STUDY

DATA & DOCUMENTATION · RESEARCH · PARTICIPANTS · ABOUT US

- Using the data
  - Getting Started
  - Using health and biomarker data
  - Using ethnicity and immigration data
  - Using Innovation Panel data
  - Using linked data
  - Data release information
- Access the data
  - Access Understanding Society data
  - Download the data
  - Access conditions
- Documentation
  - Main survey
  - Health and biomarkers
  - Innovation Panel
- Help & support**
  - FAQs
  - Training
  - User support forum
  - Citations

Home > Data and documentation


Data collected from the survey's thousands of participants is securely stored by the [UK Data Service](#), from where researchers can access it online. All of the associated documentation is available here.

Main survey	Health and biomarkers	Innovation panel
OVERVIEW	OVERVIEW	OVERVIEW
USER GUIDE	USER GUIDE	USER GUIDE
DATASET DOCUMENTATION	DATASET DOCUMENTATION	DATASET DOCUMENTATION
QUESTIONNAIRES	QUESTIONNAIRES	QUESTIONNAIRES
TECHNICAL REPORTS	FIELDWORK DOCUMENTS	TECHNICAL REPORTS
FIELDWORK DOCUMENTS		FIELDWORK DOCUMENTS
LONG TERM CONTENT PLAN		
QUALITY PROFILE		
<a href="#">SEE MORE</a>	<a href="#">SEE MORE</a>	<a href="#">SEE MORE</a>

Help and support	Citation	Data releases
<a href="#">Information for new users</a> <a href="#">How to access the data</a> <a href="#">Training courses</a> <a href="#">User Support Forum</a>	Bibliographic citations and acknowledgements.	Information about latest releases of data and updates.
<a href="#">SEE MORE</a>	<a href="#">SEE MORE</a>	<a href="#">SEE MORE</a>

# User Forum



Sign in Register

Search:

Home Projects Help

## Understanding Society User Support

Overview Activity Roadmap Issues News Wiki

Sign in Register

Search:

### Overview



We offer a range of user support. You can search through our [FAQ](#), [Issues](#) raised by other users in the past or raise a [New issue](#). For new data users, we offer [training courses](#) in the handling and analysis of longitudinal data.

**We aim to respond to all queries within 10 working days.**

The FAQ are currently being comprehensively overhauled and updated and we will communicate the updated FAQ section once it is complete

- [How to raise an issue](#)
- [FAQ](#)
- [Current & closed issues](#)
- [Report a new issue](#)
- [Training courses](#)

We are keen to hear about any data issues and experiences that you have as this will help us build the best possible knowledge database for the UKHLS and BHPS data sets.

#### Issue tracking

- Support: 17 open / 675

[View all issues](#)

#### Members

Manager: Alita Nandi, Gundi Knies, Jon Nears, Olena Kaminska, Victoria Nolan

#### Latest news

##### Waves 1-5 Quality Profile

*Waves 1-5 Quality Profile now available online*

Added by Victoria Nolan about 1 year ago

##### New "How To" Guide for using the Support Forum

*How to get started with raising an issue*

Added by Victoria Nolan about 1 year ago

##### Intro to Understanding Society using Stata 17-18 Nov 2016

*Free training course*

Added by Victoria Nolan about 1 year ago

##### Intro to Understanding Society using Stata 14-15 April 2016

*Free training course*

Added by [Victoria Nolan](#) about 1 year ago

##### Wave 5 is out!

Added by Redmine Admin over 1 year ago

[View all news](#)