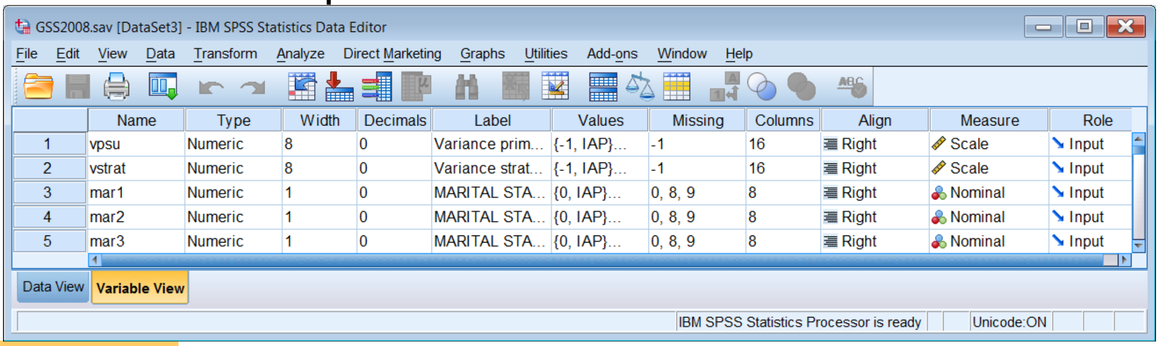
Data sets and documents used for tutorials and exercises

NORC General Social Survey 2008 [Draft only 8 January 2014]

First encounter with GSS2008.SAV[1]

Here's what I tend to do with new-to-me SPSS saved files the first time I open them. The file **GSS2008.SAV** arrived from the Roper Center in a zip file: when first opened it looked like this:

GSS2008.SAV as first opened

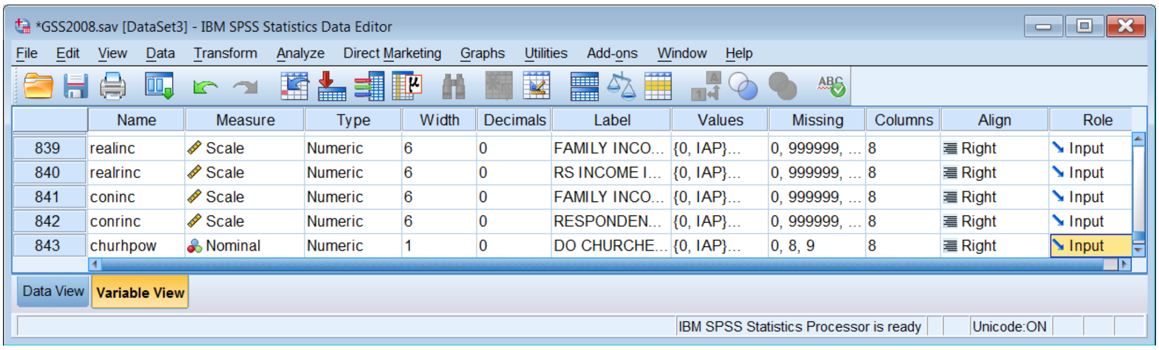

GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	vpsu	Numeric	8	0	Variance prim...	{-1, IAP}...	-1	16	Right	Scale	Input
2	vstrat	Numeric	8	0	Variance strat...	{-1, IAP}...	-1	16	Right	Scale	Input
3	mar1	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
4	mar2	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
5	mar3	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input

Variable View

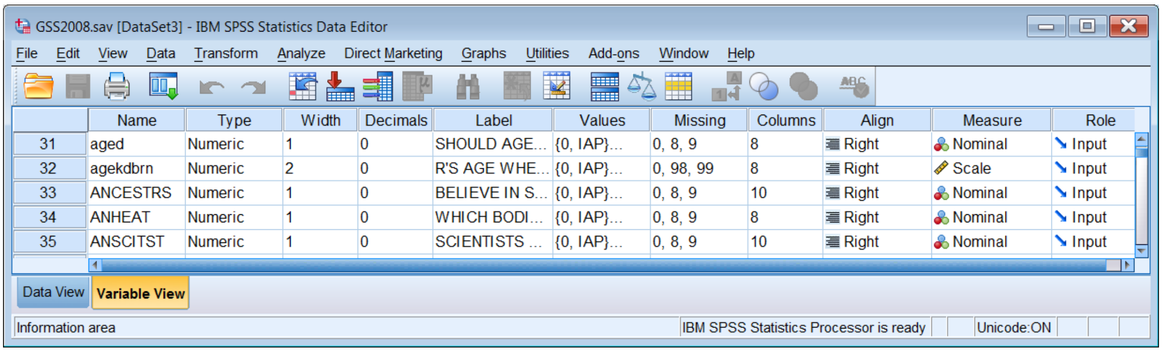
First, a quick check on file contents:

In Variable View, press [Ctrl] + [End] to see how many variables there are (843):


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Role
839	realinc	Scale	Numeric	6	0	FAMILY INCO...	{0, IAP}...	0, 999999, ...	8	Right	Input
840	realrinc	Scale	Numeric	6	0	RS INCOME I...	{0, IAP}...	0, 999999, ...	8	Right	Input
841	coninc	Scale	Numeric	6	0	FAMILY INCO...	{0, IAP}...	0, 999999, ...	8	Right	Input
842	conrinc	Scale	Numeric	6	0	RESPONDEN...	{0, IAP}...	0, 999999, ...	8	Right	Input
843	churhpow	Nominal	Numeric	1	0	DO CHURCHE...	{0, IAP}...	0, 8, 9	8	Right	Input

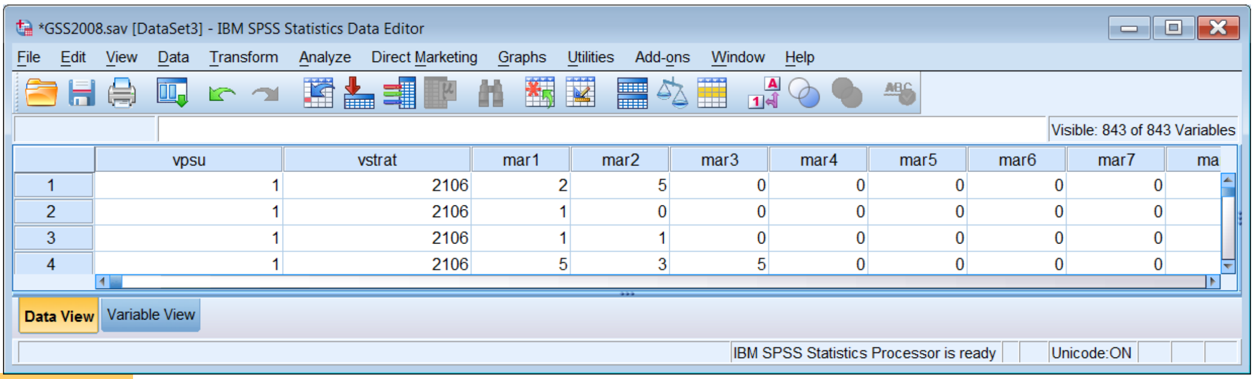
After a quick scroll up and down the file just to see what's there, and inspecting a few labels, missing values etc., I notice that, in the file as distributed, apart from the first 13 rows, variables are in alphabetical order: most variable names are in lower case (as above) but some are in UPPER CASE as are most variable labels:


GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
Information area
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
31	aged	Numeric	1	0	SHOULD AGE...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
32	agekdbrn	Numeric	2	0	R'S AGE WHE...	{0, IAP}...	0, 98, 99	8	Right	Scale	Input
33	ANCESTRS	Numeric	1	0	BELIEVE IN S...	{0, IAP}...	0, 8, 9	10	Right	Nominal	Input
34	ANHEAT	Numeric	1	0	WHICH BODI...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
35	ANSCITST	Numeric	1	0	SCIENTISTS ...	{0, IAP}...	0, 8, 9	10	Right	Nominal	Input

[1] gss2008.sav (SPSS *.sav file for full NORC 2008 survey: 843 variables, 2023 cases) distributed by the Roper Center.

Switch to Data View:



*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor

Visible: 843 of 843 Variables

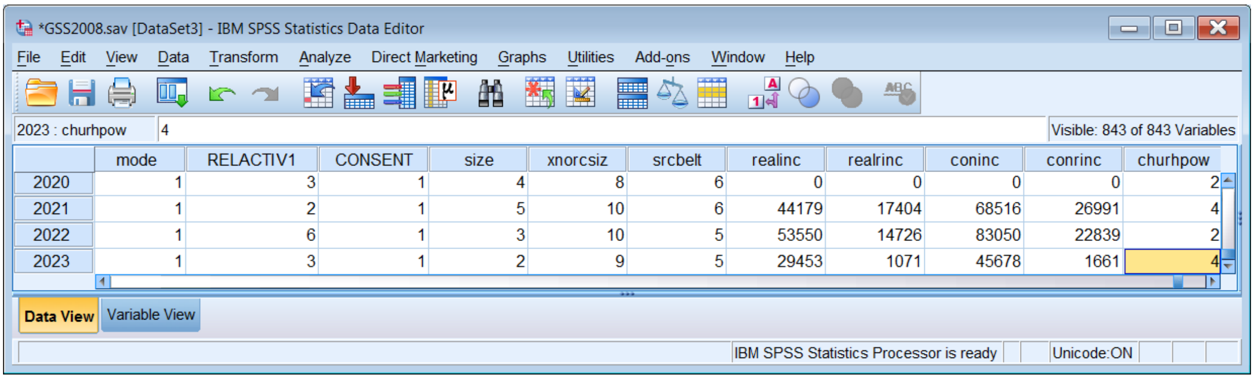
	vpsu	vstrat	mar1	mar2	mar3	mar4	mar5	mar6	mar7	ma
1	1	2106	2	5	0	0	0	0	0	
2	1	2106	1	0	0	0	0	0	0	
3	1	2106	1	1	0	0	0	0	0	
4	1	2106	5	3	5	0	0	0	0	

IBM SPSS Statistics Processor is ready

Unicode:ON

Data View

Press [Ctrl] + [End] to see how many cases there are (2023):



*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor

2023 : churhpow 4

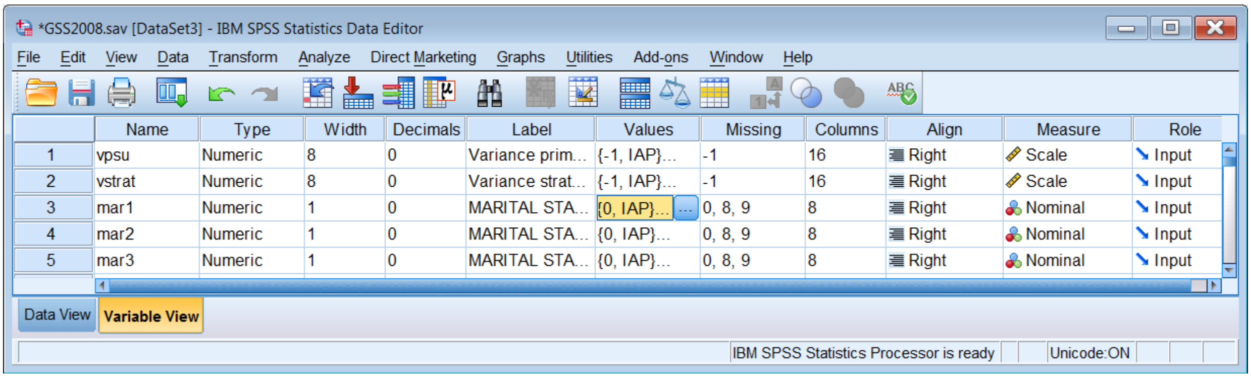
Visible: 843 of 843 Variables

	mode	RELACTIV1	CONSENT	size	xnorcsiz	srcbelt	realinc	realrinc	coninc	conrinc	churhpow
2020	1	3	1	4	8	6	0	0	0	0	2
2021	1	2	1	5	10	6	44179	17404	68516	26991	4
2022	1	6	1	3	10	5	53550	14726	83050	22839	2
2023	1	3	1	2	9	5	29453	1071	45678	1661	4

IBM SPSS Statistics Processor is ready

Unicode:ON

For historical reasons, GSS2008.sav has variable names with a maximum length of 8 characters, but SPSS can now have much longer names. Most variable labels and value labels used to have a maximum of 40 characters and 20 characters respectively, but both can also now be much longer (one variable label in this file is 97 characters long!). Regardless of whether variable names are in UPPER CASE or lower case, most variable and value labels are in UPPER CASE, but others are in Mixed Case. I wonder why?



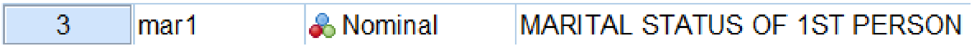
*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	vpsu	Numeric	8	0	Variance prim...	{-1, IAP}...	-1	16	Right	Scale	Input
2	vstrat	Numeric	8	0	Variance strat...	{-1, IAP}...	-1	16	Right	Scale	Input
3	mar1	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
4	mar2	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
5	mar3	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input

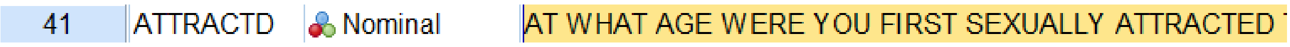
IBM SPSS Statistics Processor is ready

Unicode:ON

It looks as if variables with names in lower case are original and have UPPER CASE value labels. Variables with names in UPPER CASE seem to be either derived or used in secondary analysis and have Mixed Case value labels.

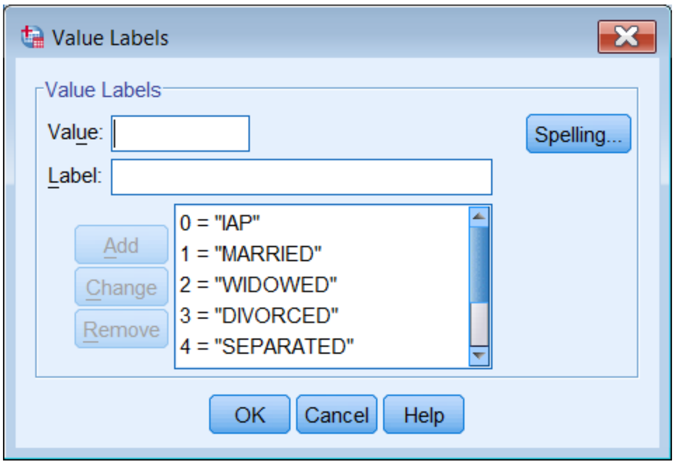


3	mar1	Nominal	MARITAL STATUS OF 1ST PERSON



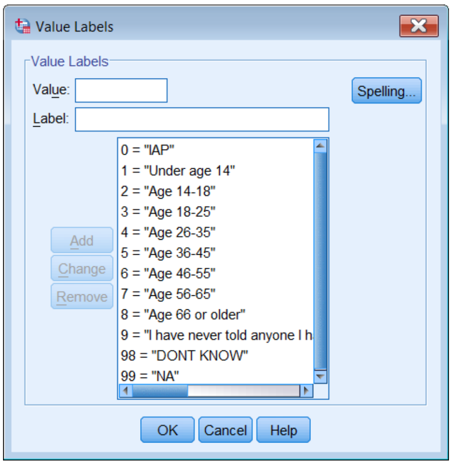
41	ATTRACTD	Nominal	AT WHAT AGE WERE YOU FIRST SEXUALLY ATTRACTED T

mar1
[MARITAL STATUS OF 1ST PERSON]



0 = "IAP"
1 = "MARRIED"
2 = "WIDOWED"
3 = "DIVORCED"
4 = "SEPARATED"

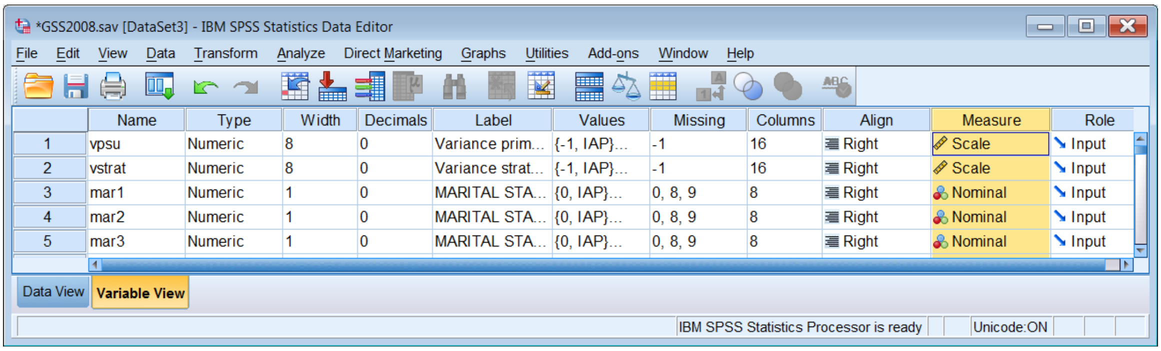
ATTRACTD
[AT WHAT AGE WERE YOU FIRST SEXUALLY ATTRACTED TO SAME SEX?]



0 = "IAP"
1 = "Under age 14"
2 = "Age 14-18"
3 = "Age 18-25"
4 = "Age 26-35"
5 = "Age 36-45"
6 = "Age 46-55"
7 = "Age 56-65"
8 = "Age 66 or older"
9 = "I have never told anyone I h
98 = "DONT KNOW"
99 = "NA"

After browsing around the file I tend to change the column positions around, moving the (to me) more important columns to the left. Most SPSS files I deal with are from questionnaire surveys in which all variables are Numeric with no decimal places (ie Integer): any String variables will have been converted to numeric and the only variables with decimal places will be sample weights, height in metres, weight in kilograms, distances or derived scores. Thus I'm not immediately interested in attributes such as **Type**, **Width**, **Decimals**, **Align** or **Role**, and so will move **Measure**, **Label**, **Value** and **Missing** further to the left. This can be done by changing the SPSS settings to change the order or even suppress some attributes, but for now it's easier to do it by highlighting the attributes and dragging them to a new position:

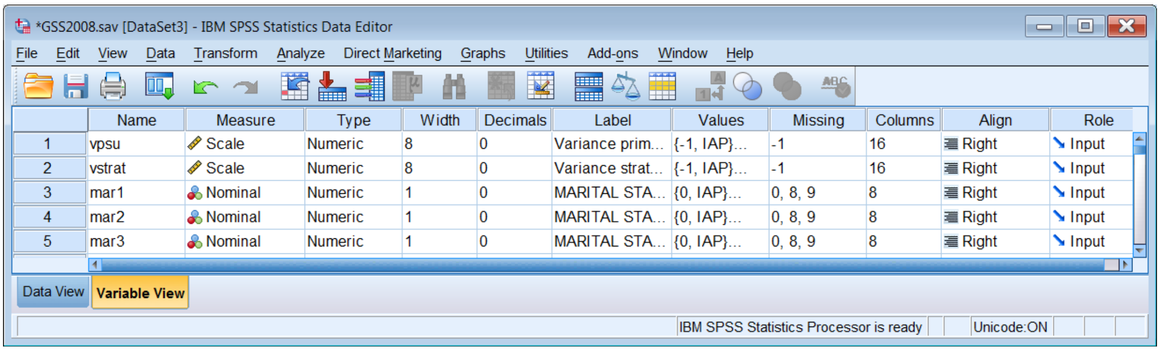
Left click once on **Measure** to highlight the whole column:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	vpsu	Numeric	8	0	Variance prim...	{-1, IAP}...	-1	16	Right	Scale	Input
2	vstrat	Numeric	8	0	Variance strat...	{-1, IAP}...	-1	16	Right	Scale	Input
3	mar1	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
4	mar2	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
5	mar3	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input

IBM SPSS Statistics Processor is ready   Unicode:ON

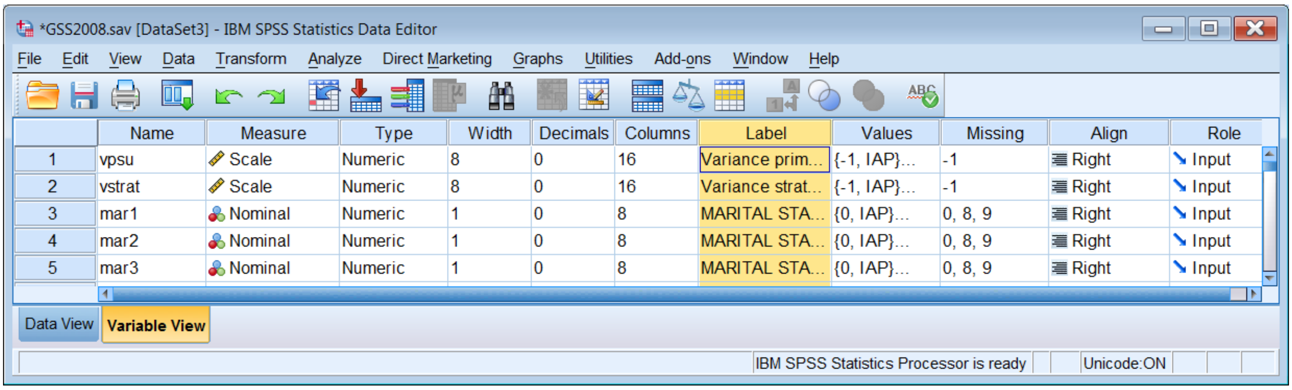
Click and hold down the left mouse button on **Measure** to drag it to its new position after **Name** (a thin vertical red line at the left of each column indicates the destination currently reached).



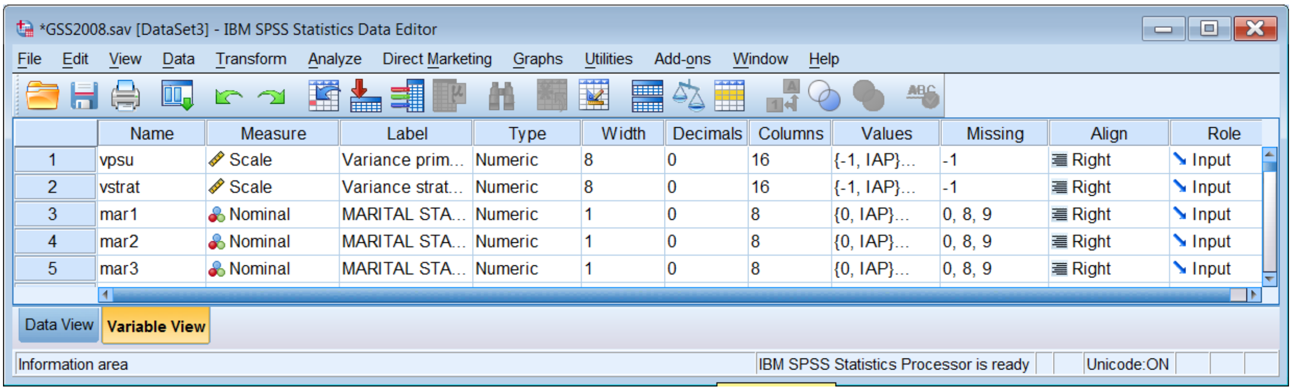
	Name	Measure	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Role
1	vpsu	Scale	Numeric	8	0	Variance prim...	{-1, IAP}...	-1	16	Right	Input
2	vstrat	Scale	Numeric	8	0	Variance strat...	{-1, IAP}...	-1	16	Right	Input
3	mar1	Nominal	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Input
4	mar2	Nominal	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Input
5	mar3	Nominal	Numeric	1	0	MARITAL STA...	{0, IAP}...	0, 8, 9	8	Right	Input

IBM SPSS Statistics Processor is ready   Unicode:ON

Repeat for **Label**:

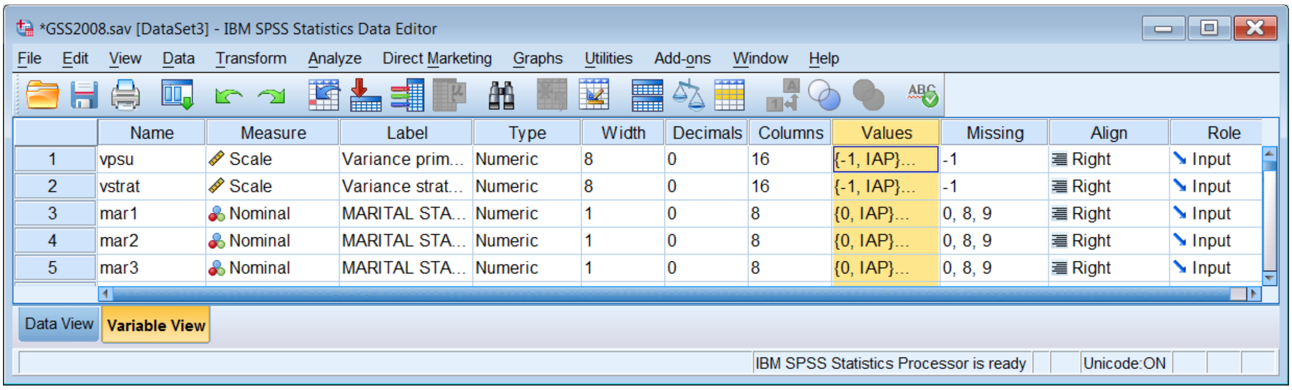

*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Type	Width	Decimals	Columns	Label	Values	Missing	Align	Role
1	vpsu	Scale	Numeric	8	0	16	Variance prim...	{-1, IAP}...	-1	Right	Input
2	vstrat	Scale	Numeric	8	0	16	Variance strat...	{-1, IAP}...	-1	Right	Input
3	mar1	Nominal	Numeric	1	0	8	MARITAL STA...	{0, IAP}...	0, 8, 9	Right	Input
4	mar2	Nominal	Numeric	1	0	8	MARITAL STA...	{0, IAP}...	0, 8, 9	Right	Input
5	mar3	Nominal	Numeric	1	0	8	MARITAL STA...	{0, IAP}...	0, 8, 9	Right	Input

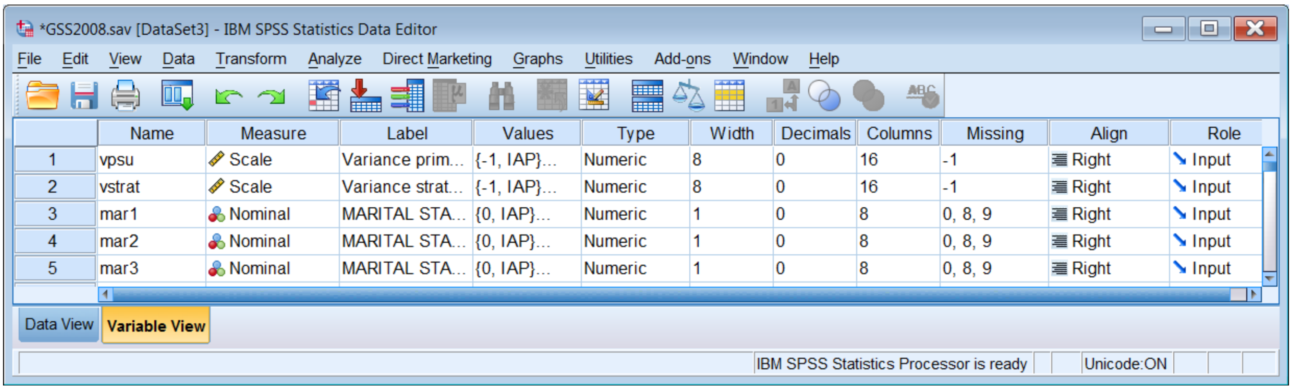

*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
Information area
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Type	Width	Decimals	Columns	Values	Missing	Align	Role
1	vpsu	Scale	Variance prim...	Numeric	8	0	16	{-1, IAP}...	-1	Right	Input
2	vstrat	Scale	Variance strat...	Numeric	8	0	16	{-1, IAP}...	-1	Right	Input
3	mar1	Nominal	MARITAL STA...	Numeric	1	0	8	{0, IAP}...	0, 8, 9	Right	Input
4	mar2	Nominal	MARITAL STA...	Numeric	1	0	8	{0, IAP}...	0, 8, 9	Right	Input
5	mar3	Nominal	MARITAL STA...	Numeric	1	0	8	{0, IAP}...	0, 8, 9	Right	Input

Repeat for **Values**

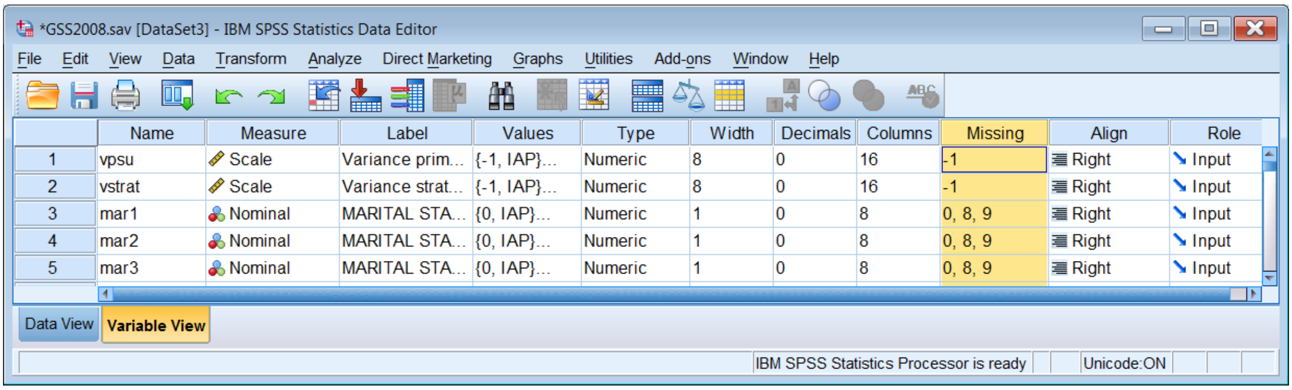

*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Type	Width	Decimals	Columns	Values	Missing	Align	Role
1	vpsu	Scale	Variance prim...	Numeric	8	0	16	{-1, IAP}...	-1	Right	Input
2	vstrat	Scale	Variance strat...	Numeric	8	0	16	{-1, IAP}...	-1	Right	Input
3	mar1	Nominal	MARITAL STA...	Numeric	1	0	8	{0, IAP}...	0, 8, 9	Right	Input
4	mar2	Nominal	MARITAL STA...	Numeric	1	0	8	{0, IAP}...	0, 8, 9	Right	Input
5	mar3	Nominal	MARITAL STA...	Numeric	1	0	8	{0, IAP}...	0, 8, 9	Right	Input

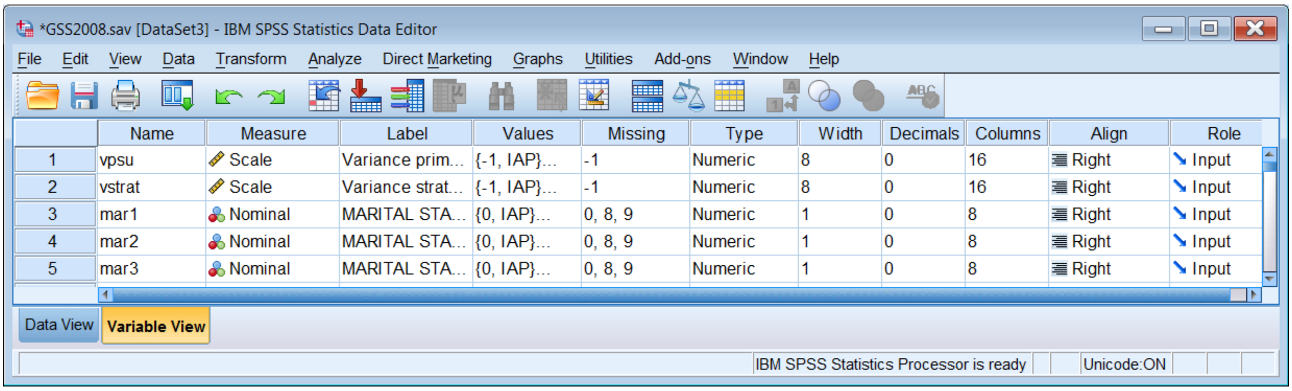

*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Values	Type	Width	Decimals	Columns	Missing	Align	Role
1	vpsu	Scale	Variance prim...	{-1, IAP}...	Numeric	8	0	16	-1	Right	Input
2	vstrat	Scale	Variance strat...	{-1, IAP}...	Numeric	8	0	16	-1	Right	Input
3	mar1	Nominal	MARITAL STA...	{0, IAP}...	Numeric	1	0	8	0, 8, 9	Right	Input
4	mar2	Nominal	MARITAL STA...	{0, IAP}...	Numeric	1	0	8	0, 8, 9	Right	Input
5	mar3	Nominal	MARITAL STA...	{0, IAP}...	Numeric	1	0	8	0, 8, 9	Right	Input

and **Missing**

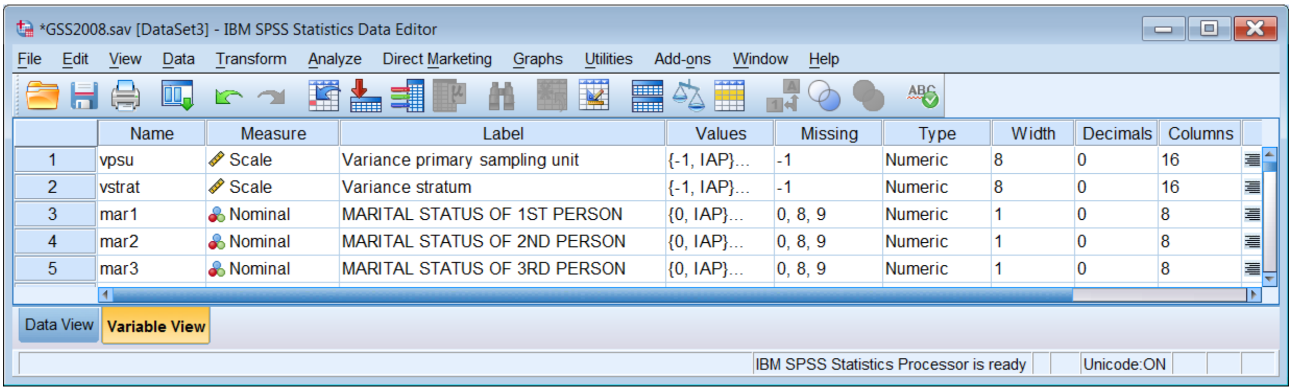

*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Values	Type	Width	Decimals	Columns	Missing	Align	Role
1	vpsu	Scale	Variance prim...	{-1, IAP}...	Numeric	8	0	16	-1	Right	Input
2	vstrat	Scale	Variance strat...	{-1, IAP}...	Numeric	8	0	16	-1	Right	Input
3	mar1	Nominal	MARITAL STA...	{0, IAP}...	Numeric	1	0	8	0, 8, 9	Right	Input
4	mar2	Nominal	MARITAL STA...	{0, IAP}...	Numeric	1	0	8	0, 8, 9	Right	Input
5	mar3	Nominal	MARITAL STA...	{0, IAP}...	Numeric	1	0	8	0, 8, 9	Right	Input


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

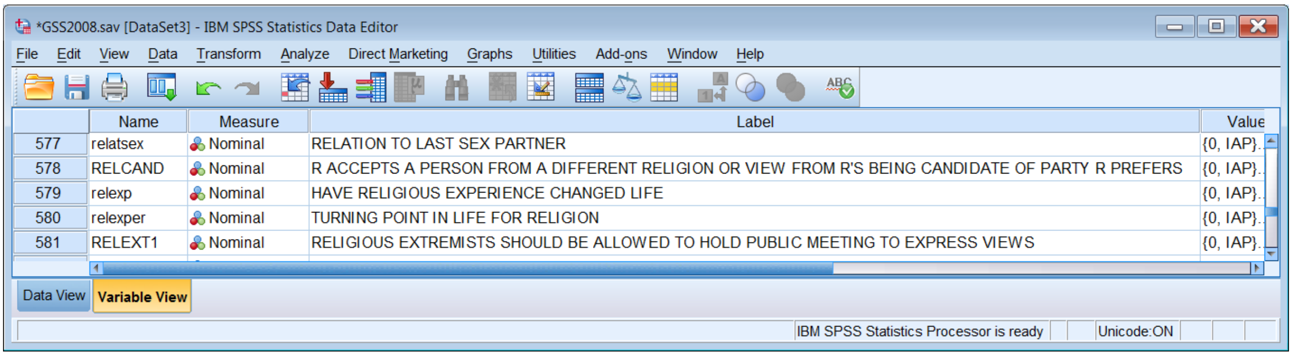
	Name	Measure	Label	Values	Missing	Type	Width	Decimals	Columns	Align	Role
1	vpsu	Scale	Variance prim...	{-1, IAP}...	-1	Numeric	8	0	16	Right	Input
2	vstrat	Scale	Variance strat...	{-1, IAP}...	-1	Numeric	8	0	16	Right	Input
3	mar1	Nominal	MARITAL STA...	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right	Input
4	mar2	Nominal	MARITAL STA...	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right	Input
5	mar3	Nominal	MARITAL STA...	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right	Input

Once I've got the columns in my preferred order, I also widen the **Labels** and **Values** columns so that I can see the longest text. In the **Label** column header, drag the right hand column separator sideways to see the full text of the variable labels:


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Values	Missing	Type	Width	Decimals	Columns
1	vpsu	Scale	Variance primary sampling unit	{-1, IAP}...	-1	Numeric	8	0	16
2	vstrat	Scale	Variance stratum	{-1, IAP}...	-1	Numeric	8	0	16
3	mar1	Nominal	MARITAL STATUS OF 1ST PERSON	{0, IAP}...	0, 8, 9	Numeric	1	0	8
4	mar2	Nominal	MARITAL STATUS OF 2ND PERSON	{0, IAP}...	0, 8, 9	Numeric	1	0	8
5	mar3	Nominal	MARITAL STATUS OF 3RD PERSON	{0, IAP}...	0, 8, 9	Numeric	1	0	8

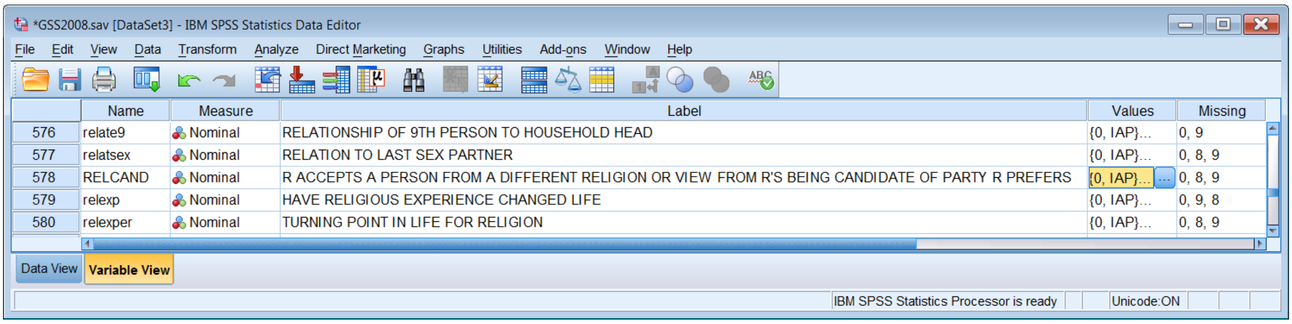
You'll need to scroll down to find the longest label, then widen the column even more:


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Value
577	relatsex	Nominal	RELATION TO LAST SEX PARTNER	{0, IAP}...
578	RELCAND	Nominal	R ACCEPTS A PERSON FROM A DIFFERENT RELIGION OR VIEW FROM R'S BEING CANDIDATE OF PARTY R PREFERS	{0, IAP}...
579	relexp	Nominal	HAVE RELIGIOUS EXPERIENCE CHANGED LIFE	{0, IAP}...
580	relexper	Nominal	TURNING POINT IN LIFE FOR RELIGION	{0, IAP}...
581	RELEXT1	Nominal	RELIGIOUS EXTREMISTS SHOULD BE ALLOWED TO HOLD PUBLIC MEETING TO EXPRESS VIEWS	{0, IAP}...

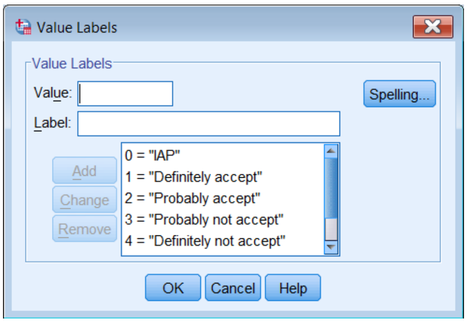
As you can see, it barely fits on the screen: that's a good reason for you to **keep labels short!**

In many surveys you then need to widen the Data Editor and do the same for **Values**, but in this file all you can see in the **Values** column is the label for the lowest missing value, 0 = IAP. If you widen the Data Editor to include the **Values** and **Missing** columns, then in the **Values** column click the box for variable **RELCAND**:


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

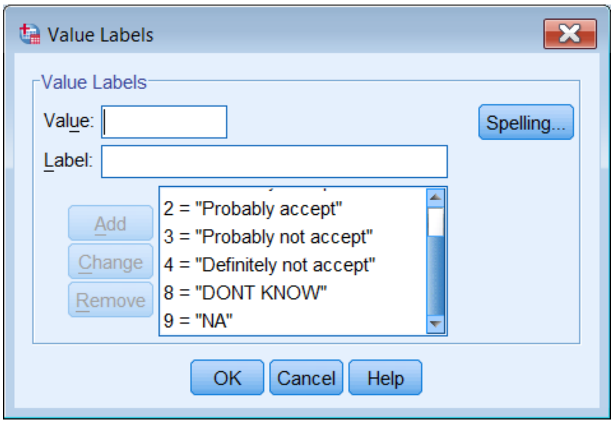
	Name	Measure	Label	Values	Missing
576	relate9	Nominal	RELATIONSHIP OF 9TH PERSON TO HOUSEHOLD HEAD	{0, IAP}...	0, 9
577	relatsex	Nominal	RELATION TO LAST SEX PARTNER	{0, IAP}...	0, 8, 9
578	RELCAND	Nominal	R ACCEPTS A PERSON FROM A DIFFERENT RELIGION OR VIEW FROM R'S BEING CANDIDATE OF PARTY R PREFERS	{0, IAP}... ...	0, 8, 9
579	relexp	Nominal	HAVE RELIGIOUS EXPERIENCE CHANGED LIFE	{0, IAP}...	0, 9, 8
580	relexper	Nominal	TURNING POINT IN LIFE FOR RELIGION	{0, IAP}...	0, 8, 9

click on (blue square)

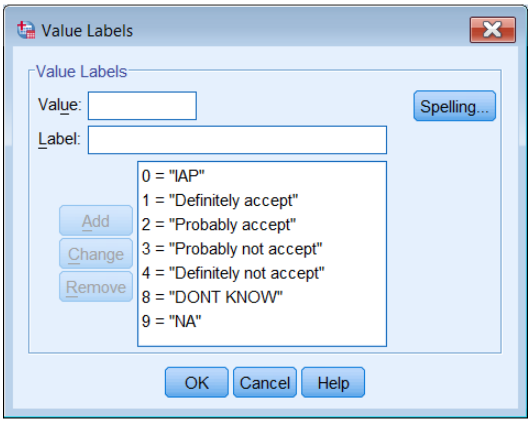


0 = "IAP"
1 = "Definitely accept"
2 = "Probably accept"
3 = "Probably not accept"
4 = "Definitely not accept"

To see the other labels you can either scroll down or drag the lower edge of the window down:

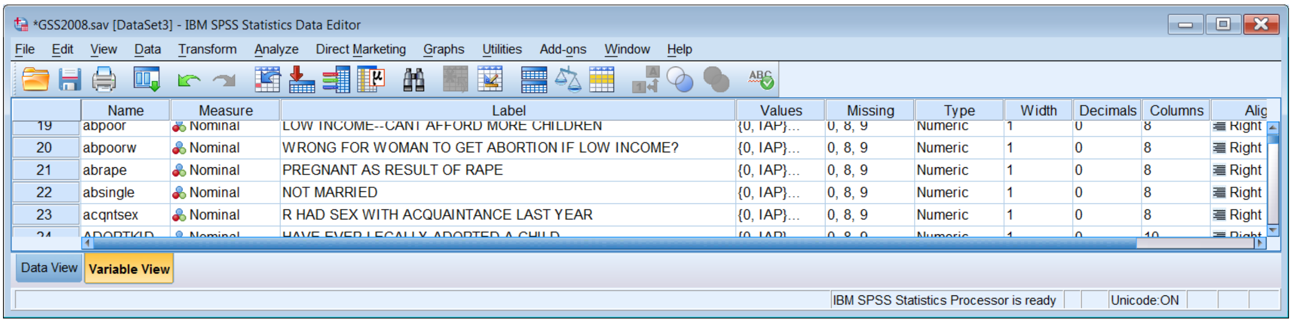


2 = "Probably accept"
3 = "Probably not accept"
4 = "Definitely not accept"
8 = "DONT KNOW"
9 = "NA"



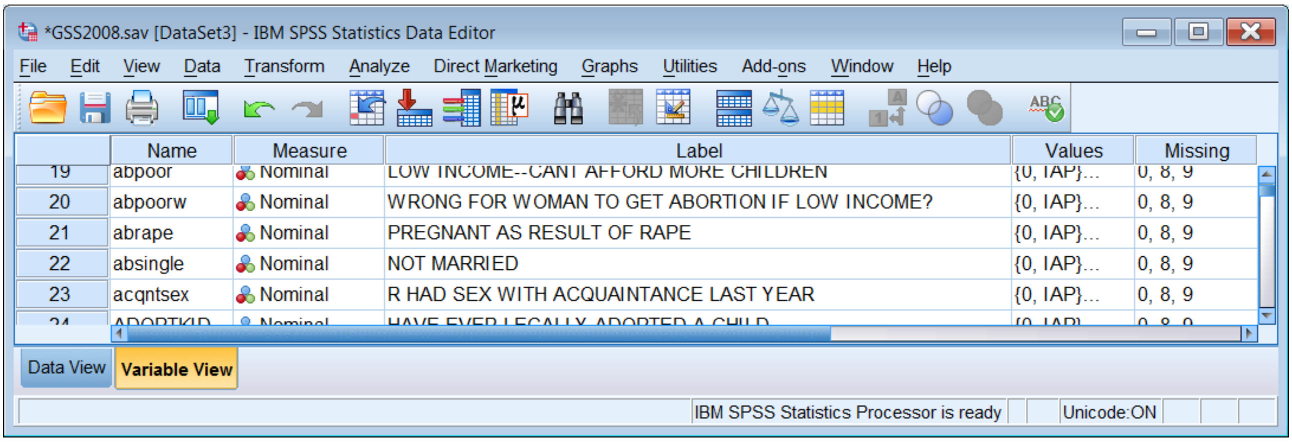
0 = "IAP"
1 = "Definitely accept"
2 = "Probably accept"
3 = "Probably not accept"
4 = "Definitely not accept"
8 = "DONT KNOW"
9 = "NA"

It's a bit silly to keep the **Labels** column as wide as this so reduce it to something which will accommodate most labels:


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	Name	Measure	Label	Values	Missing	Type	Width	Decimals	Columns	Alig
19	abpoor	Nominal	LOW INCOME--CANT AFFORD MORE CHILDREN	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right
20	abpoorw	Nominal	WRONG FOR WOMAN TO GET ABORTION IF LOW INCOME?	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right
21	abrape	Nominal	PREGNANT AS RESULT OF RAPE	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right
22	absingle	Nominal	NOT MARRIED	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right
23	acqntsex	Nominal	R HAD SEX WITH ACQUAINTANCE LAST YEAR	{0, IAP}...	0, 8, 9	Numeric	1	0	8	Right
24	ADOPTKID	Nominal	HAVE EVER LEGALLY ADOPTED A CHILD	{0, IAP}...	0, 8, 9	Numeric	1	0	10	Right

and reduce the **Data Editor** window as you don't really need the other columns (yet).


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

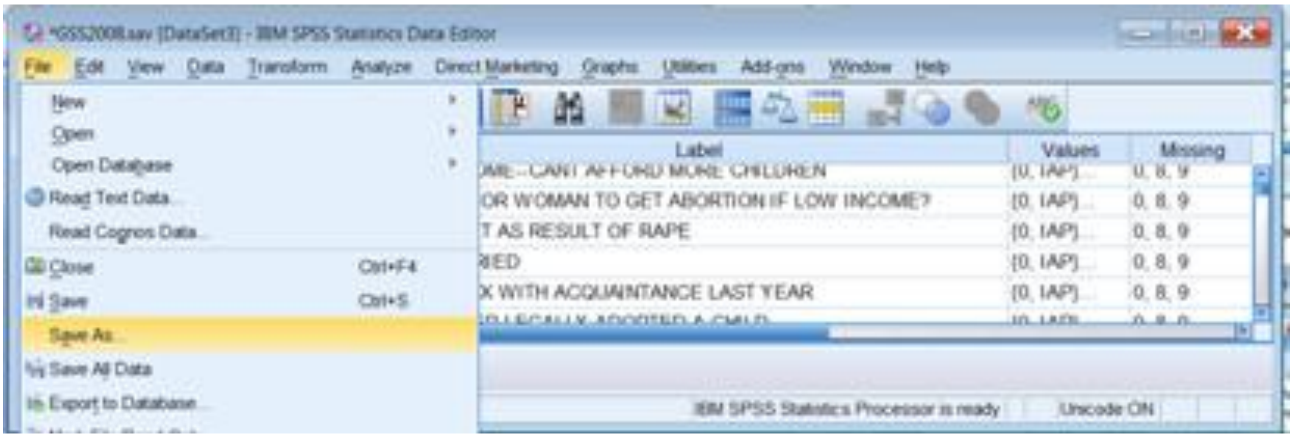
	Name	Measure	Label	Values	Missing
19	abpoor	Nominal	LOW INCOME--CANT AFFORD MORE CHILDREN	{0, IAP}...	0, 8, 9
20	abpoorw	Nominal	WRONG FOR WOMAN TO GET ABORTION IF LOW INCOME?	{0, IAP}...	0, 8, 9
21	abrape	Nominal	PREGNANT AS RESULT OF RAPE	{0, IAP}...	0, 8, 9
22	absingle	Nominal	NOT MARRIED	{0, IAP}...	0, 8, 9
23	acqntsex	Nominal	R HAD SEX WITH ACQUAINTANCE LAST YEAR	{0, IAP}...	0, 8, 9
24	ADOPTKID	Nominal	HAVE EVER LEGALLY ADOPTED A CHILD	{0, IAP}...	0, 8, 9

At this point I need to save the working file, but with a different name. In fact I should have made a copy of the file and worked on that. Throughout this session I have been breaking my golden rule:

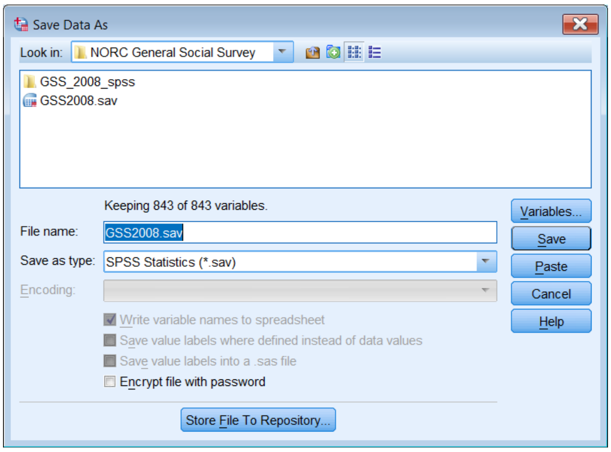
Never make alterations to an original file! Always make a copy and work on that.

In this case the original file is always available from the Roper Center, but best practice for beginners, especially if they are processing their own survey, is to make the original a read-only file and always work on a copy. I usually append my initials to the file name:

File > Save as


*GSS2008.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
New
Open
Open Database
Read Text Data...
Read Cognos Data...
Close Ctrl+F4
Save Ctrl+S
Save As...
Save All Data
Export to Database...
IBM SPSS Statistics Processor is ready
Unicode:ON

Change GSS2008 to gss2008jfh and click on Save



Look in: NORC General Social Survey

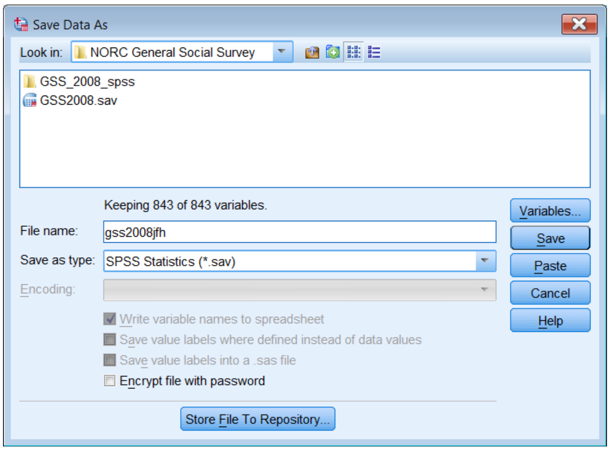
Keeping 843 of 843 variables.

File name: GSS2008.sav

Save as type: SPSS Statistics (*.sav)

Encoding:

☑ Write variable names to spreadsheet
☐ Save value labels where defined instead of data values
☐ Save value labels into a .sas file
☐ Encrypt file with password



Look in: NORC General Social Survey

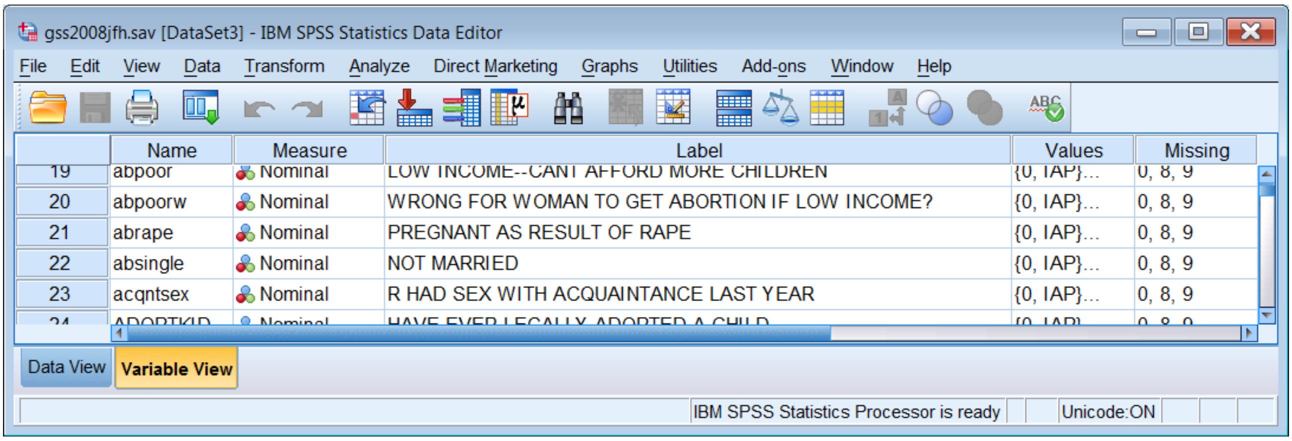
Keeping 843 of 843 variables.

File name: gss2008jfh

Save as type: SPSS Statistics (*.sav)

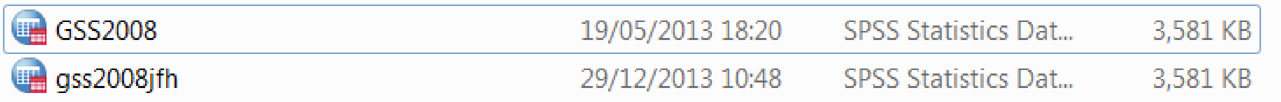
Encoding:

☑ Write variable names to spreadsheet
☐ Save value labels where defined instead of data values
☐ Save value labels into a .sas file
☐ Encrypt file with password



	Name	Measure	Label	Values	Missing
19	abpoor	Nominal	LOW INCOME--CANT AFFORD MORE CHILDREN	{0, IAP}...	0, 8, 9
20	abpoorw	Nominal	WRONG FOR WOMAN TO GET ABORTION IF LOW INCOME?	{0, IAP}...	0, 8, 9
21	abrape	Nominal	PREGNANT AS RESULT OF RAPE	{0, IAP}...	0, 8, 9
22	absingle	Nominal	NOT MARRIED	{0, IAP}...	0, 8, 9
23	acqntsex	Nominal	R HAD SEX WITH ACQUAINTANCE LAST YEAR	{0, IAP}...	0, 8, 9
24	ADOPTKID	Nominal	HAVE EVER LEGALLY ADOPTED A CHILD	{0, IAP}	0, 8, 9

The original **GSS2008.sav** is still there and unaltered: I am now working on the new edition:



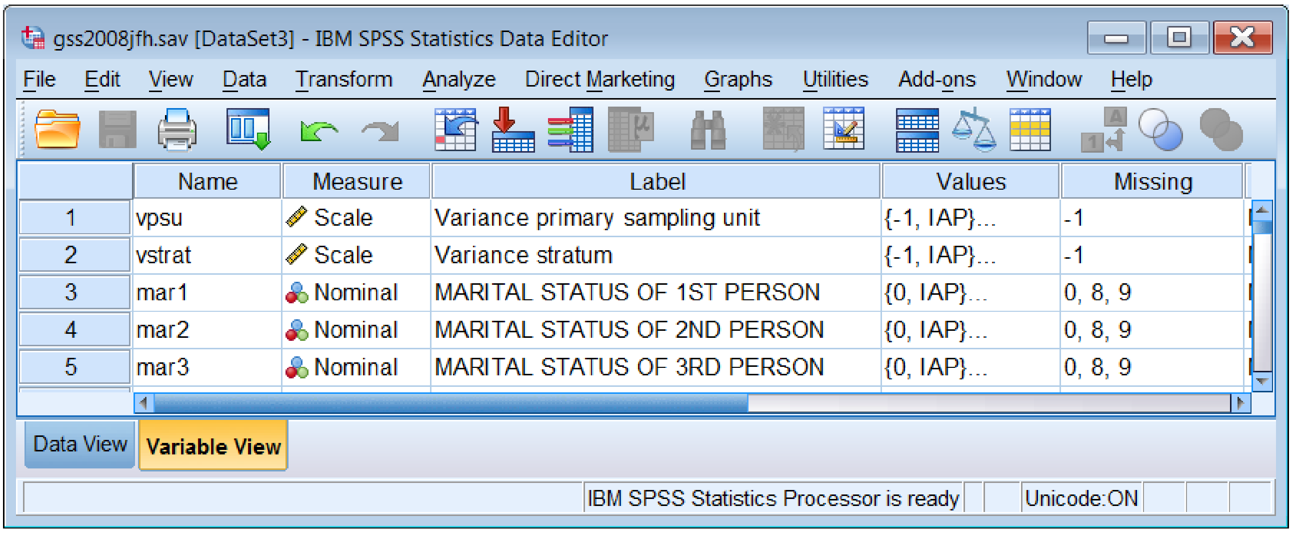
GSS2008	19/05/2013 18:20	SPSS Statistics Dat...	3,581 KB
gss2008jfh	29/12/2013 10:48	SPSS Statistics Dat...	3,581 KB

It's probably a good idea for me to save this commentary file as well, especially given the current inclement weather and a susceptibility in rural Normandy to power cuts! So now I've also been breaking another golden rule:

When writing anything in Word or SPSS syntax . .
Save your work every 10 lines or so!

Ctrl + S

Making labels more aesthetic

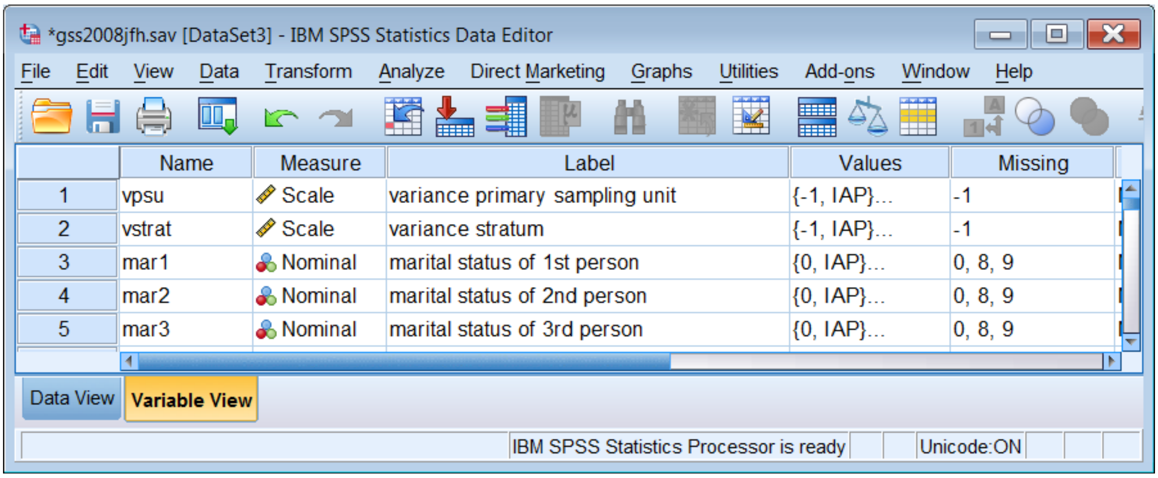


	Name	Measure	Label	Values	Missing
1	vpsu	Scale	Variance primary sampling unit	{-1, IAP}...	-1
2	vstrat	Scale	Variance stratum	{-1, IAP}...	-1
3	mar1	Nominal	MARITAL STATUS OF 1ST PERSON	{0, IAP}...	0, 8, 9
4	mar2	Nominal	MARITAL STATUS OF 2ND PERSON	{0, IAP}...	0, 8, 9
5	mar3	Nominal	MARITAL STATUS OF 3RD PERSON	{0, IAP}...	0, 8, 9

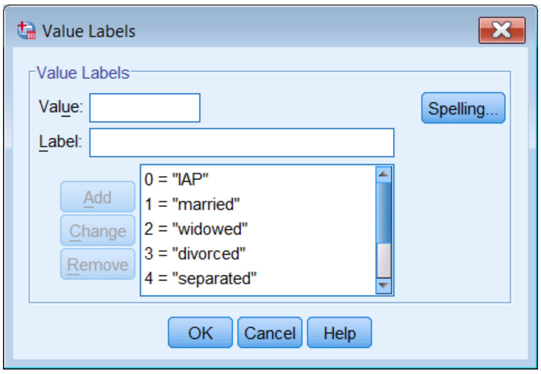
The variable and value labels are pretty ugly in UPPER CASE, but they can be changed to Mixed case using Python code supplied by Jon Peck (senior Software engineer, IBM/SPSS) to perform similar operations on other data sets. Our first effort is run in two stages, followed by a Ctrl + H substitution: slightly ungainly, but it works. The code needs to be refined to avoid the need for the second chunk of Python.

The following code can be run from inside the SPSS syntax editor. It changes all alphabetic characters to lower case except the value labels IAP, DK and NA for values declared as missing.

```
/*Changes all letters to lower case.
begin program.
import spss, spssaux
vardict = spssaux.VariableDict()
for var in vardict:
    var.VariableLabel = var.VariableLabel.lower()
    vallabels = var.ValueLabels
    for k,v in vallabels.items():
        if not v in ['IAP', 'DK', 'NA']:
            vallabels[k] = v.lower()
    var.ValueLabels = vallabels
end program.
```



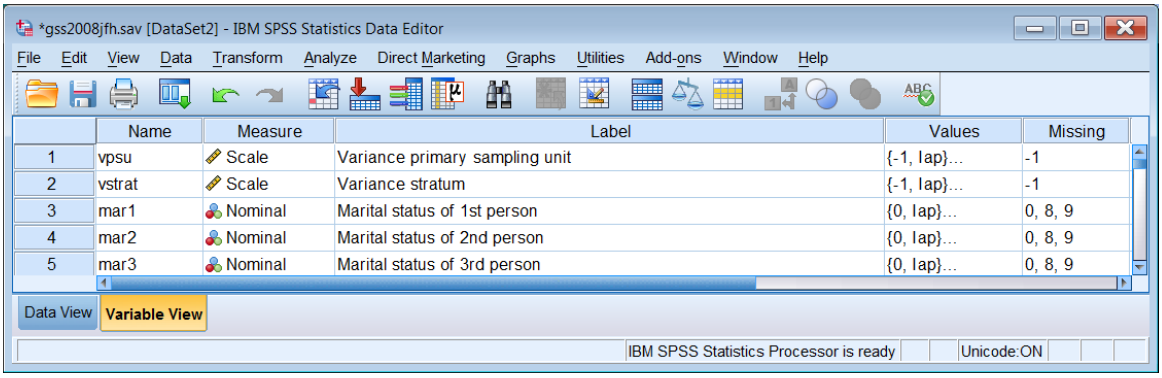
	Name	Measure	Label	Values	Missing
1	vpsu	Scale	variance primary sampling unit	{-1, IAP}...	-1
2	vstrat	Scale	variance stratum	{-1, IAP}...	-1
3	mar1	Nominal	marital status of 1st person	{0, IAP}...	0, 8, 9
4	mar2	Nominal	marital status of 2nd person	{0, IAP}...	0, 8, 9
5	mar3	Nominal	marital status of 3rd person	{0, IAP}...	0, 8, 9



0 = "IAP"
1 = "married"
2 = "widowed"
3 = "divorced"
4 = "separated"

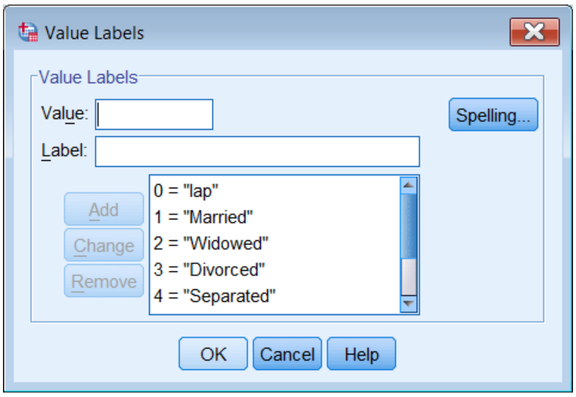
This next chunk changes all the first letters of each label back to UPPER CASE, but for some reason also changes IAP, DK and NA to Iap, Dk and Na

```
/*Restores first letter of first word to upper case
begin program.
import spss, spssaux
vd = spssaux.VariableDict()
for v in vd:
    varlabel = v.VariableLabel
    if varlabel:
        v.VariableLabel = varlabel.capitalize()
    vallbls = v.ValueLabels
    for k in vallbls:
        vallbls[k] = vallbls[k].capitalize()
    if vallbls:
        v.ValueLabels = vallbls
end program.
```


*gss2008jfh.sav [DataSet2] - 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	vpsu	Scale	Variance primary sampling unit	{-1, Iap}...	-1
2	vstrat	Scale	Variance stratum	{-1, Iap}...	-1
3	mar1	Nominal	Marital status of 1st person	{0, Iap}...	0, 8, 9
4	mar2	Nominal	Marital status of 2nd person	{0, Iap}...	0, 8, 9
5	mar3	Nominal	Marital status of 3rd person	{0, Iap}...	0, 8, 9

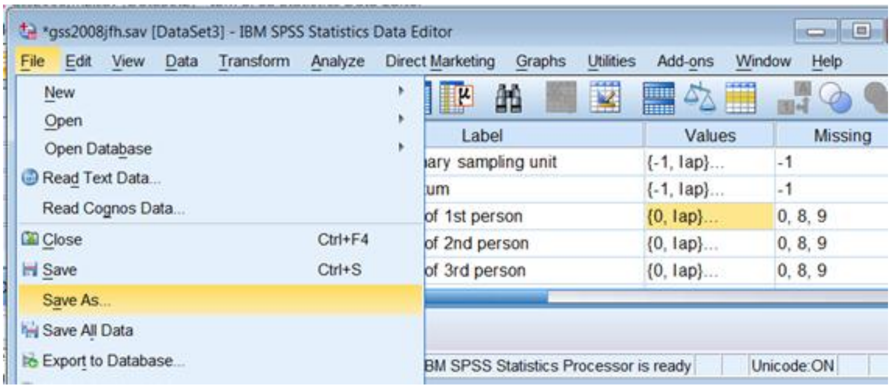
IBM SPSS Statistics Processor is ready | Unicode:ON



0 = "Iap"
1 = "Married"
2 = "Widowed"
3 = "Divorced"
4 = "Separated"

With a bit of tweaking the Python could be made to retain the labels IAP, DK and NA, but I don't know how to do that, so there's another way. First I need to save my work, so:

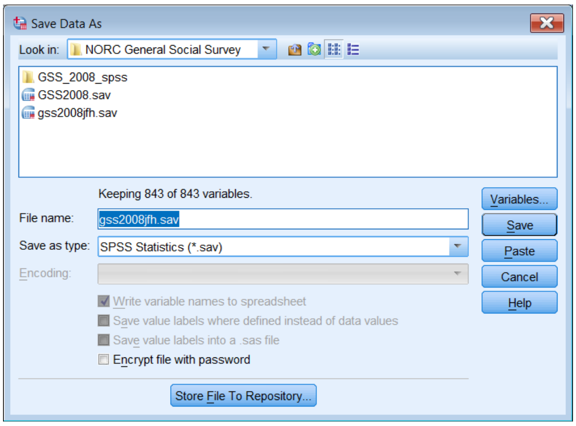
File > Save As


*gss2008jfh.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
New
Open
Open Database
Read Text Data...
Read Cognos Data...
Close Ctrl+F4
Save Ctrl+S
Save As...
Save All Data
Export to Database...

Label	Values	Missing
ary sampling unit	{-1, Iap}...	-1
um	{-1, Iap}...	-1
of 1st person	{0, Iap}...	0, 8, 9
of 2nd person	{0, Iap}...	0, 8, 9
of 3rd person	{0, Iap}...	0, 8, 9

BM SPSS Statistics Processor is ready | Unicode:ON

Change gss2008jfh.sav to gss2008jfh2.sav



Look in: NORC General Social Survey

GSS_2008_spss
GSS2008.sav
gss2008jfh.sav

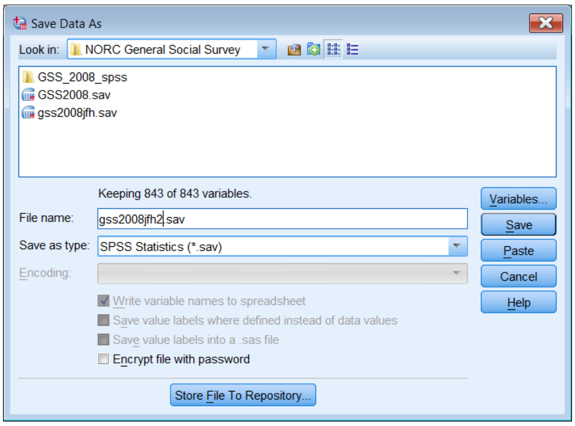
Keeping 843 of 843 variables.

File name: gss2008jfh.sav

Save as type: SPSS Statistics (*.sav)

Encoding:

☑ Write variable names to spreadsheet
☐ Save value labels where defined instead of data values
☐ Save value labels into a .sas file
☐ Encrypt file with password



Look in: NORC General Social Survey

GSS_2008_spss
GSS2008.sav
gss2008jfh.sav

Keeping 843 of 843 variables.

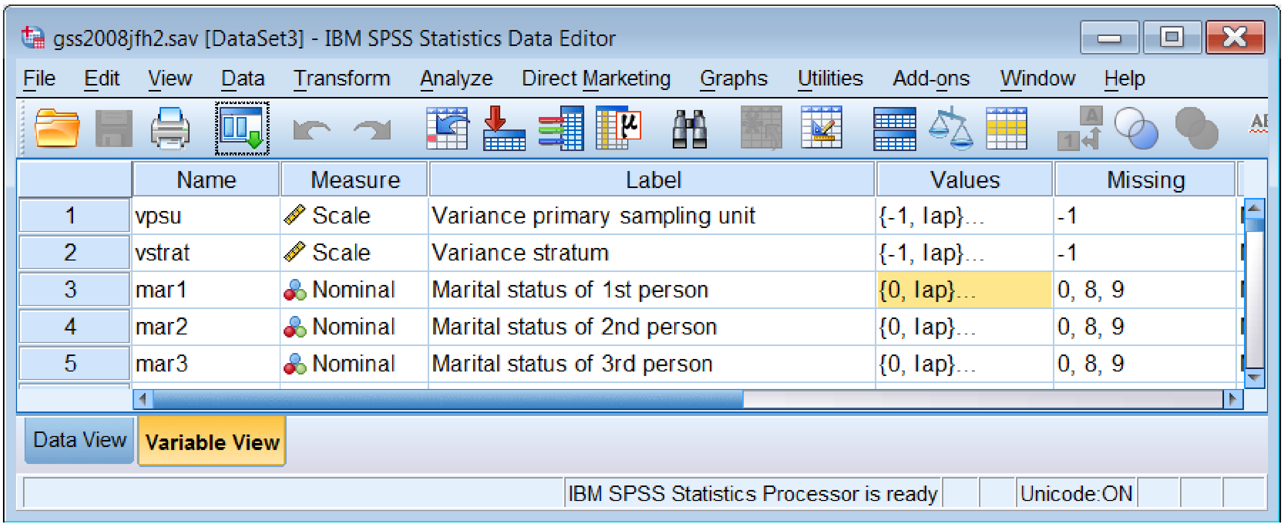
File name: gss2008jfh2.sav

Save as type: SPSS Statistics (*.sav)

Encoding:

☑ Write variable names to spreadsheet
☐ Save value labels where defined instead of data values
☐ Save value labels into a .sas file
☐ Encrypt file with password

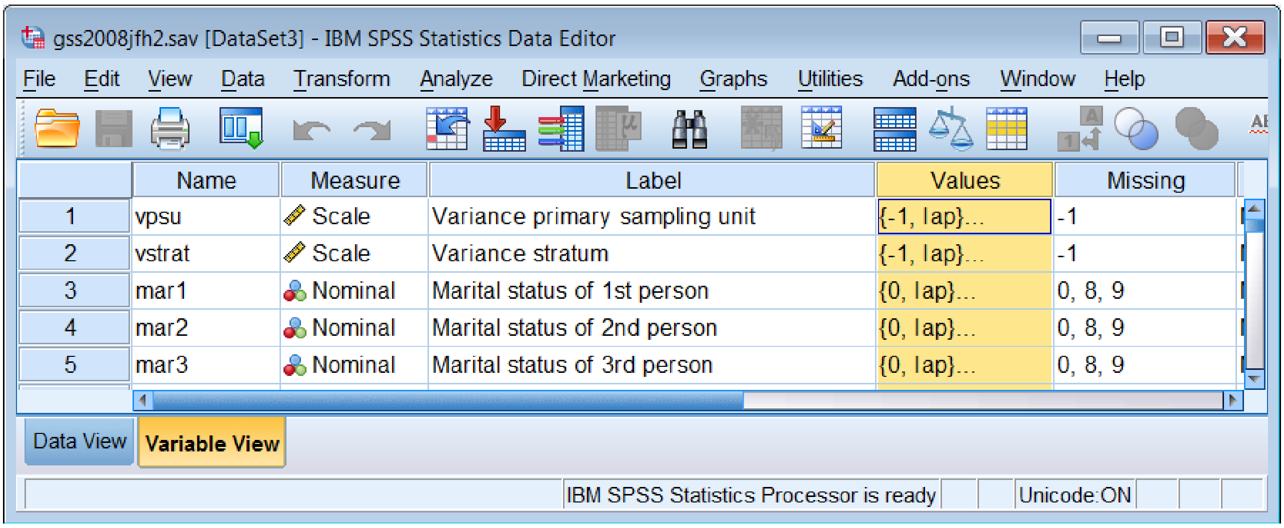
. . and press Save


gss2008jfh2.sav [DataSet3] - 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	vpsu	Scale	Variance primary sampling unit	{-1, Iap}...	-1
2	vstrat	Scale	Variance stratum	{-1, Iap}...	-1
3	mar1	Nominal	Marital status of 1st person	{0, Iap}...	0, 8, 9
4	mar2	Nominal	Marital status of 2nd person	{0, Iap}...	0, 8, 9
5	mar3	Nominal	Marital status of 3rd person	{0, Iap}...	0, 8, 9

Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

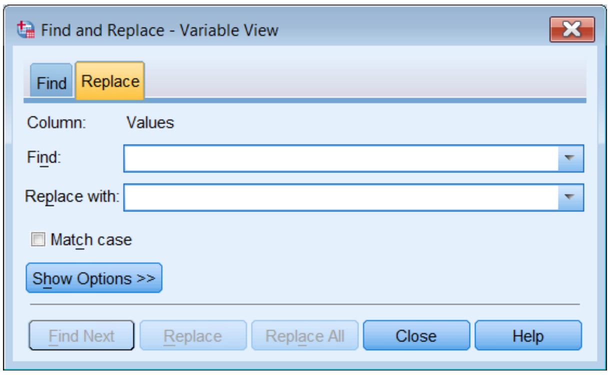
Left click on **Values** to highlight the column:


gss2008jfh2.sav [DataSet3] - 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	vpsu	Scale	Variance primary sampling unit	{-1, Iap}...	-1
2	vstrat	Scale	Variance stratum	{-1, Iap}...	-1
3	mar1	Nominal	Marital status of 1st person	{0, Iap}...	0, 8, 9
4	mar2	Nominal	Marital status of 2nd person	{0, Iap}...	0, 8, 9
5	mar3	Nominal	Marital status of 3rd person	{0, Iap}...	0, 8, 9

Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

Then use Ctrl + H (find and replace) to change Iap, Dk and Na back to UPPER CASE.

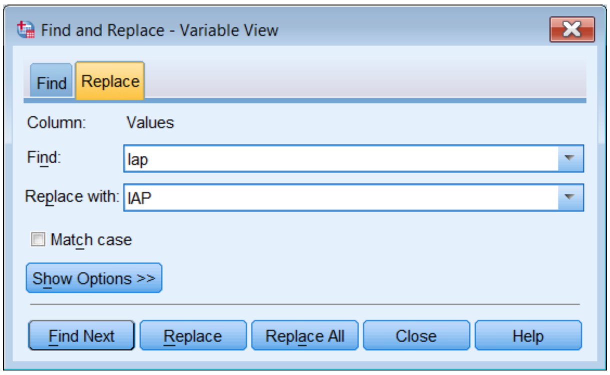


Column: Values

Find:

Replace with:

□ Match case

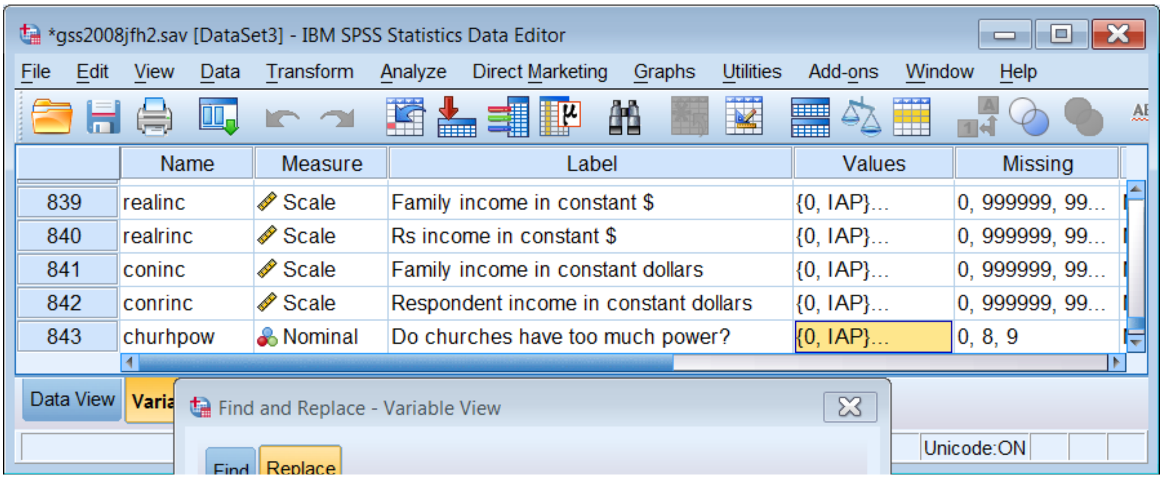


Column: Values

Find: Iap

Replace with: IAP

□ Match case

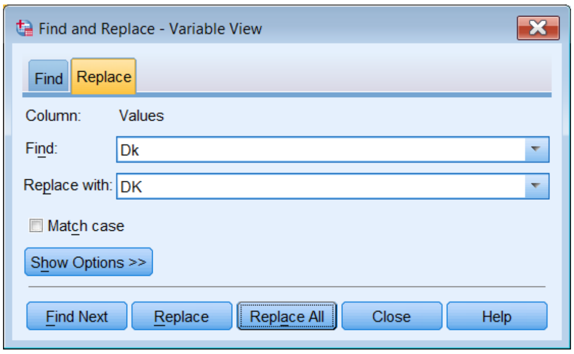


*gss2008jfh2.sav [DataSet3] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Varia
Unicode:ON

	Name	Measure	Label	Values	Missing
839	realinc	Scale	Family income in constant $	{0, IAP}...	0, 999999, 99...
840	realrinc	Scale	Rs income in constant $	{0, IAP}...	0, 999999, 99...
841	coninc	Scale	Family income in constant dollars	{0, IAP}...	0, 999999, 99...
842	conrinc	Scale	Respondent income in constant dollars	{0, IAP}...	0, 999999, 99...
843	churhpow	Nominal	Do churches have too much power?	{0, IAP}...	0, 8, 9

Find and Replace - Variable View
Find
Replace

Actually this seems to work if only one cell is highlighted in the **Values** column:

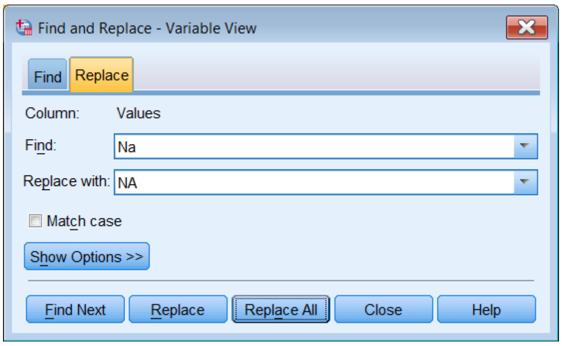


Column: Values

Find: Dk

Replace with: DK

□ Match case

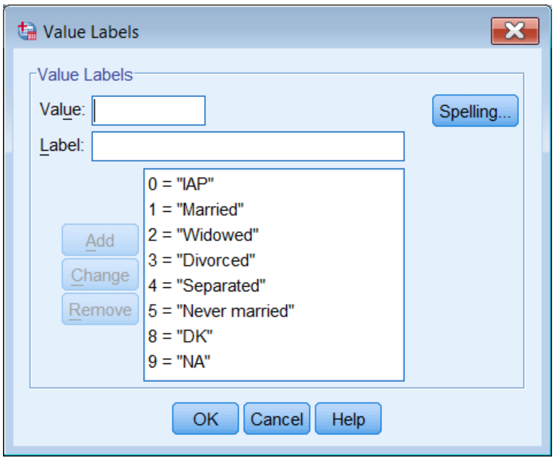


Column: Values

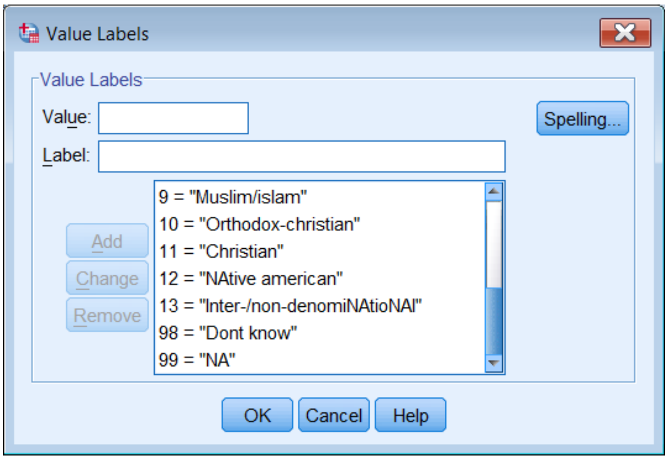
Find: Na

Replace with: NA

□ Match case



0 = "IAP"
1 = "Married"
2 = "Widowed"
3 = "Divorced"
4 = "Separated"
5 = "Never married"
8 = "DK"
9 = "NA"



9 = "Muslim/islam"
10 = "Orthodox-christian"
11 = "Christian"
12 = "NAtive american"
13 = "Inter-/non-denomiNAtioNAl"
98 = "Dont know"
99 = "NA"

. . but there's always the danger of changing similar strings elsewhere in the file (eg NAtive and NAtioNAl above)!! Religion and other variables may need to be corrected by hand, since some words need initial upper case letters (eg Christian, American and Islam) so it may be safer to write syntax with VALUE LABELS based on the entries in the **Missing** column.

However, all this would be much easier in Python code, so I asked Jon Peck if he can merge the two sets of code. He duly obliged with the following code:

```
* This program capitalizes all variable and value labels and
replaces words listed in the uppercase variable with their
all-caps equivalent in variable and value labels.
```

```
begin program.
import spss, spssaux, re
# terms to capitalize as free-standing words - list in uppper case
# or whatever case variant is desired
uppercase = ["IAP", "DK", "NA"]

vd = spssaux.VariableDict()
for v in vd:
   varlabel = v.VariableLabel
   if varlabel:
      varlabel = varlabel.capitalize()
      for w in uppercase:
         sexpr = r"(?i)\b%s\b" % w
         varlabel = re.sub(sexpr, w, varlabel)
      v.VariableLabel = varlabel
   vallbls = v.ValueLabels
   for k in vallbls:
      vallbls[k] = vallbls[k].capitalize()
      for w in uppercase:
         sexpr = r"(?i)\b%s\b" % w
         vallbls[k] = re.sub(sexpr, w, vallbls[k])
   if vallbls:
      v.ValueLabels = vallbls
end program.
```

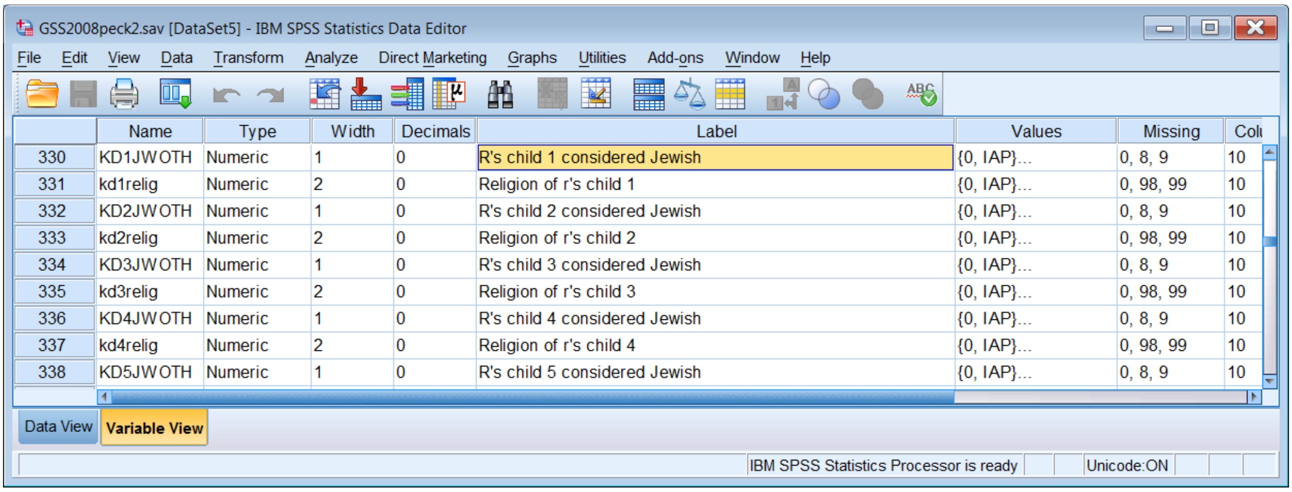
As well as the code he sent instructions for amending it to include words to be left in their original state, such as acronyms, countries, denominations etc. which I can add myself, eg:

```
uppercase = ["IAP", "DK", "NA", "Christian", "Jew", "Jewish", "America", "Islam"]
```

. . and which I can extend over more lines provided I break after a comma, eg:

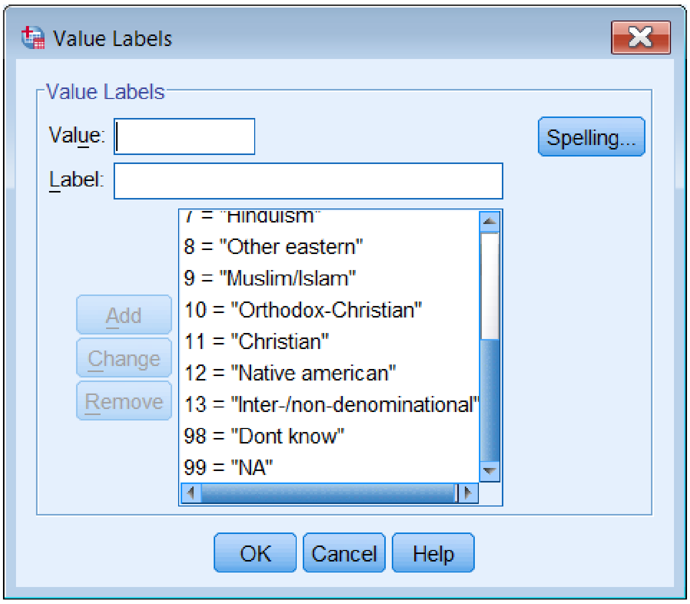
```
uppercase = ["IAP", "DK", "NA", "Christian", "Jew",
 "Jewish", "America", "Islam"]
```

This what the file looks like so far:



	Name	Type	Width	Decimals	Label	Values	Missing	Colu
330	KD1JWOTH	Numeric	1	0	R's child 1 considered Jewish	{0, IAP}...	0, 8, 9	10
331	kd1relig	Numeric	2	0	Religion of r's child 1	{0, IAP}...	0, 98, 99	10
332	KD2JWOTH	Numeric	1	0	R's child 2 considered Jewish	{0, IAP}...	0, 8, 9	10
333	kd2relig	Numeric	2	0	Religion of r's child 2	{0, IAP}...	0, 98, 99	10
334	KD3JWOTH	Numeric	1	0	R's child 3 considered Jewish	{0, IAP}...	0, 8, 9	10
335	kd3relig	Numeric	2	0	Religion of r's child 3	{0, IAP}...	0, 98, 99	10
336	KD4JWOTH	Numeric	1	0	R's child 4 considered Jewish	{0, IAP}...	0, 8, 9	10
337	kd4relig	Numeric	2	0	Religion of r's child 4	{0, IAP}...	0, 98, 99	10
338	KD5JWOTH	Numeric	1	0	R's child 5 considered Jewish	{0, IAP}...	0, 8, 9	10

As you can see "Jewish" has been left as was, both in the variable labels above and the value labels below (in which "Islam" and "Christian" are also preserved:



7 = "Hinduism"
8 = "Other eastern"
9 = "Muslim/Islam"
10 = "Orthodox-Christian"
11 = "Christian"
12 = "Native american"
13 = "Inter-/non-denominational"
98 = "Dont know"
99 = "NA"

This leaves me to comb through the file looking for terms like "American" (left as "american" in the above screenshot) names of states, "USA", "United Sates" and acronyms for organisations. There are quite a few labels like these to be picked up:

conjudge	Confid. in united states supreme court
conlegis	Confidence in congress

This seems like a lot of work, but it only has to be done once and it's great fun watching the task bar as SPSS hurtles through the file making the requested changes.

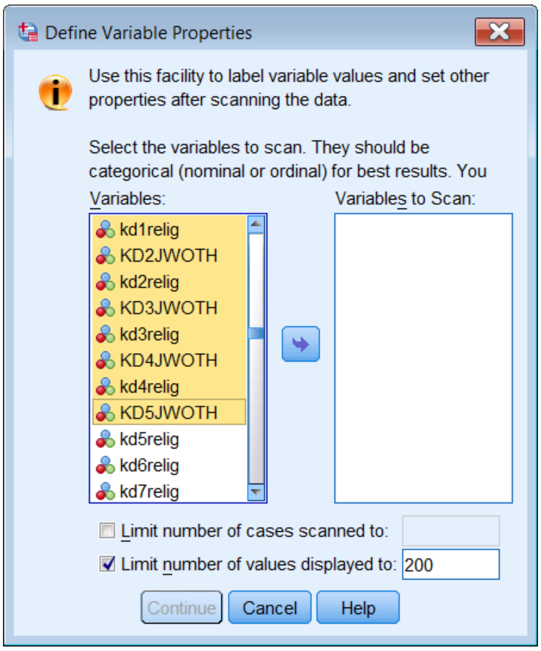
I then ran into a serious problem. The distributed file has all variables apart from the first 13 in alphabetical order. Checking the SPSS settings to set them to File Order does not change the sequence. For teaching purposes (and to some extent for secondary analysis) I need to find a questionnaire and then have the variables in questionnaire order. This has proved virtually impossible as the only questionnaires produced by NORC are in CAPI format. Refusing to admit defeat I went to the SDA site (Berkeley) indicated on the GSS website and discovered a wealth of facilities for generating raw data files and SPSS syntax files for every GSS wave up to 2010. From this I extracted *.txt files for the raw data and (automatically generated) SPSS setup for 2008. Minor modification to the setup file enabled me to create a new version of the 2008 SPSS saved file, but with many fewer variables than the NORC version, and in a completely different order.

The sheer speed of all this was amazing, but I need to do more exploration to get the variables I want in the order I want them, so that it tallies with the original questionnaire order. Generating a "proper" questionnaire from the CAPI *.pdf file is no trivial task and may yet prove fruitless, which rather defeats the point, at least for teaching.

However, it may now be possible to generate and scroll through syntax files generated by SDA, picking out acronyms and other obvious words then copying them into the Python code, adding the double primes. Probably not: for variable labels it's quicker in the **Data Editor** in Variable View to scroll up and down the **Label** column and either change the text or make a note of words to include in the Python code. For value labels it's quicker to use:

Data > Define Variable Properties

and search all the labels in batches.



Define Variable Properties

Use this facility to label variable values and set other properties after scanning the data.

Select the variables to scan. They should be categorical (nominal or ordinal) for best results. You

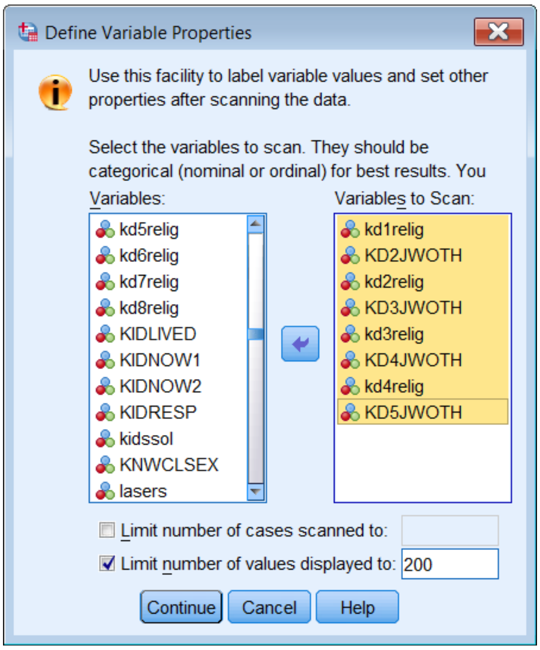
Variables:

kd1relig
KD2JWOTH
kd2relig
KD3JWOTH
kd3relig
KD4JWOTH
kd4relig
KD5JWOTH
kd5relig
kd6relig
kd7relig

Variables to Scan:

☐ Limit number of cases scanned to:

☑ Limit number of values displayed to: 200



Define Variable Properties

Use this facility to label variable values and set other properties after scanning the data.

Select the variables to scan. They should be categorical (nominal or ordinal) for best results. You

Variables:

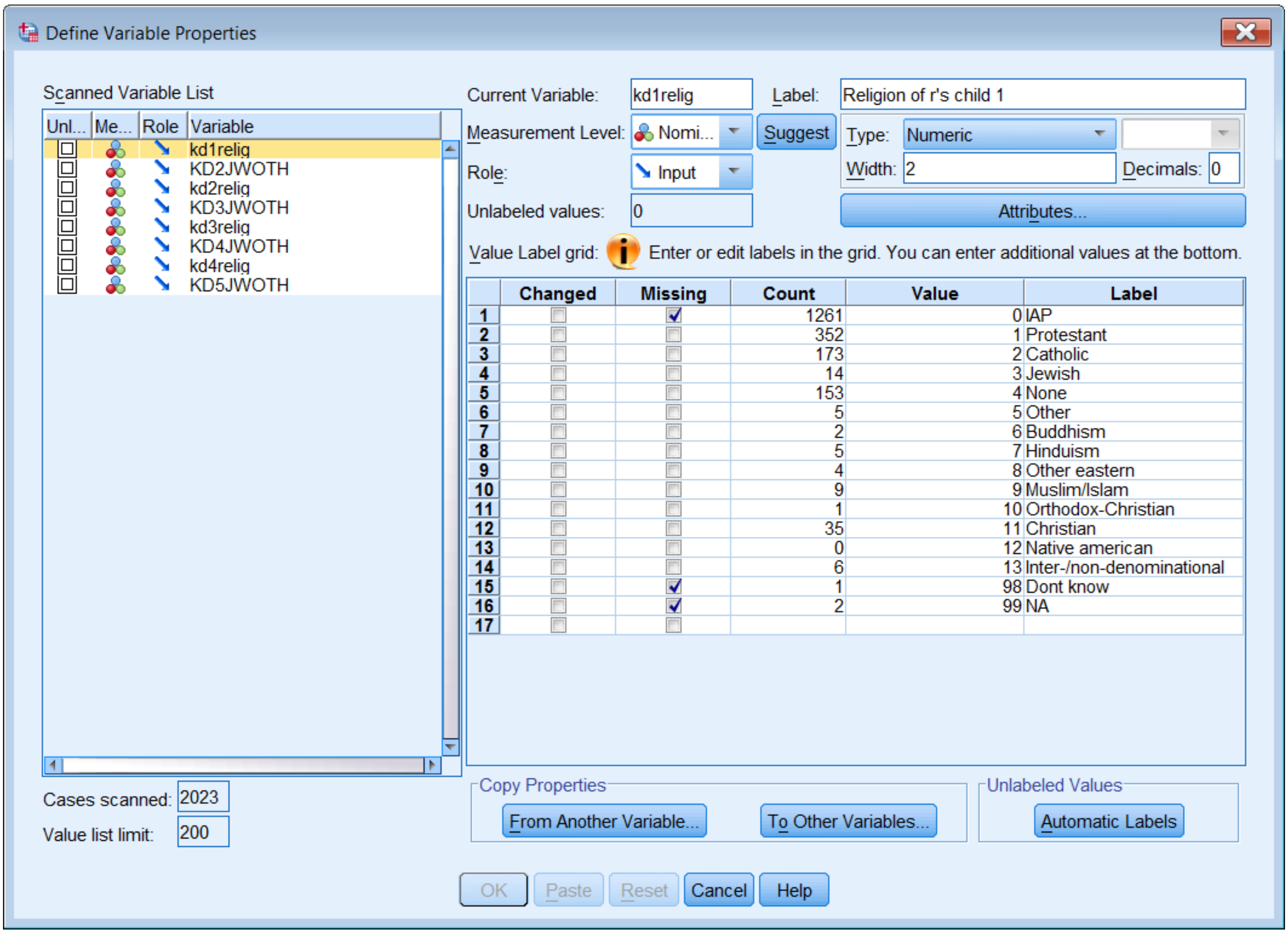
kd5relig
kd6relig
kd7relig
kd8relig
KIDLIVED
KIDNOW1
KIDNOW2
KIDRESP
kidssol
KNWCLSEX
lasers

Variables to Scan:

kd1relig
KD2JWOTH
kd2relig
KD3JWOTH
kd3relig
KD4JWOTH
kd4relig
KD5JWOTH

☐ Limit number of cases scanned to:

☑ Limit number of values displayed to: 200



Define Variable Properties

Scanned Variable List

Unl...	Me...	Role	Variable
☐			kd1relig
☐			KD2JWOTH
☐			kd2relig
☐			KD3JWOTH
☐			kd3relig
☐			KD4JWOTH
☐			kd4relig
☐			KD5JWOTH

Cases scanned: 2023

Value list limit: 200

Current Variable: kd1relig

Label: Religion of r's child 1

Measurement Level: Nomi...

Type: Numeric

Role: Input

Width: 2

Decimals: 0

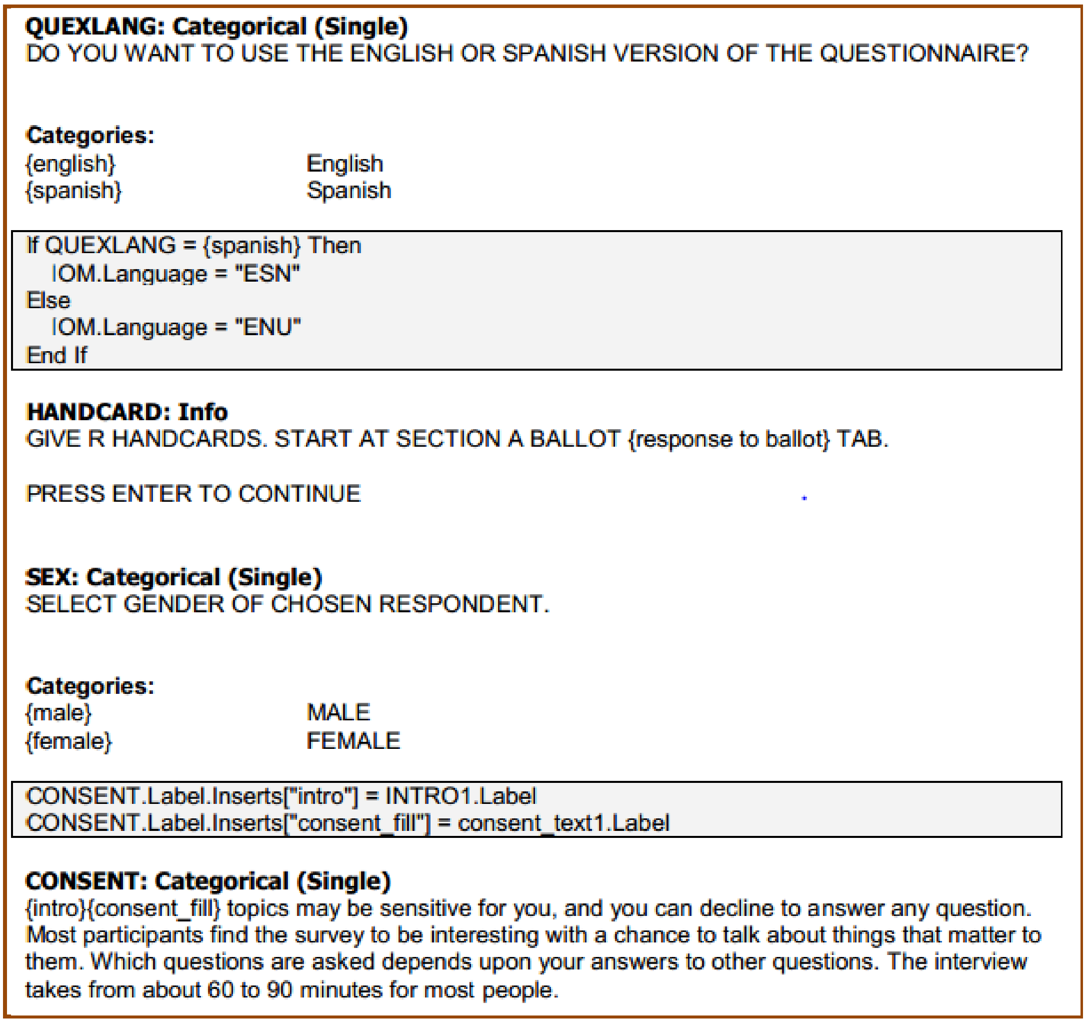
Unlabeled values: 0

Value Label grid: Enter or edit labels in the grid. You can enter additional values at the bottom.

	Changed	Missing	Count	Value	Label
1	☐	☑	1261	0	IAP
2	☐	☐	352	1	Protestant
3	☐	☐	173	2	Catholic
4	☐	☐	14	3	Jewish
5	☐	☐	153	4	None
6	☐	☐	5	5	Other
7	☐	☐	2	6	Buddhism
8	☐	☐	5	7	Hinduism
9	☐	☐	4	8	Other eastern
10	☐	☐	9	9	Muslim/Islam
11	☐	☐	1	10	Orthodox-Christian
12	☐	☐	35	11	Christian
13	☐	☐	0	12	Native american
14	☐	☐	6	13	Inter-/non-denominational
15	☐	☑	1	98	Dont know
16	☐	☑	2	99	NA
17	☐	☐			

Back to the questionnaire problem. Tom Smith of NORC tells me quite bluntly, " The GSS questionnaire is CAPI and never exists in a text format like what you are referring to." This leaves me to deal with a 247 page pdf file which is searchable, but which is not dynamic: you can't click on a variable or a page number, you have to scroll up and down looking for what you need. Not exactly motivating for beginners!

Users are therefore left with the CAPI version with pages like:



QUEXLANG: Categorical (Single)
DO YOU WANT TO USE THE ENGLISH OR SPANISH VERSION OF THE QUESTIONNAIRE?

Categories:

{english}	English
{spanish}	Spanish

```
If QUEXLANG = {spanish} Then
   IOM.Language = "ESN"
Else
   IOM.Language = "ENU"
End If
```

HANDCARD: Info
GIVE R HANDCARDS. START AT SECTION A BALLOT {response to ballot} TAB.

PRESS ENTER TO CONTINUE

SEX: Categorical (Single)
SELECT GENDER OF CHOSEN RESPONDENT.

Categories:

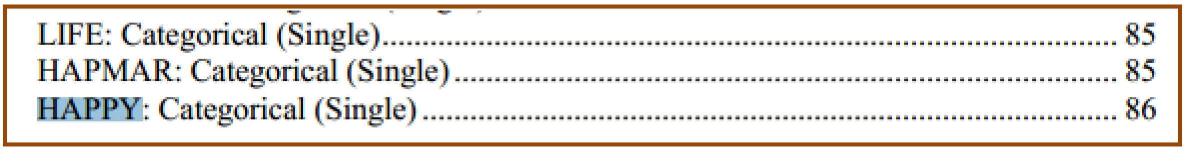
{male}	MALE
{female}	FEMALE

```
CONSENT.Label.Inserts["intro"] = INTRO1.Label
CONSENT.Label.Inserts["consent_fill"] = consent_text1.Label
```

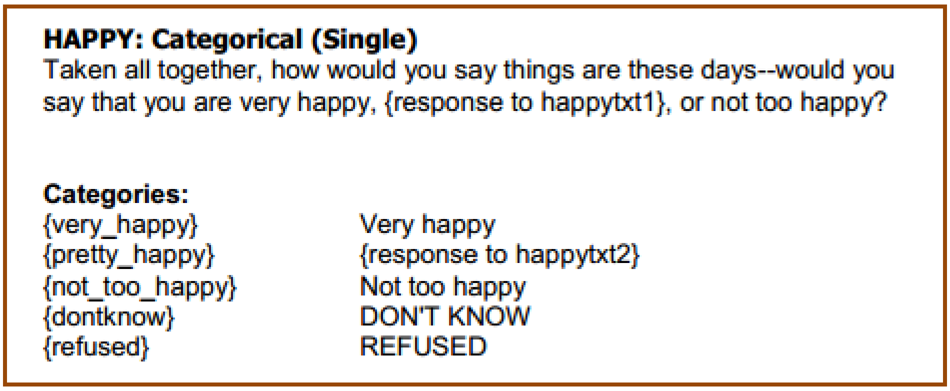
CONSENT: Categorical (Single)
{intro}{consent_fill} topics may be sensitive for you, and you can decline to answer any question. Most participants find the survey to be interesting with a chance to talk about things that matter to them. Which questions are asked depends upon your answers to other questions. The interview takes from about 60 to 90 minutes for most people.

As an example, take a question that has been asked in one form or another for decades in GSS and other surveys:

> Taking all things together, how would you say you are these days? Would you say you are very happy, pretty happy or not too happy these days?



Scrolling to page 86 finds:

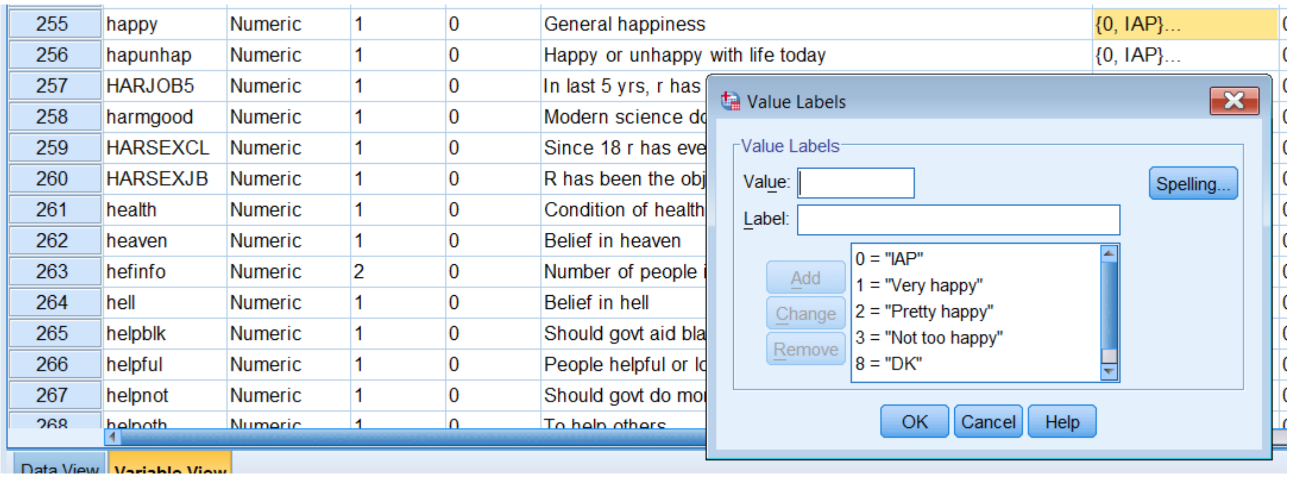


HAPPY: Categorical (Single)
Taken all together, how would you say things are these days--would you say that you are very happy, {response to happytxt1}, or not too happy?

Categories:

{very_happy}	Very happy
{pretty_happy}	{response to happytxt2}
{not_too_happy}	Not too happy
{dontknow}	DON'T KNOW
{refused}	REFUSED

In the SPSS file the variable names and labels tally, but the variables are not in the same order:



255	happy	Numeric	1	0	General happiness	{0, IAP}...
256	hapunhap	Numeric	1	0	Happy or unhappy with life today	{0, IAP}...
257	HARJOB5	Numeric	1	0	In last 5 yrs, r has	
258	harmgood	Numeric	1	0	Modern science do	
259	HARSEXCL	Numeric	1	0	Since 18 r has eve	
260	HARSEXJB	Numeric	1	0	R has been the obj	
261	health	Numeric	1	0	Condition of health	
262	heaven	Numeric	1	0	Belief in heaven	
263	hefinfo	Numeric	2	0	Number of people i	
264	hell	Numeric	1	0	Belief in hell	
265	helpblk	Numeric	1	0	Should govt aid bla	
266	helpful	Numeric	1	0	People helpful or lo	
267	helpnot	Numeric	1	0	Should govt do mor	
268	helpoth	Numeric	1	0	To help others	

Value Labels

Value Labels

Value:

Label:

0 = "IAP"
1 = "Very happy"
2 = "Pretty happy"
3 = "Not too happy"
8 = "DK"

This leaves users with three problems: first to find the variables they want, second to find the original question wording, third to find the order in which the questions were asked.

For tutors and instructors, there's a fourth problem: how to generate a "pseudo-questionnaire" with the questions in the original order, the full text of questions as asked and a saved SPSS file which tallies with these. To do this for the full survey is beyond my resources of time and goodwill, but probably worth doing for sub-sets of variables used in my recommended textbooks or which I find interesting to use as examples, either because they reflect my interest in subjective social indicators, or because they enable interesting methodological exercises.

We shall see.