Commentary on Multi-purpose Social Research Survey 1975 (UKDS SN 680)

Copyright © 2014 John F Hall　　　　　　　　　　[Draft only: last updated 24 Sep 2014]

The first ever **Multi-purpose Social Research Survey** in the UK was conducted by the SSRC Survey Unit in 1975. It was intended as a bi-annual survey conducted to rigorous professional and statistical standards, and was a vehicle for questions submitted by academics for which a fee was payable. Included in the fee was a raft of demographic data.   (see Appendix for facsimile information pamphlet.) In the event only one wave was carried out as the SSRC controversially decided to close the Survey Unit. This decision most likely deterred potential clients, but two short sections were commissioned, one on off-course betting (four questions) the other on social participation (four questions). The bulk of the questionnaire replicated questions from the SSRC Quality of Life survey, with a major and expanded section on employment (38 questions) and demographic information (15 questions).

The data were deposited by the late Cathie Marsh after she moved to SPS at Cambridge, but who was my research trainee 1974 -76 and who worked on the survey with the late John Utting. The survey is archived at UKDS as SSRC Survey Unit Multi-Purpose Survey, 1975 and the distributed SPSS file **sn680.sav** contains 249 variables and 1450 cases. The file follows conventions developed for all SSRC/SU surveys (mainly **positional** variable naming and missing values 7, 8, 9 or 97, 98, 99) but is constrained by the mainframe version of SPSS then in use (Variable and value labels exclusively in UPPER CASE, 8-character limit on names, 40-character limit on variable labels, 20-character limit on value labels). Some variable labels are in Mixed Case, but these are modifications made during later analysis and/or teaching. All variable names are in lower case, but this derives from the archiving system at Essex. There do not appear to be any associated SPSS setup files. To date I have been unable to discover who commissioned the questions on off-course betting or on social participation, or to trace any publications arising from the survey as a whole.

You can download the documentation direct from UKDS (UKDA_Study_680_Information.htm and 680userguide.pdf) but to access the actual data you need to be a registered user. Once you are registered UKDS will send you a zip file from which you can extract what you need.
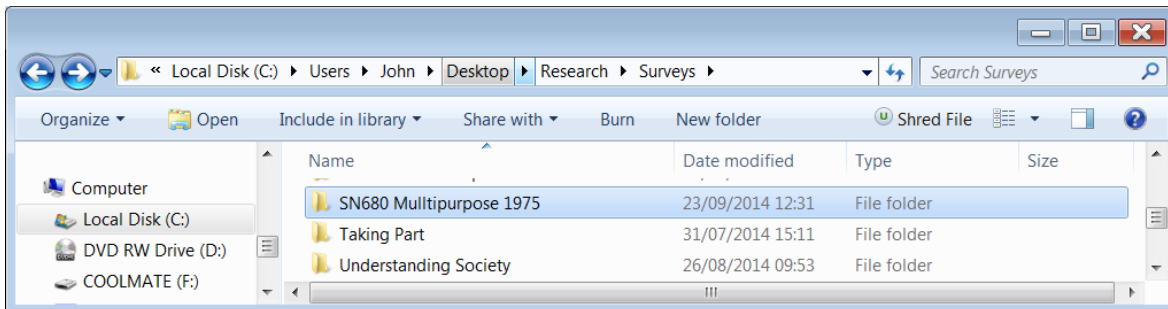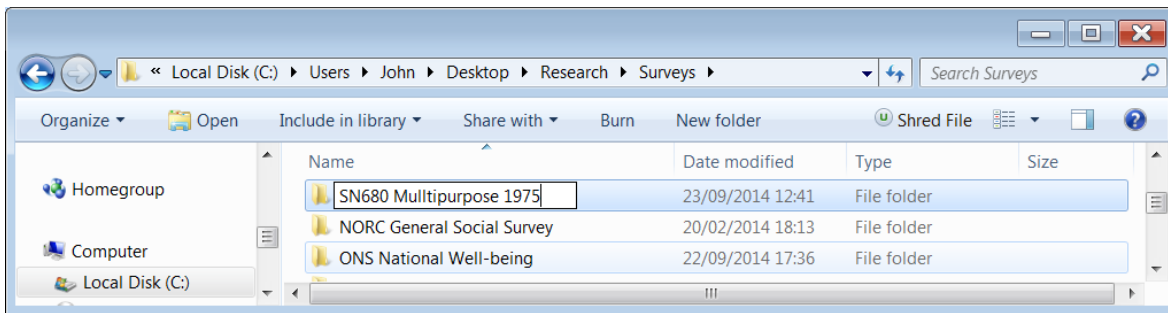
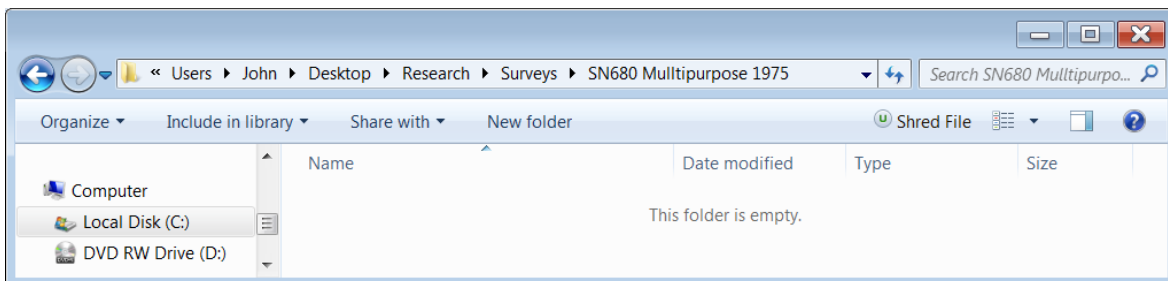Decide on a location for the survey (in my case Research >> Surveys:)

Open a New folder:

1
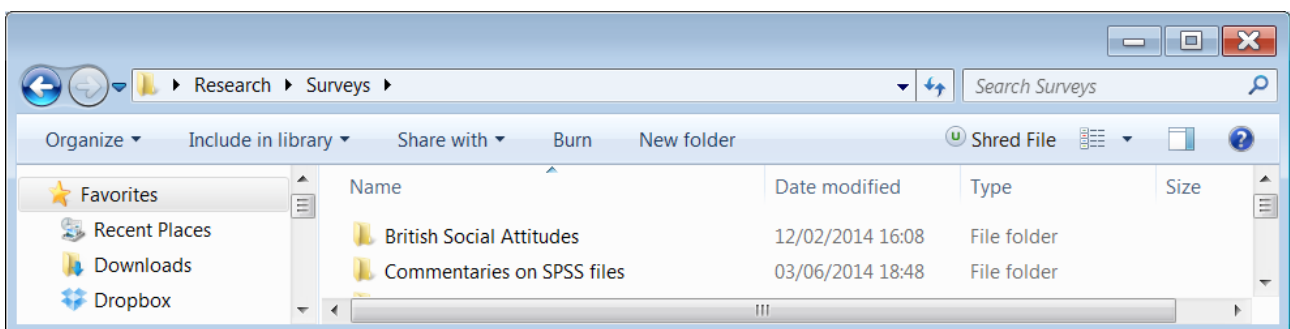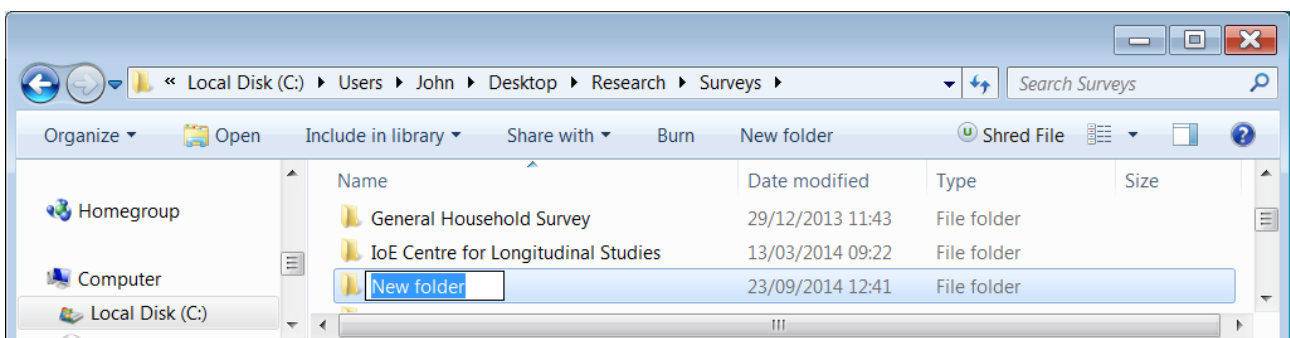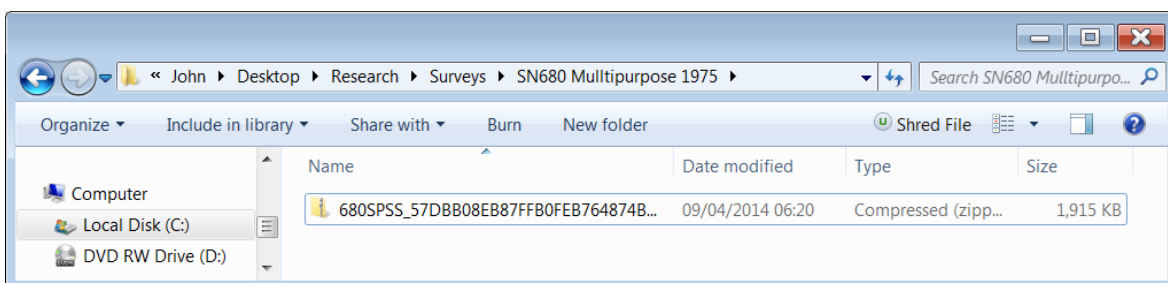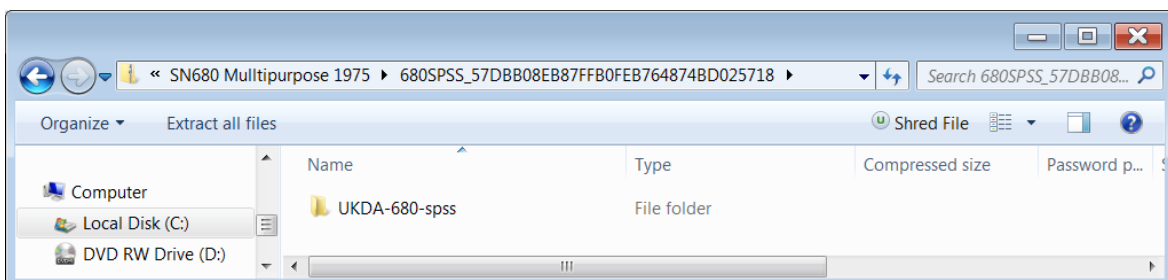
and give it a name, in my case SN680 Multipurpose 1975
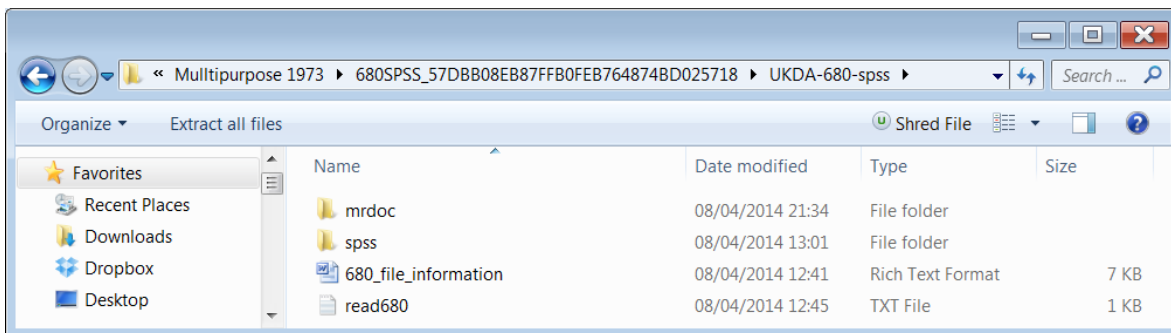

Open folder 📁 SN680 Mulltipurpose 1975

and download zip file 680SPSS_57DBB08EB87FFB0FEB764874BD025718 to the new folder:

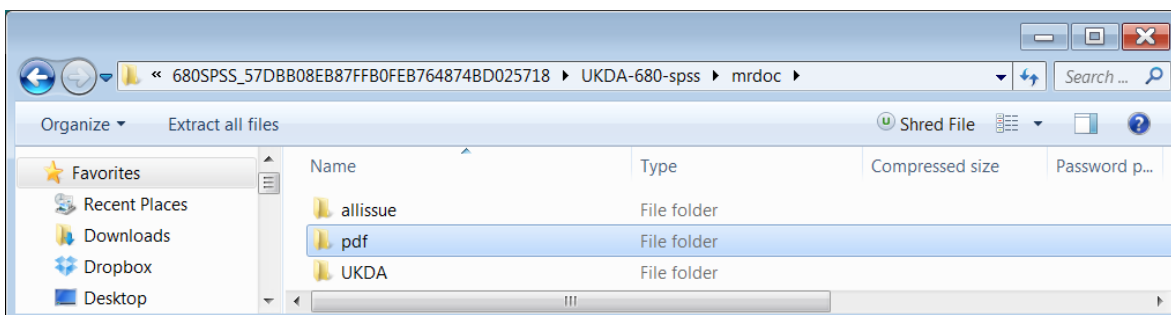Open the zip file: 📁 680SPSS_57DBB08EB87FFB0FEB764874BD025718

Open folder **UKDA-680-spss**

Name	Date modified	Type	Size
mrdoc	08/04/2014 21:34	File folder	
spss	08/04/2014 13:01	File folder	
680_file_information	08/04/2014 12:41	Rich Text Format	7 KB
read680	08/04/2014 12:45	TXT File	1 KB

Open folder **mrdoc** :

Name	Type	Compressed size	Password p...
allissue	File folder		
pdf	File folder		
UKDA	File folder		

Open folder **pdf** :

Name	Type	Compressed size	Password p...
680userguide	Adobe Acrobat Document	1,798 KB	No

Open **680userguide**

The **User Guide** is in searchable pdf format.

Bookmarks

- User Guide
  - Report on Sampling and Fieldwork
  - Questionnaire
  - Appendix - Recoding
  - Codebook
  - Notes
  - Questionnaire Instructions

3
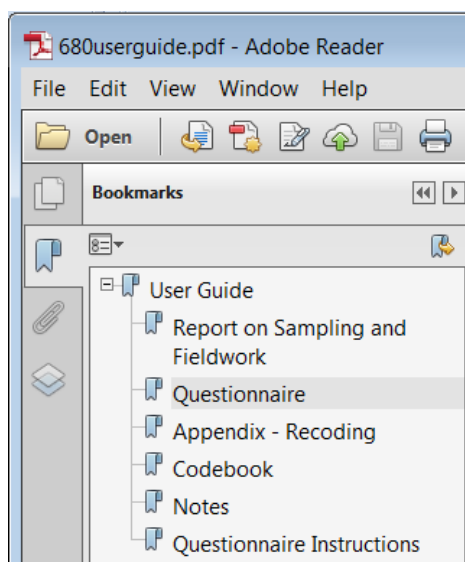
It's worth having a look at the questionnaire (pages 8-27) to get a feel for the content.

File >> Save as:

Navigate to folder 📁 SN680 Mulltipurpose 1975
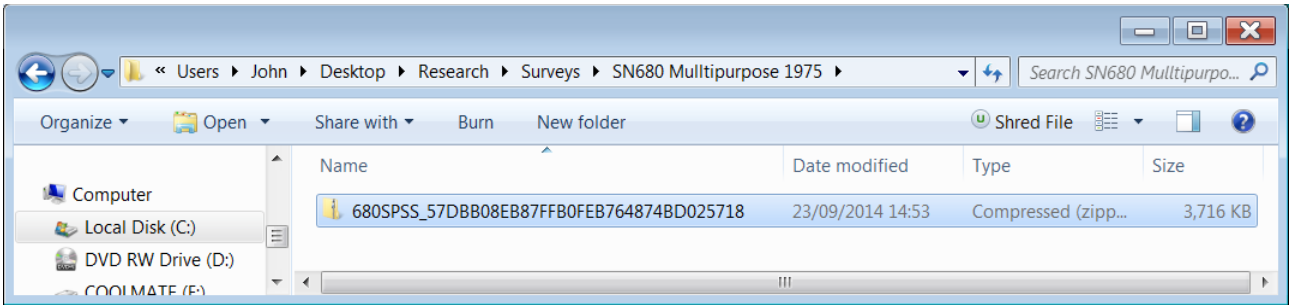
and save a copy of the user guide there. You will need access to it during data analysis.
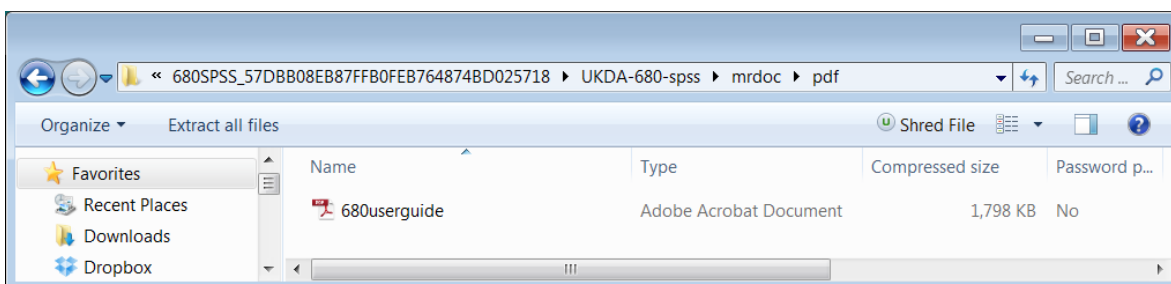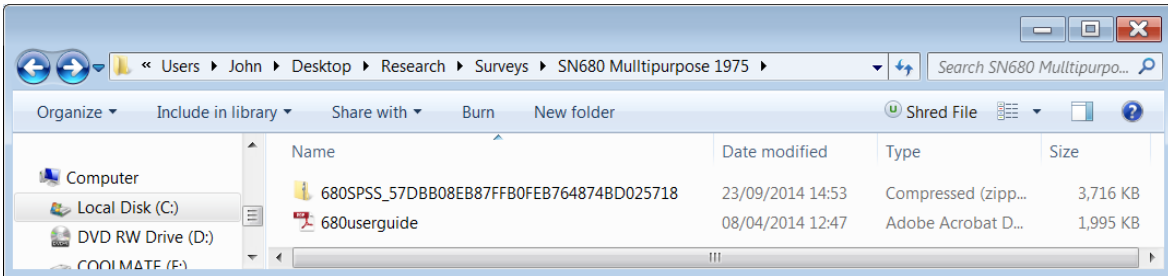

Open the zip file: 680SPSS_57DBB08EB87FFB0FEB764874BD025718
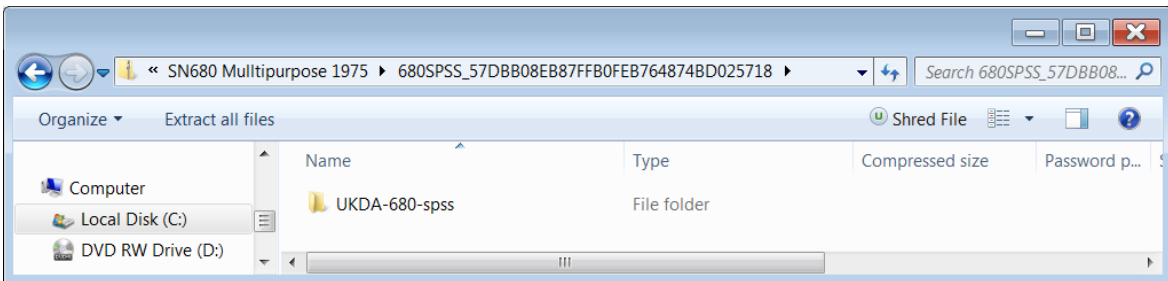
Open folder   again
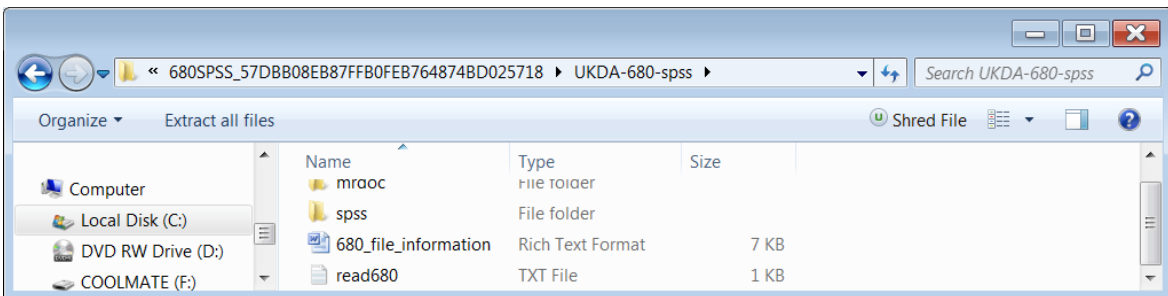
Open folder
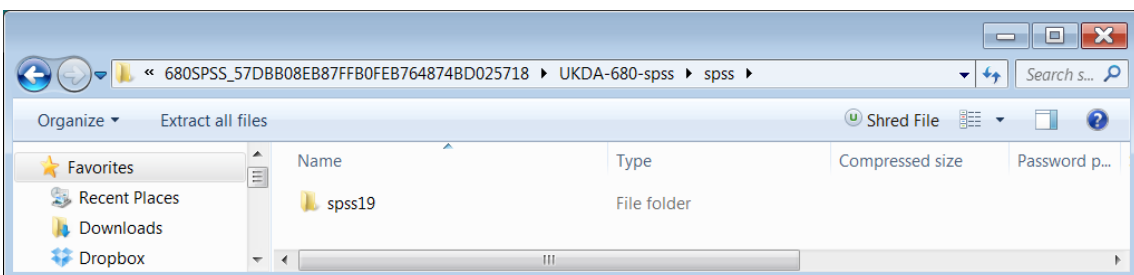
then open folder 📁 spss19
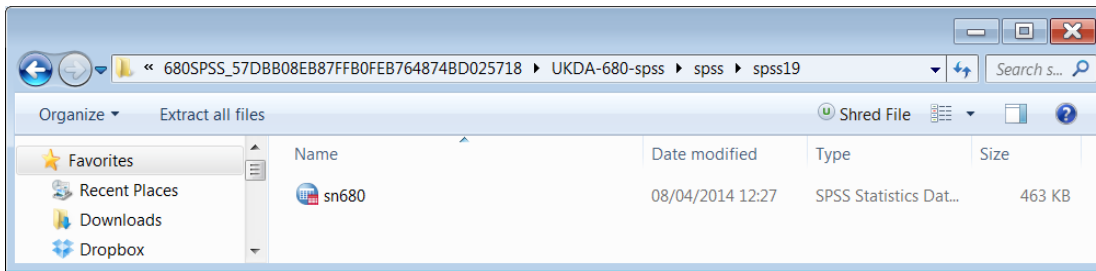
To perform the following steps you will need access to a licenced copy of SPSS, but you can still easily follow the process and logic by reading on.

Double-click on file 🖼 sn680
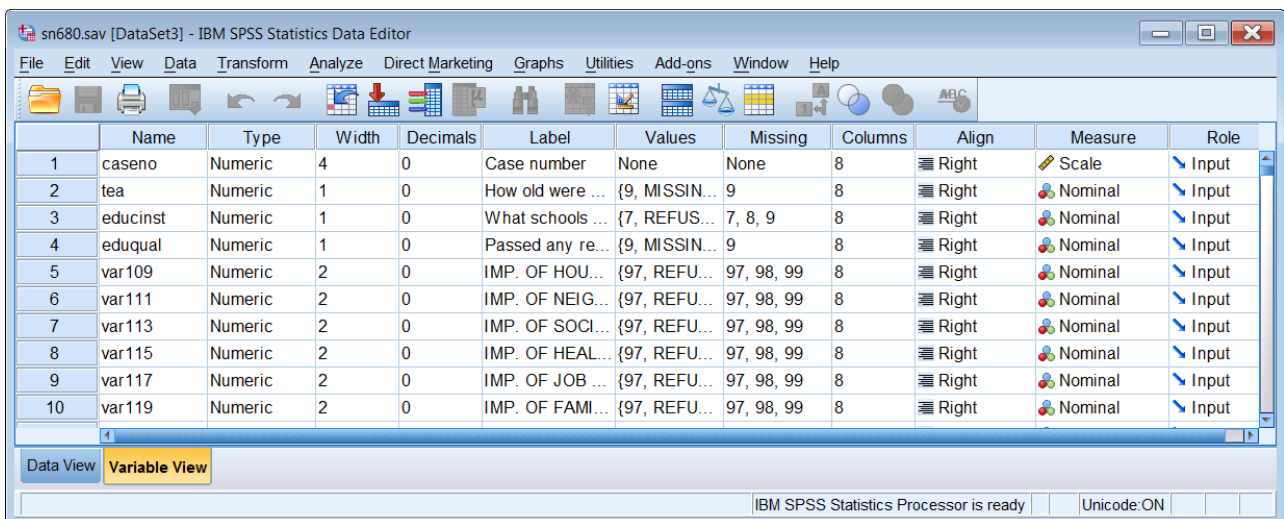
Data Editor sn680.sav in Variable View

Never work on an original file. **Make a copy** of the file:

File >> Save As
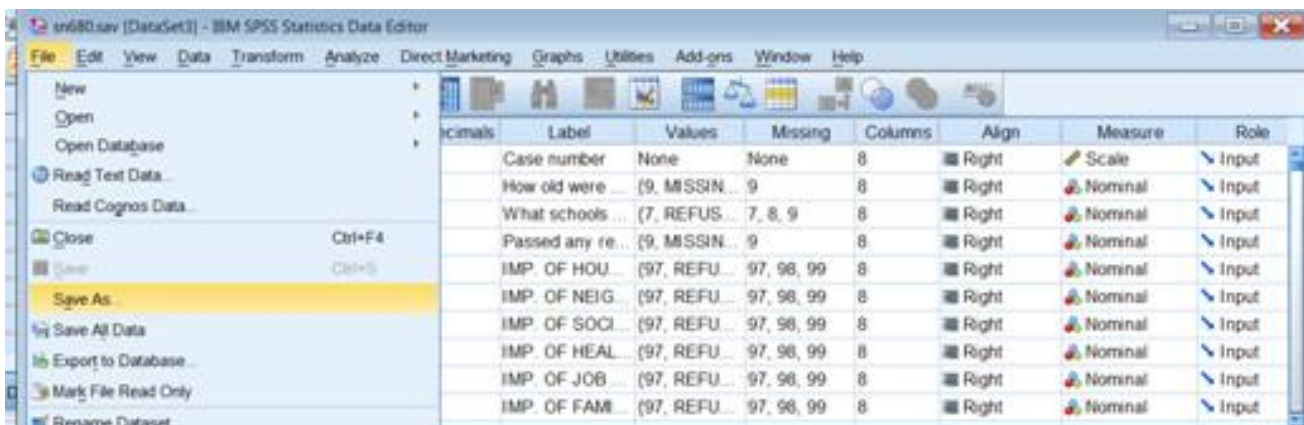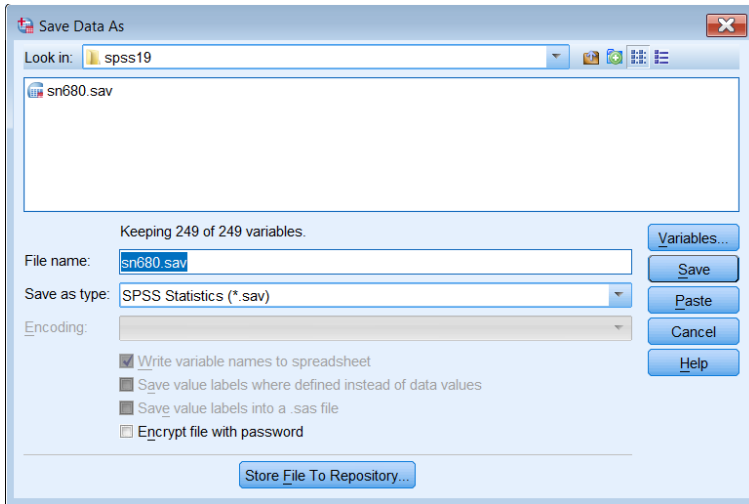
In the Look in box at top left, navigate from back to
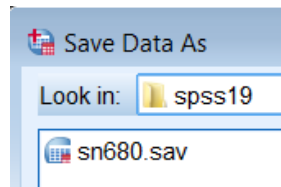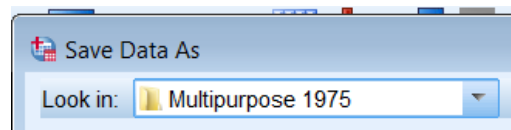
Change sn680 to a new file name:
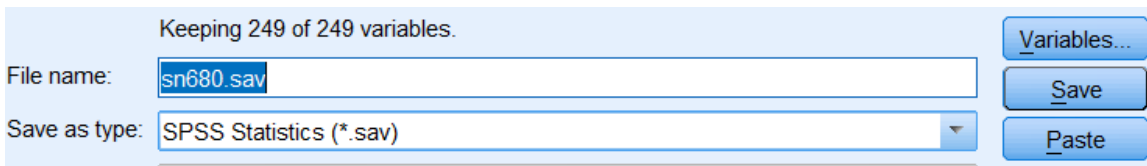
. . in my case the root sn680 plus an underscore, my initials and an edition number sn80_jfh1.sav
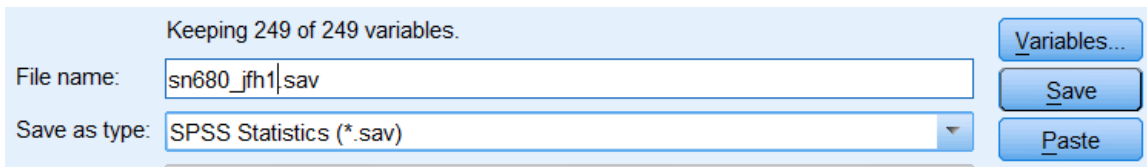
Click on [Save] to save the file with a new name to your new folder:
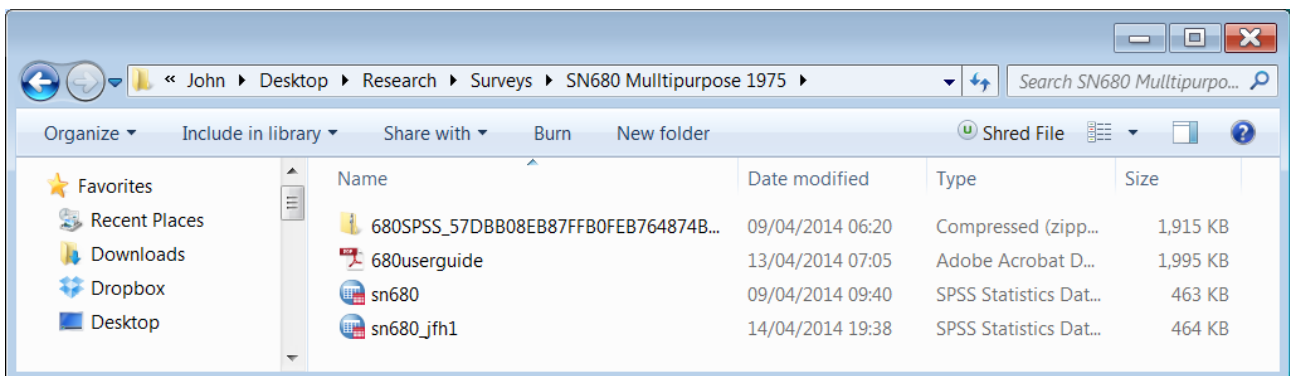
The original file remains untouched in the zip folder, your copy of the original remains untouched in folder Multipurpose 1975 and your personal copy of the file has now been saved with a new name in the same folder: SN680 Mulltipurpose 1975

Your personal copy immediately becomes the active file. This is good practice, as you should never work on an original file in case you make irretrievable errors and have to start again from scratch.

Open file 📊 sn680_jfh1   (or whatever you called your version)
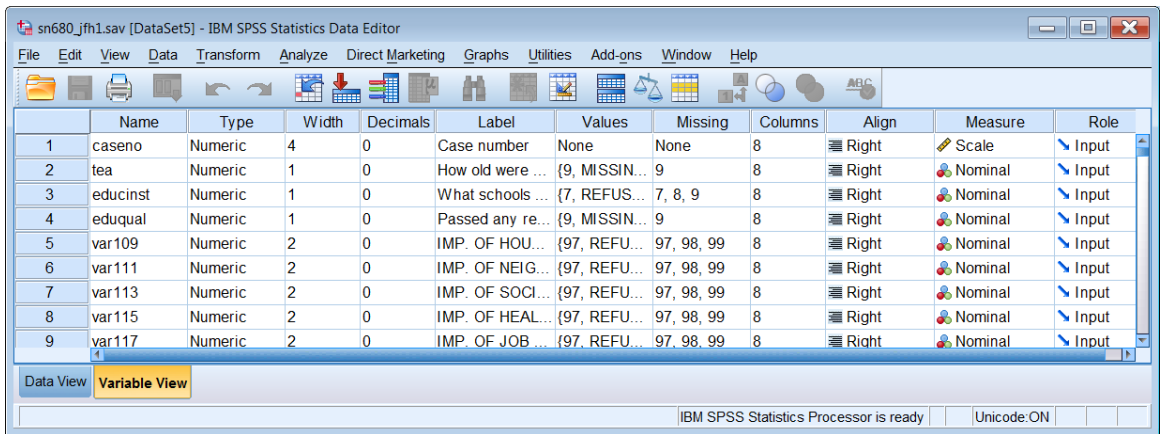
Data Editor sn680_jfh1 in Variable View

I don't like working with the SPSS default column order for variable attributes, so I always rearrange them to put the more important (to me) attributes further to the left.

View >> Customize Variable View
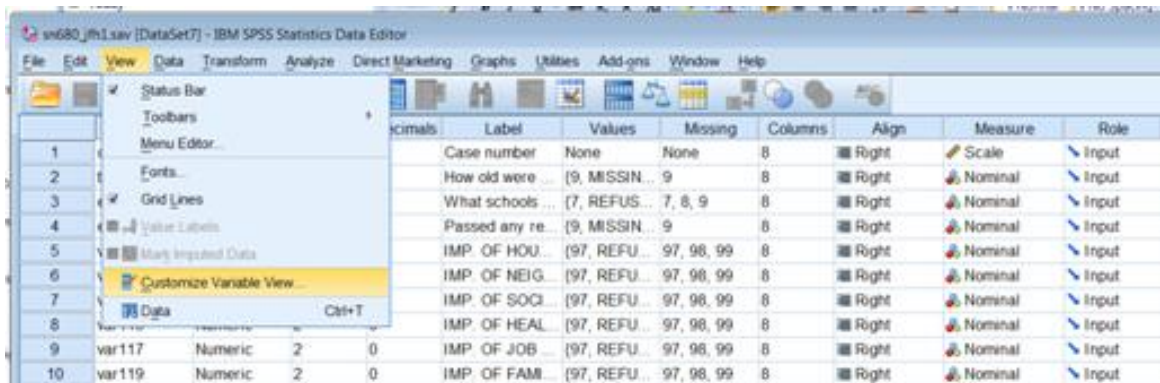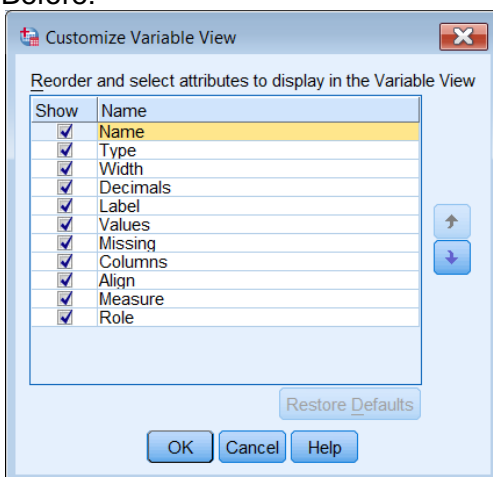
Before: After:
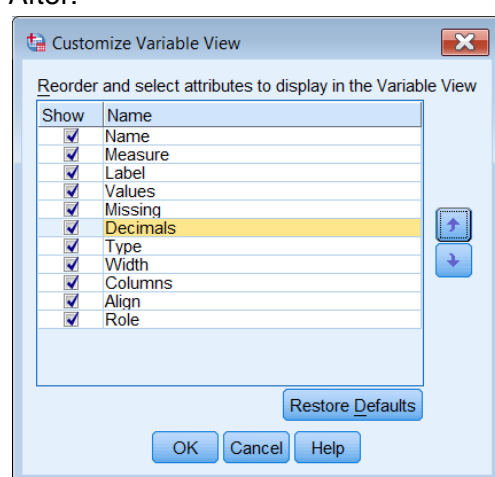
Use the blue arrows 🔽 and 🔼 to move the more important attributes to a higher priority:

Press   OK

8

File **sn680_jfh1** in Variable View with attributes reordered.

Widen the **Labels** and **Values** columns to see more information:

File **sn680_jfh1.sav** in Variable View with **Label** and **Values** columns widened

Scroll down the **Labels** and **Values** columns and widen them to accommodate the longest labels:

This display format will be retained whilst the file is open during your SPSS session, but will revert to narrow columns if the file is closed and re-opened. You don't really need **Decimals**, **Type**, **Width**, **Columns** or **Role** for now, so drag the right hand edge in to reveal only:

Name   **Measure**   **Label**   **Values**   **Missing**

9

sn680_jfh1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Measure	Label	Values	Missing
1	caseno	Scale	Case number	None	None
2	tea	Nominal	How old were you when you finished your full-time education?	{9, MISSING,DK,REF}...	9
3	educinst	Nominal	What schools and other places of f-t education have you attended since age 11?	{7, REFUSED}...	7, 8, 9
4	eduqual	Nominal	Passed any recognised examinations or completed an apprenticeship?	{9, MISSING,DK,REF}...	9
5	var109	Nominal	IMP. OF HOUSING CONDITIONS	{97, REFUSED}...	97, 98, 99
6	var111	Nominal	IMP. OF NEIGHBOURHOOD	{97, REFUSED}...	97, 98, 99
7	var113	Nominal	IMP. OF SOCIAL STATUS	{97, REFUSED}...	97, 98, 99
8	var115	Nominal	IMP. OF HEALTH	{97, REFUSED}...	97, 98, 99
9	var117	Nominal	IMP. OF JOB SATISFACTION	{97, REFUSED}...	97, 98, 99
10	var119	Nominal	IMP. OF FAMILY LIFE	{97, REFUSED}...	97, 98, 99

Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

It's important to **save your work** regularly. **Save the file** with `Ctrl`+S or File >> Save

The working file *sn680_jfh1.sav (* denotes a working file)is saved as sn680_jfh1.sav [DataSet1]

sn680_jfh1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

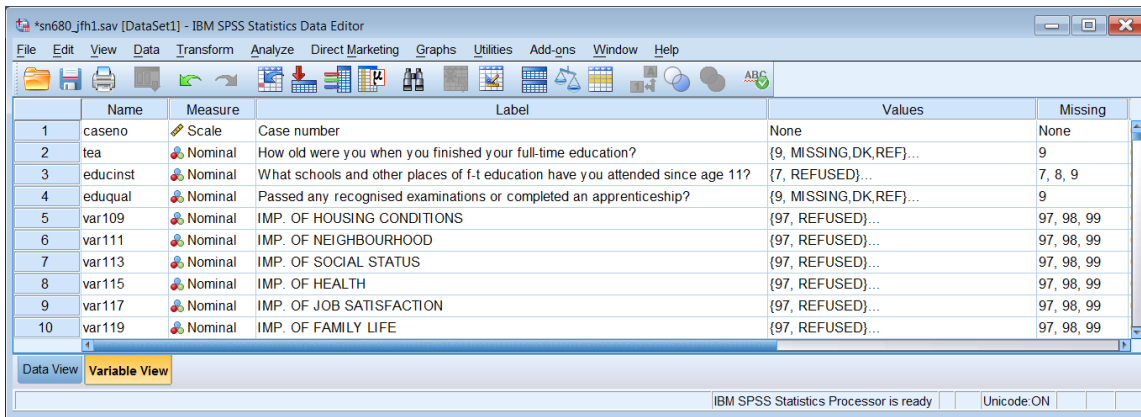|	Name	Measure	Label	Values	Missing
1	caseno	Scale	Case number	None	None
2	tea	Nominal	How old were you when you finished your full-time education?	{9, MISSING,DK,REF}...	9
3	educinst	Nominal	What schools and other places of f-t education have you attended since age 11?	{7, REFUSED}...	7, 8, 9
4	eduqual	Nominal	Passed any recognised examinations or completed an apprenticeship?	{9, MISSING,DK,REF}...	9
5	var109	Nominal	IMP. OF HOUSING CONDITIONS	{97, REFUSED}...	97, 98, 99
6	var111	Nominal	IMP. OF NEIGHBOURHOOD	{97, REFUSED}...	97, 98, 99
7	var113	Nominal	IMP. OF SOCIAL STATUS	{97, REFUSED}...	97, 98, 99
8	var115	Nominal	IMP. OF HEALTH	{97, REFUSED}...	97, 98, 99
9	var117	Nominal	IMP. OF JOB SATISFACTION	{97, REFUSED}...	97, 98, 99
10	var119	Nominal	IMP. OF FAMILY LIFE	{97, REFUSED}...	97, 98, 99

Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

Each row of the data represents a variable. If you scroll to the bottom of the file you will see that there are 249 variables:

sn680_jfh1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Measure	Label	Values	Missing
245	var647	Nominal	LEAVE JOB - JOB CONTENT	None	7, 8, 9
246	var648	Nominal	LEAVE JOB - WORKED TOO HARD	None	7, 8, 9
247	var649	Nominal	LEAVE JOB - WAY MANAGEMENT TREAT EMPLYEES	None	7, 8, 9
248	var650	Nominal	LEAVE JOB - STATUS	None	7, 8, 9
249	var651	Nominal	LEAVE JOB - FAIRNESS	None	7, 8, 9
250					

Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

Switch to Data View

sn680_jfh1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 249 of 249 Variables

	caseno	tea	educinst	eduqual	var109	var111	var113	var115	var117	var119	var121	var123	var12!
1	1	7	2	5	8	7	4	10	6	9	7	8	
2	2	1	1	0	10	8	9	10	9	10	9	10	
3	3	4	1	4	8	9	7	9	9	10	7	8	
4	4	0	1	0	9	9	7	9	9	10	8	10	
5	5	0	1	3	10	10	8	10	10	8	6	7	

Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

Each row of the data represents a case. If you scroll to the bottom of the file you will see that there are 1450 cases:

Each of these 1450 cases has data for all 249 variables, so there are 361050 cells in the complete data matrix (spreadsheet).

Go back to Variable View
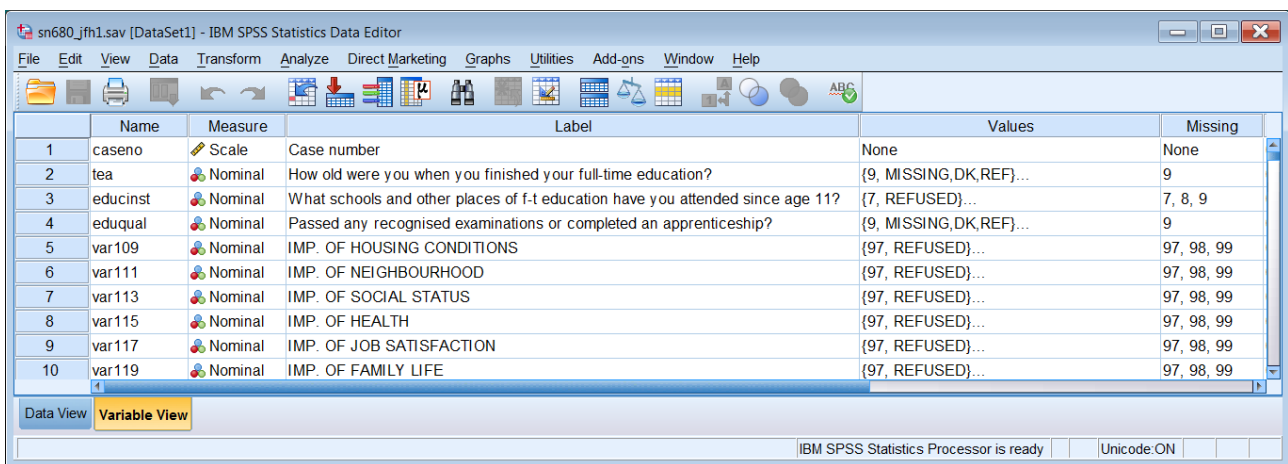
Variable Names

All the variable names are in lower case. Whilst there are a few mnemonic variable names (eg **caseno tea educinst eduqual**) most names start with **var** followed by three digits **ddd** (eg **var109 var111** . . . **var650 var651**) Clearly, these names do not tally with the row numbers (see right); they are positional. The original data were punched on 80-column Hollerith cards, six cards per case. In those days all good questionnaires were already marked up for data-preparation by having card and column numbers clearly indicated in the margins. The data for Q.1 (Terminal Education Age: see questionnaire extract above) were punched in record 1 column 6 and would originally have been called **var106**.

It was required SSRC/SU practice that names start with **var** followed by three digits **ddd**, where the first digit indicates the record number and the last two digits indicate the start column in that record. This meant that researchers looking at **varxxx** could easily find the relevant question in the questionnaire and vice-versa. Thus a researcher interested in var224:

. . could immediately look through the questionnaire to find the question used (Q.40c on p11 of the questionnaire).

```
(STILL USING CARD 3)
c.  All things considered, how satisfied
    or dissatisfied are you overall with          (RECORD NO.)
                            your job?_ _ _ _ _        _____  (24)(25)
```

Conversely, a researcher interested in Q.12 about betting on horse races:

No.	Question	Answer	Code	Skip to
	ASK ALL RESPONDENTS, INCLUDING THOSE SKIPPED FROM Q.5/6, ALL REMAINING QUESTIONS IN SURVEY		(6)	
'2.	Many people, at some time in their lives, bet money on things like cards, football pools, raffle tickets or horse-racing. Do you ever back horses either at the race course or without going to the races?	Yes, on course only	1	
		Yes, off course only	2	43
		Yes, both	3	
		No, neither	4	45

(77)-(80) (1)-(4) (5)
7502   Repeat   3

- 13 -

. . . can immediately find the relevant data by looking at the data-prep instructions which indicate record 3 column 6:

(77)-(80) (1)-(4) (5)
7502   Repeat   3

Code   **Code**

(6)   (6)

. . . and looking in the SPSS file for a variable called **var306**.

var306	Nominal	DO YOU EVER BET

For a fuller explanation of this convention, see 1.3.1 Conventions for Naming Variables in SPSS

[**NB**: This particular questionnaire does not have card numbers indicated afresh on each page, but fieldwork agencies (in this case Social Policy Research) all had their own procedures for questionnaire layout and data-prep.]

Some questions for which variables have mnemonic names can be complex to find. One solution would be to add question numbers to the beginning of the variable labels. For instance:

167	occstath	Nominal	HOH in paid employment?

can easily be amended to:

167	occstath	Nominal	Q.49: HOH in paid employment?

. . but naming the variable **var425** would also be acceptable. At least the variables in this file are all in questionnaire order: this is not always the case.
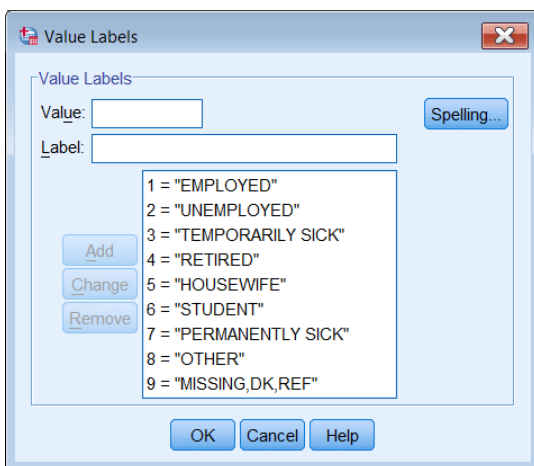
Labels

Most of the variable labels and all the value labels are in UPPER CASE.

occstat	Nominal	Are you at present working for pay?	{1, EMPLOYED}...
oug	Scale	What job do you do?	{997, REFUSED}...
classr	Nominal	Registrar General's social class classification	{7, REFUSED}...
empstat	Nominal	Registrar General's employment status classification	{8, INAPPLICABLE}..
industry	Nominal	What does the firm/organisation you work for actually do?	{97, REFUSED}...
var137	Nominal	SELF-EMPLOYED OR NOT	{7, REFUSED}...
var138	Nominal	NO. OF PEOPLE ONE IS IN CONTROL OF AT WORK	{8, INAPPLICABLE}..
var139	Nominal	SHORT-TIME	{7, REFUSED}...

Full value labels are present for only a few classification and household grid variables. If you scroll down the **Values** column, you will see that most values displayed start at 7, 97 or 997 (Refused). Value labels have to be added manually using the **ADD VALUE LABELS** command in syntax or by using Data >> Define Variable Properties.

One variable with value labels already saved is **occstat** (row 15).

The only other variables with a complete set of value labels are **tenure** (row 116) **house** (117) the sex of each household occupant **var329 var332** etc (119 ff) **occstath** (167) **sex** (181) **area** (183)

Missing values are present, correct and consistent for all variables, 7, 8 or 9 for single digit values, 97, 98, 99 for 2-digit values or 997, 998 and 999 for 3-digit.:

{7, REFUSED}...
{97, REFUSED}...
{997, REFUSED}...

Measurement Levels

Measurement levels originally specified, if any, have been lost during processing at UKDS (apparently normal as they produce Stata and *.tab versions as well, plus documentation). The ones in this file are those automatically assigned by SPSS as Nominal or Scale, depending on the number of discrete values encountered.

If you scroll up and down the **Level** column you will see there are no variables specified as Ordinal. The measurement levels therefore need to be checked and corrected. This can be done manually directly in the **Data Editor**, but that leaves no audit trail. It can also be done using Data >> Define Variable Properties. This is probably more accurate for beginners, but is time consuming and produces separate sets of syntax commands for each variable edited, which, if you use **PASTE**, leads to enormously long and repetitive syntax. The latter can always be edited down, but my own preference is to use the **VARIABLE LEVEL** command in syntax.

Proceed to check data (in topics, to keep output small and manageable). Start with the classification data (demographics):

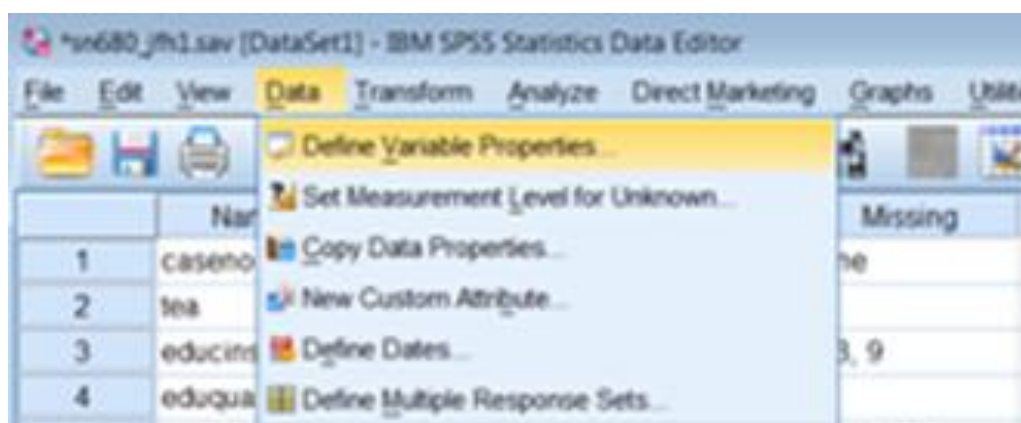| 2	tea	Nominal	How old were you when you finished your full-time education?
3	educinst	Nominal	What schools and other places of f-t education have you attended since age 11?
4	eduqual	Nominal	Passed any recognised examinations or completed an apprenticeship?

15	occstat	Nominal	Are you at present working for pay?
16	oug	Scale	What job do you do?
17	classr	Nominal	Registrar General's social class classification
18	empstat	Nominal	Registrar General's employment status classification
19	industry	Nominal	What does the firm/organisation you work for actually do?

165	numberhh	Nominal	NUMBER OF PEOPLE IN HOUSEHOLD
166	personno	Nominal	PERSON NUMBER OF THE RESPONDENT INTERVIEWED
167	occstath	Nominal	HOH in paid employment?
168	classh	Nominal	What job does HOH do?
169	religion	Nominal	When you were a child, what was the religious faith of your parents?
170	race	Nominal	RACE OF RESPONDENT
171	maristat	Nominal	Marital status
172	incomeh	Nominal	Household weekly income
173	incomer	Nominal	Respondent personal weekly income

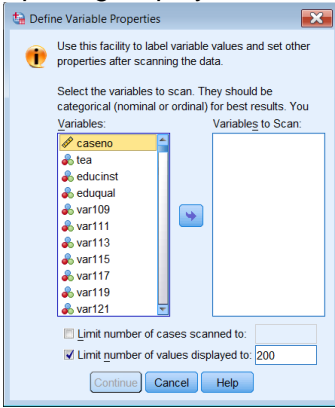| 180	age	Scale	AGE OF RESPONDENT
181	sex	Nominal	SEX OF RESPONDENT
182	houstat	Nominal	RESPONDENT STATUS IN HOUSEHOLD
183	area	Scale	AREA CODE

This could be done with **FREQUENCIES**, but for a quick check which displays all attributes, and also a count for each value encountered:
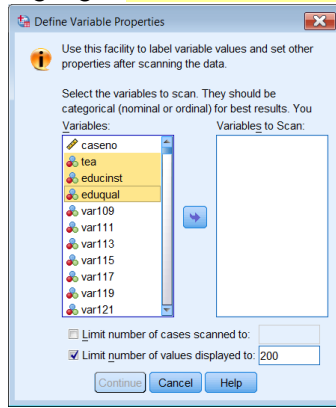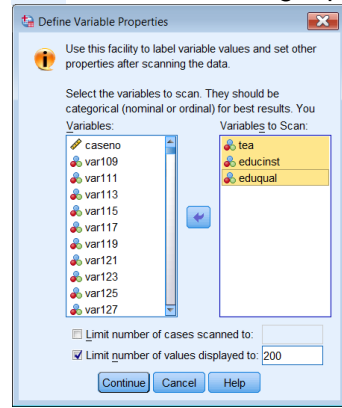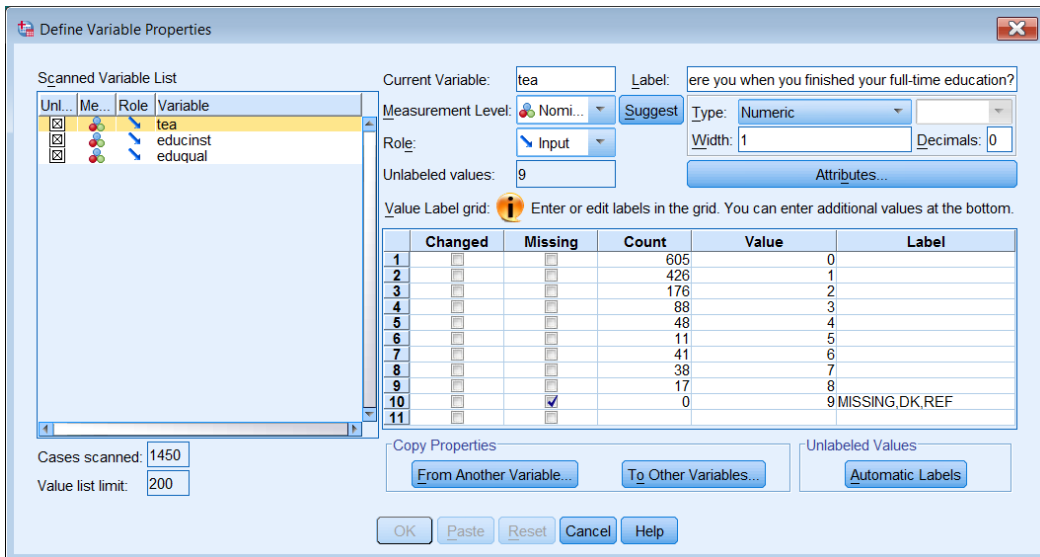
Data >> Define Variable Properties

14

Opening display | Highlight <mark>variables of interest</mark> | to transfer to right pane


Click on [Continue] :

tea   How old were you when you finished your full-time education?

DVP always opens with a display of attributes for the first variable in the list. As you can see, for **tea**, in the **Missing** column, the 10th box is checked (for value 9): there are no value labels except for value 9 (MISSING DK REF) but **tea** has no missing cases, so there is a 0 in the **Count** column for value 9. Clicking on any other variable in the left pane produces the display for that variable.
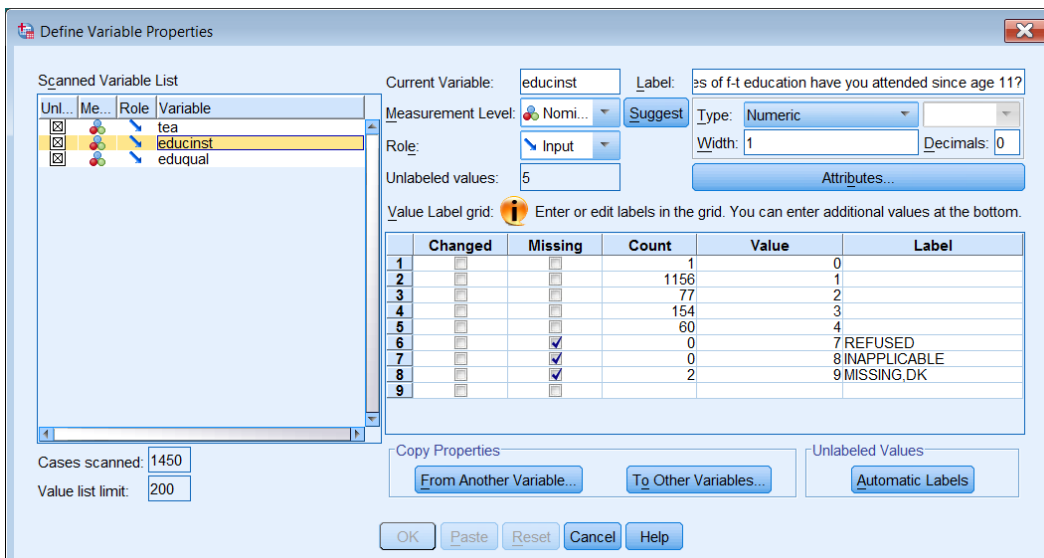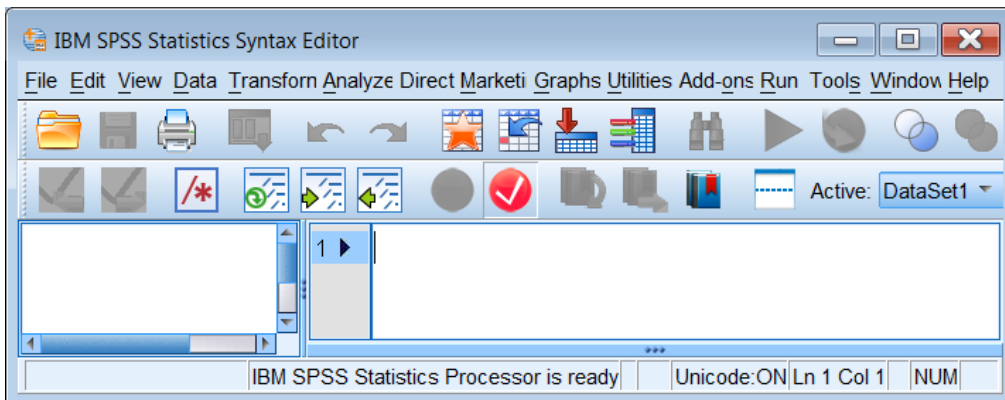

educinst   What schools and other places of f-t education have you attended since age 11?
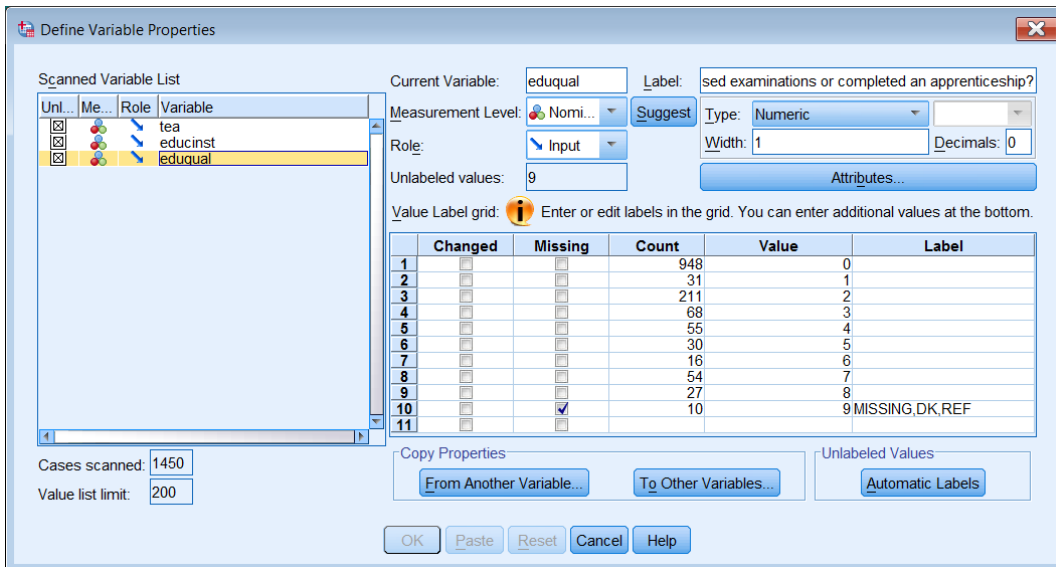
eduqual   Passed any recognised examinations or completed an apprenticeship?

To add the value labels for these three variables:

File >> New >> Syntax:

Carefully type in the following syntax: [Don't forget the full stops!]

```
title 'Value labels for SN 680' .
add value labels
  tea
    0 '14 or under'  1 '15'   2 '16'   3 '17'   4 '18' 5 '19'
    6 '20 or 21'   7 '22 or over'   8 'Still studying'
  /educinst
    0 'None'      1 'School  only'     2 'School and Tralning college'
    3 'School and TechnIcal college'     4 'School and University'
  /eduqual
    0 'None'   1 'CSE'     2 'GCE 0-Level/ONC/OND/C&G etc'
    3 'Full apprenticeship'      4 'A-Level/HSC'
    5 'Teacher Training'      6 'HNC/HND'
    7 'Degree'      8 'Other' .
```
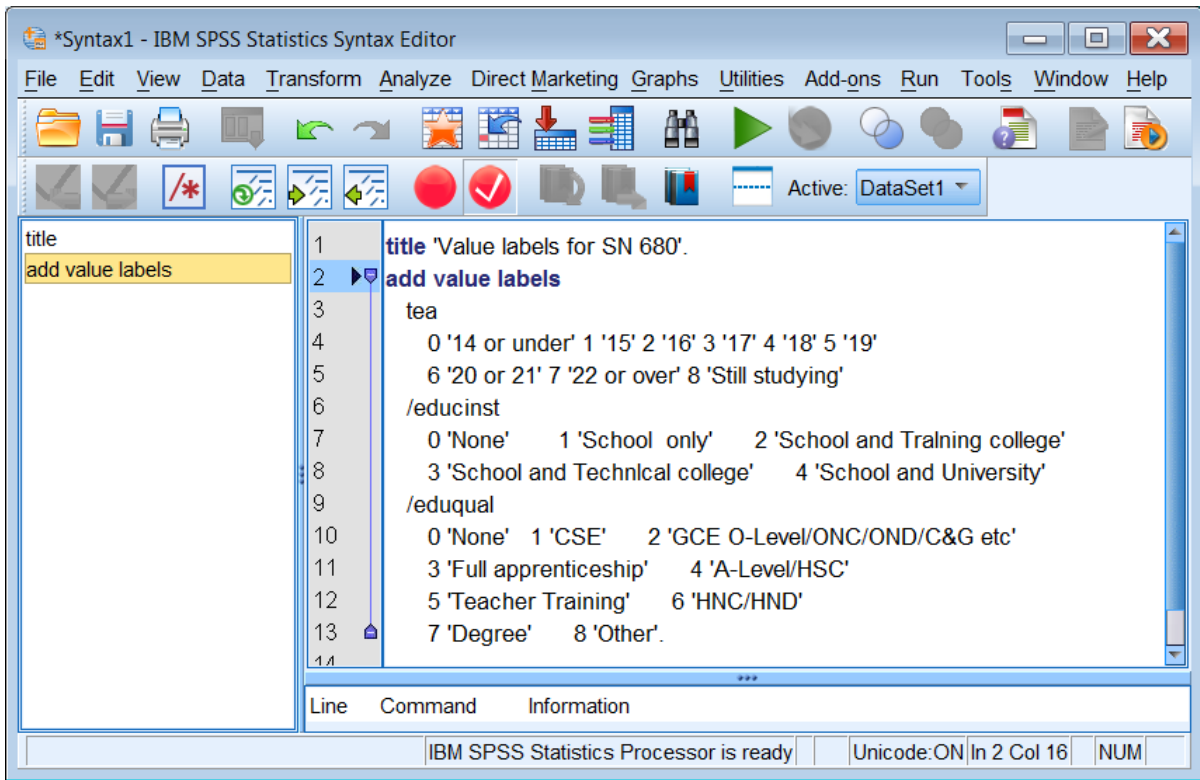
Run >> All:

The **Values** column has changed to display the first non-missing value for each variable: ↓↓

2	tea	👥 Nominal	How old were you when you finished your full-time education?	{1, 14 or under}…
3	educinst	👥 Nominal	What schools and other places of f-t education have you attended since age 11?	{0, None}…
4	eduqual	👥 Nominal	Passed any recognised examinations or completed an apprenticeship?	{0, None}…

You can check each variable by clicking on the cell under **Values**:

and then on the blue square:

If you can't see all the labels, drag the bottom edge of the window down until you can:
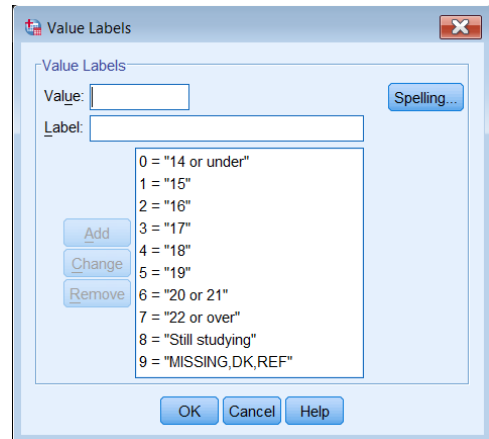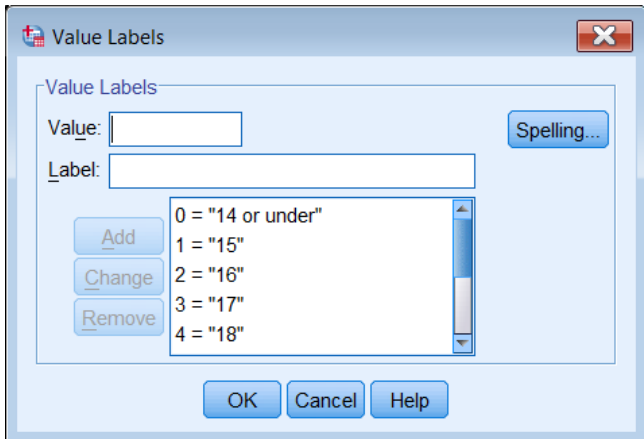
You can do this for any variable in the file, and it's a useful quick check to see what value labels are actually present for each variable.

For point-and-click addicts, this facility also offers an alternative, if tedious, method of editing or adding value labels:
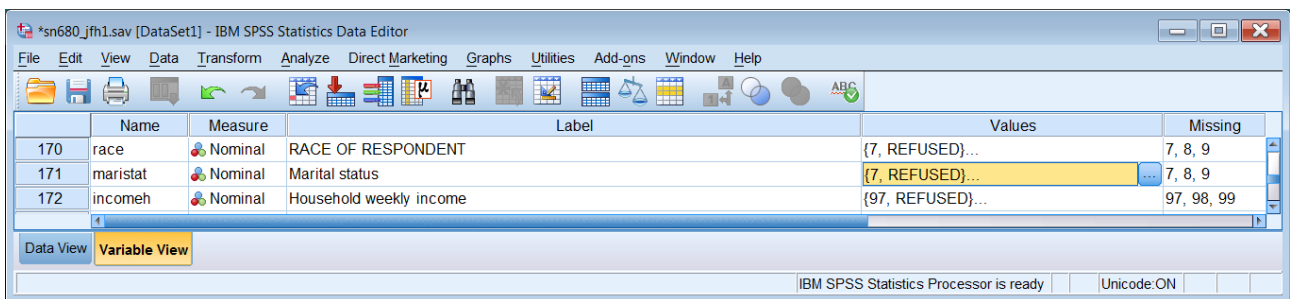
Click on the blue square
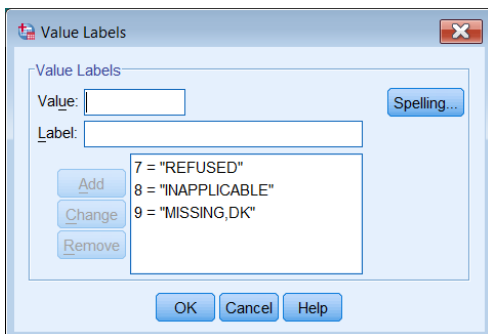
To add labels for the other values, we first need to find the original question. If data-prep layout is indicated on the questionnaire, it is quite straightforward to find positional variable names, However, with mnemonic variable names, this takes some time as it necessitates scrolling through the questionnaire. Variable **maristat** is one of a group of variables sandwiched between **var421** and **var435**. It is followed by **incomeh** and **incomer**, both of which have values running to two digits (the **Missing** column in the **Data Editor** shows 97 and 98 as missing). Working backwards **maristat** must originally have come from record 4, column 30 (ie **var430** in positional terms) so scrolling through the questionnaire looking for column (30) on card 4, we find Q.53:

18

53.	Are you single, married, divorced, separated or widowed?			(30)
			Single	1
			Married	2
			Divorced/separated	3
			Widowed	4

We can now fill in the missing labels.

Write 1 in the Value box and
Single in the Label box, then click on [Add]
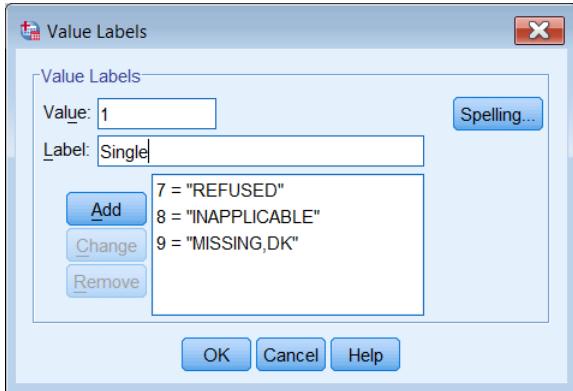
Write 2, then Married and click on [Add]
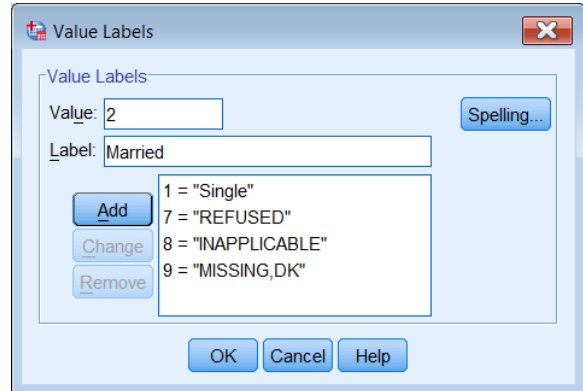
Write 3, then Separated/Divorced:
then click on [Add]
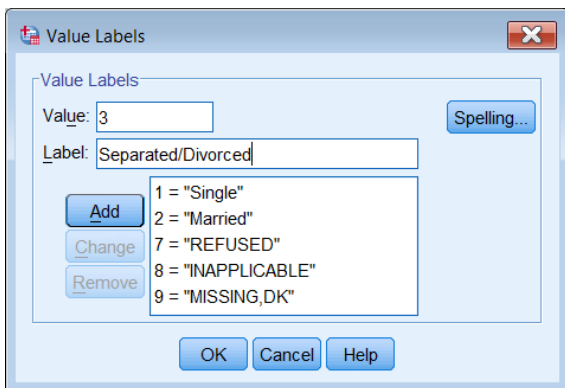
Write 4, then Widowed:
then click on [Add]
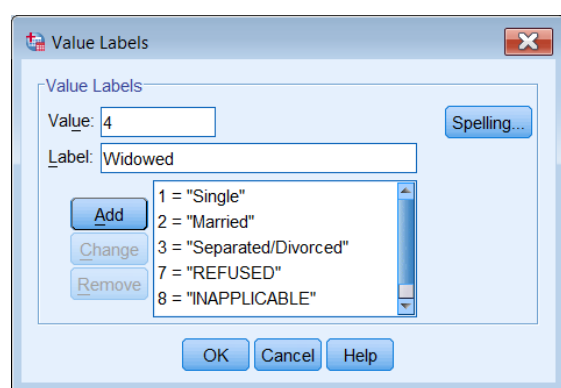
Click on [OK]
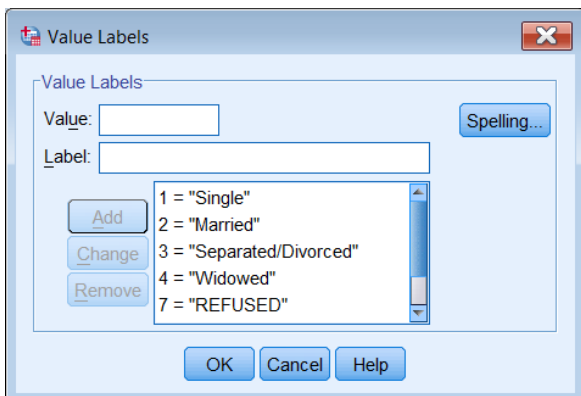
Doing this for every variable may be error free, but is tedious, time-consuming, and, unless you use **PASTE**, leaves no audit trail. Syntax is easier and quicker (see p 16 above).

19

The first value for **maristat** now appears under **Values**
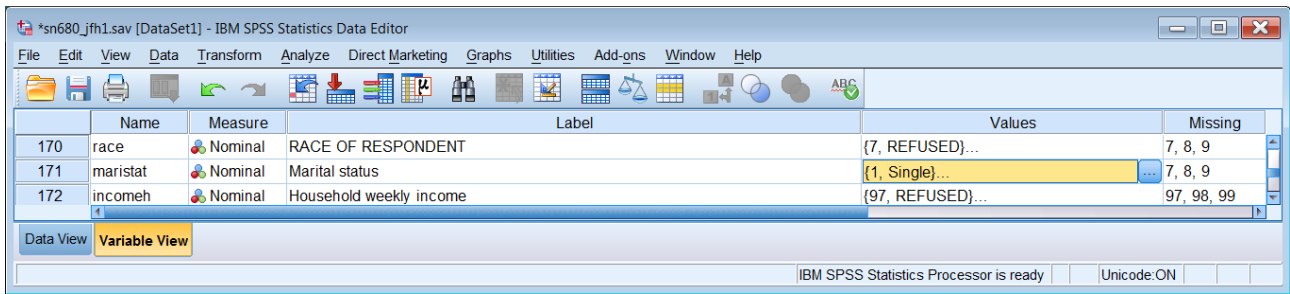
A quick check on your labelling is:

Data >> Define Variable Properties again:
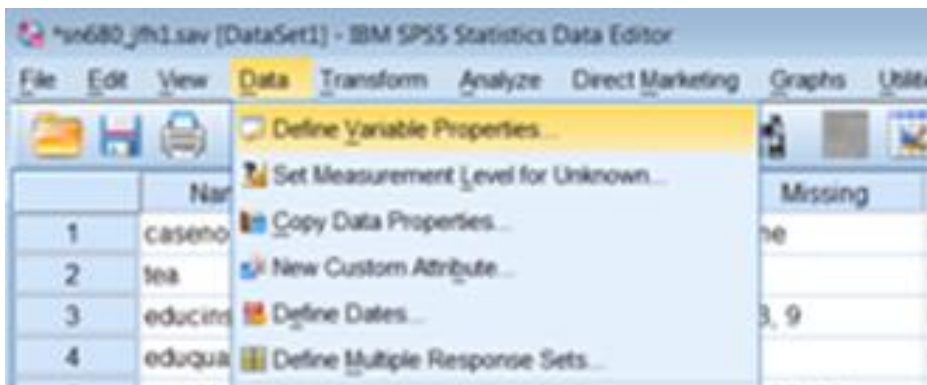
Opening display   Highlight variables of interest   ➡ to transfer to right pane
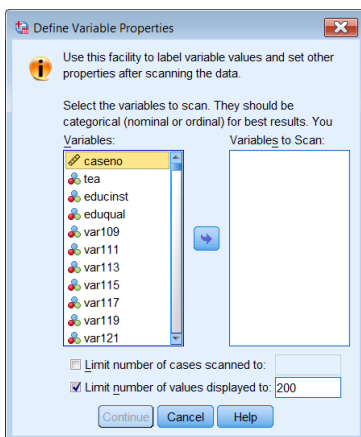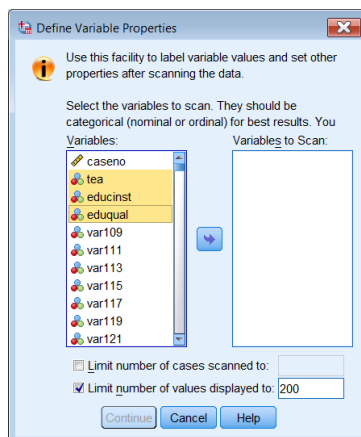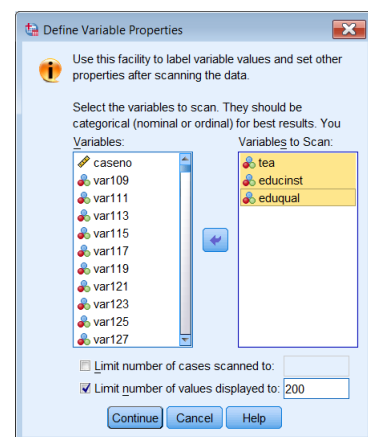
Click on Continue

The **Label** column now has a complete set of value labels: ↓↓

DVP display for **tea** (Q.1: Terminal Education Age)

DVP display for **educinst** (Q.2: Educational institutions attended full-time since age 11)

DVP display for **eduqual** (Q.3: Qualifications obtained from education or training)

 [NB: The original question allowed for multiple-response, but this variable seems to be derived from combinations of these and has a single value indicating the highest qualification mentioned]

Appendices

JFH to Jack Kneeshaw
cc UK Data Service Collections Development Team, Roger.Stafford, (Natcen) Jen Beaumont (ONS) Jacqueline Carter (Manchester) Vanessa Higgins (Manchester)

I've now explored, and written comments on, a number of SPSS files distributed by ESDS (and some by Roper) for major survey series, including: (Commentaries downloadable from; http://surveyresearch.weebly.com/spss-files-and-documentation-used-for-tutorials-and-exercises.html)

British Social Attitudes
Commentary on SPSS file for British Social Attitudes 2011

Notes on British Social Attitudes 2004 teaching data set
(as used by Marsh and Elliott, 2008)

Understanding Society
Commentary on Understanding Society 2010

NORC General Social Survey (GSS)
Commentary on full NORC General Social Survey 2008

Commentary on subset of General Social Survey 2008
(as used by Sweet & Grace-Martin)

Commentary on GSS 2008 SPSS files for Babbie et al
(as used by Babbie, Halley, Wagner & Zaino)

I've also been looking at:

ONS Opinions Survey: Well-being module (April 2011)
(Data set and user guide from Cathie Marsh Centre for Census and Survey Research, Manchester)

Taking Part
Three waves 2005-2008 (SN 5717, 6272 and 6273)

I have observed several properties some or all of them have in common.

1: Unless measurement levels have been specified before deposit, they are invariably incorrect and seem to derive from the automatic levels "heuristically" allocated by SPSS depending on the values encountered. These will be Nominal for variables with a narrow range of values and Scale for everything else.

2: Missing values may be specified, but are incomplete and sometimes inconsistent: values which should be treated as missing are still encountered. When value labels are specified for such values they are tabulated as non-missing.

3: Values which should be integer have two decimal places (f8.2 being the default format for SPSS).

4: (Less serious and mainly cosmetic) Some variable and value labels are inordinately long. In some files they also contain errors of spelling and grammar.

In commentaries I demonstrate various ways of exploring these files by copying them to a new file and making the necessary modifications directly in the Data Editor, by using Data >> Define Variable Properties, or by writing a new *.sps setup file with new specifications using at least the following SPSS commands:

VARIABLE LEVEL
FORMATS
MISSING VALUES
ADD VALUE LABELS

(and advise users to do the same, preferably before doing any analysis.

This leads me to wonder how the SPSS files are generated. Do they come ready-made from the depositors, are they generated by CAPI programs or later at Essex (possibly via Stata)? Are there any associated SPSS *.sps setup files? If so, who actually writes them?

Jack Kneeshaw to JFH 19 March 10:38

Please see http://www.data-archive.ac.uk/curate/process and http://www.data-archive.ac.uk/curate/archive-quality - and links off those pages - for a full description of what we do to the data we receive.

JFH to Jack Kneeshaw

Many thanks for these. I'm busy packing for an early departure to UK tomorrow and will examine them closely on my return.

For my ASSESS presentation at York in October, I'm going to need full SPSS files for selected surveys to be uploaded/copied to their computer(s) for use on Thur 30th (by me for testing) and Fri 31st for talk and workshop. These will be used to demonstrate how to check and improve file contents and structures and also the Python code for dealing with lower and upper case labelling. Delegates will be given USB sticks with my paper, slideshows etc., but not the actual data. For the workshop session Survey Analysis for Beginners I'll also need at least one data set available. I can always use one of my own (easier to follow when reading data from ASCII files) but a more recent one would probably be more interesting, even though CAPI is less easy to follow in a very short session.

What's the procedure to be followed? Do I have to seek depositors'permissions as well as get delegates to sign ESDS confidentiality agreements?

JFH to UKDS 8 April

Datasets for usages:

78349 - Comparison of response distributions
67995 - Secondary analysis of data from Quality of Life surveys

I've checked the documentation, but do you still have any files, even raw data in 80-column ASCII format, and SPSS setups for SN680?

680   SSRC Survey Unit Multi-Purpose Survey, 1975
         Standard formats not available.Order via Other media or emailhelp@ukdataservice.ac.uk

(QTHELP-6129) Request for data files for SN 680

UKDS to JFH 8 April 12:12

```
Thank you for contacting the help desk.

SN 680 is currently in the processes of being converted to SPSS. This will take a
couple of weeks.

If you would like to place an order for SN680 in the data catalogue, we will let
you know when it is ready.
```

JFH to UKDS 8 April 17:19

Many thanks for this info. Please place an order for me for SN 680. If you need any help with the reconstruction, let me know.

UKDS automatic procedures tend to require adjustments to files to set correct measurement levels and missing values, but I'm used to this with other files I've downloaded (eg British Social Attitudes,

Understanding Society). Sometimes variable and value labels also need adjustment and this could well be the case with SN 680 as it was done in the days before SPSS could handle mixed case text, so names and labels will probably all be in UPPER CASE, with variable labels to 40 printing characters and value labels to 20 (only 16 in column headers). In those days, to generate automatic variable names using the TO facility, variables had to be called VARddd TO VARddd, later still allowing different (combinations of) letters (eg Q1 to Q25 etc) but even today you can't use Q1a to Q1g.

I can make these adjustments and return the file to UKDS for you to archive.

JFH to UKDS 9 April

UKDS to JFH 8 April 21:12

Thank you for your reply and offer of help.

I have passed on your comments to our data team.

I am unable to create the order at this end. You will need to place the order at http://discover.ukdataservice.ac.uk/catalogue/?sn=680&type=Data%20catalogue , and then add it to the usage of your choice.

Once we have your order I can make the data available when its ready.

JFH to UKDS 9 April 15:20

I followed your order instructions and downloaded a *.tab file, but Windows wouldn't open it. SPSS was also listed, so I went back in and clicked on that. I expected to get nothing, but I now have the SPSS file for SN 280. No idea why I couldn't do this yesterday, but everything went smoothly with a new login.

There are 249 variables and 1450 cases in the *.sav file.

Examples below are extracts for the first 8 variables in the file.

Variable names are a mixture of mnemonic and positional types: they are all in lower case, but SPSS is case insensitive for these:

caseno
tea
educinst
eduqual
var109
var111
var113
var115

Variable labels are a mixture of UPPER CASE and Mixed Case:

Case number
How old were you when you finished your full-time education?
What schools and other places of f-t education have you attended since age 11?
Passed any recognised examinations or completed an apprenticeship?
IMP. OF HOUSING CONDITIONS
IMP. OF NEIGHBOURHOOD
IMP. OF SOCIAL STATUS

IMP. OF HEALTH

Value labels are all in UPPER CASE

None
{9, MISSING,DK,REF}...
{7, REFUSED}...
{9, MISSING,DK,REF}...
{97, REFUSED}...
{97, REFUSED}...
{97, REFUSED}...
{97, REFUSED}...

Missing values appear to have been correctly and consistently declared:

None
9
7, 8, 9
9
97, 98, 99
97, 98, 99
97, 98, 99
97, 98, 99

Measurement levels are those heuristically allocted by SPSS:

Scale
Nominal
Nominal
Nominal
Nominal
Nominal
Nominal
Nominal

will need to be checked and corrected.

I'll work my way through the file and make any necessary modifications, including converting variable and value labels to mixed case using a chunk of Python code used on other surveys, written to my specifications by Jon Peck (Senior Software Engineer, IBM/SPSS). When it's finished I'll send the new version back to you.

I may inadvertently have ordered stuff on CD, so perhaps you could check and cancel that if necessary.

Just out of interest what software would I need to open the *.tab file?

JFH to UKDS 13 April

SSRC Survey Unit Multi-Purpose Survey, 1975 [SN 680]

The SPSS file has value labels for only a few classification and household grid variables. If you scroll down the **Values** column, you will see that most values displayed start at 7 or 97 (Refused). I'll have to use syntax to add value labels manually for all variables although I might be able to find a short cut by copy/pasting from the user guide. With over 200 variables this may take some time.

Appendix

Facsimile of pamphlet for Multi-purpose Social Research Survey

SSRC Survey Unit
Multipurpose Survey

The SSRC Survey Unit is organizing a multipurpose survey for academic social researchers. This will be a national survey with a sample of 2,000 adults. It will use a rigorously designed probability sample with a low level of clustering (100 wards, 20 electors per ward). Participants, who must be individual academics, educational institutions or other non-profit bodies, will be able to buy interviewing time in surveys which will be carried out at regular intervals – initially twice a year, in April and October. The total questionnaire will be such that it will not take, on average, more than 45 minutes interviewing time.

POSSIBLE ADVANTAGES TO PARTICIPANTS ARE:

1. They will be able to obtain national data for comparison with their own local or regional studies.

2. The survey can be used to pretest the strength or weakness of research hypotheses, in advance of more intensive study.

3. It will be possible over a period of time to accumulate information about minority groups, based on a series of comparatively large national samples.

4. The surveys will be designed to assist the study of trends, by rotating the sample areas so as to facilitate comparisons between successive surveys.

Fieldwork will be spread over a period of 4–5 weeks, normally starting in mid-April and mid-October each year, and the results will be available within a month of the completion of fieldwork. Participants will be provided with a clean dataset, consisting of the replies to their own questions and the standard classificatory information, on punched cards or magnetic tape.

A more detailed specification of the survey will be found overleaf.

Some Technical Details

Area The Survey will cover Great Britain (not N. Ireland) excluding the most northerly parts of Scotland, north of the Caledonian Canal.

Sample A sample of named individuals aged 18 and over, drawn from the Electoral Register, with provision for a random substitute from the same address if, and only if, the entire household to which the named individual belonged has moved away.

The sample will be a two-stage sample.

The first stage selects 100 local government wards with probability approximately proportional to the number of electors; the second stage will aim at 20 successful interviews per ward, spread evenly through the ward. The lists for the selection of local government areas will be arranged so as to incorporate a degree of stratification:

a) by region;

b) into — conurbations, sub-divided into high and low employment areas;

   — urban areas, sub-divided by population;

   — rural areas, sub-divided by population density;

c) by listing according to a social grade index.

The sample areas will be "rotated" in the same way as in the government's General Household Survey (see article in *SSRC Newsletter 23*). One quarter of the wards will be replaced in each successive survey by a newly selected ward from the same stratum. Thus successive surveys will have three-quarters of the wards in common; and, once the survey is in full operation, each selected ward will remain in the sample for four successive rounds (but with a new sample of electors each time). This should facilitate the study of change over time.

Full details of the sample, including the names of the selected wards, will be available to subscribers before each survey is undertaken.

The Question- naire This will include questions related to the Survey Unit's own research, questions requested by subscribers and standard classificatory questions. The Survey Unit must reserve the right to reject questions which cannot be fitted in to make a satisfactory questionnaire. All subscribers will be given a broad indication of the questionnaire content as early as possible and will be shown the final questionnaire before fieldwork begins. They will have the right to withdraw if they do not consider it satisfactory.

Classifi- catory infor- mation The standard classificatory information, to be provided to every subscriber, will be broadly comparable with that obtained in the National Readership Survey and will include information about other members of the household, as well as the respondent (e.g. age, sex, marital status, educational background and occupation status — whether employed full- or part-time, retired, etc.). If the respondent is not head of household, we shall obtain details of occupation for both respondent and head. We shall ask about income (grouped), for both respondent and household.

Standard output This will be a pack of punched cards, with data from all the subscriber's own questions and the standard classificatory information.

Timing Fieldwork will normally begin in mid-April (depending on the date of Easter) and mid-October each year. To achieve this the questionnaire must be finalized a month earlier; i.e. the final date for receipt of questions will be about end-February and end-August. In most cases,

preliminary discussion will need to have started at least a month earlier: by the end of January for the April survey, and the end of July for the October survey.

Data packs will be available by about the end of June and December respectively.

Field-work control
The Survey Unit does not have its own body of interviewers. Fieldwork, coding and preparation of the data pack will be contracted to a commercial organization. The Unit will, however, ensure that there is adequate supervision of these operations and that the fieldwork control includes a check by personal recall on a sample of respondents.

Scale of Charges

The basic fees to participants will be as follows (all subject to VAT at standard rate):

Entry Fee: £150

Question fee (up to 6 pre-codes per question): £100

Additional pre-codes (up to 6 per question): £25

Use of showcards (prompt lists): £25

Open ended questions are charged double

Discounts

Repeated participation:	20% discount for second and subsequent entries with the same questions
Minority samples:	33% reduction for questions asked only of half, or less, of the sample (e.g. M or F only)
Batteries of scaled questions:	by negotiation.

Special requirements

The Unit will negotiate appropriate rates for special requirements such as increased sample size in particular areas, special output formats, etc.

Any of the above charges are subject to variation by agreement. The most important variable elements in the cost of the survey (and therefore of the charge to participants) are the interviewing and coding time required. Charges may be varied, therefore, if a particular participant's questions make unusually light or heavy demands on interviewing and coding time.

Further information about the Multi-purpose Survey may be obtained from:

John Utting
Deputy Director
SSRC Survey Unit
Regent House
89 Kingsway
London WC2B 6RH
Tel: 01-405 6491