ASSESS 2006

Old Dog, Old Tricks

Using SPSS syntax to avoid the mouse trap

John F Hall[1]

(photo JFH 25 Dec 2005)

1: Introduction

I was (and at heart still am) a survey researcher. From 1970 until 1992 (when I took early retirement) this was at senior level. I specialised in advisory, design and collaborative survey work, getting value for money, and in research rescue jobs. According to the late Dr Mark Abrams, I “moved easily and widely in the world of social research”, developing subjective social indicators (“Quality of Life”) and conducting surveys (for clients or on research grants). From 1973 until 1992, and even after retirement, I provided expert advice on, and training in, the management and analysis of social survey data using SPSS.

From 1972 to 1993 I used SPSS on an almost daily basis (exclusively in syntax mode and on a range of computers (CDC 2000, ICL1900 and 2900 series, Dec10, Dec20 and finally a Vax cluster) to process and analyse dozens of surveys, either for clients or as part of professional and academic research programmes. In 1974 I organised the international conference at LSE, **Social Science Data and the New SPSS** (the final design exercise preparing for a new release) and in 1978 set up and chaired the **UK SPSS Users’ Group** (UKSUG, a predecessor of ASSESS) and edited its newsletter. I have trained or advised hundreds of researchers and students to use SPSS.

This presentation draws on my experiences of using SPSS, not only on surveys I have worked on, but also from the hands-on postgraduate course, **Survey Analysis Workshop**, (part-time, evening) which I developed and taught at the then Polytechnic of North London (PNL) from 1976 to 1992. Most students, at least in the early days, came with little or no previous experience of statistics or computers: many of them could not even type!. I have lived in France since 1994, but maintain an active interest in the development of, and training in, social research methods in the UK, and in the **Mark Abrams Prize**[2], awarded annually by the Social Research Association for the best piece of work linking survey research, social theory and/or social policy.

In 2001 the Social Research Association, in return for a 300 word review, offered a copy of Julie Pallant’s **SPSS Survival Manual**[3] which, motivated in equal measure by curiosity, vanity and greed, I duly requested. Apart from a short consultancy in 2000 for the Institute of Employment Studies[4], I had been out of serious SPSS action since 1993. All my previous experience of SPSS had been on mainframes in batch mode (with some interactive on the Vax) using VMS and EDT: my PC experience was limited to DOS, VPPlanner and WordStar4 (keyboard only, no mouse). I had barely got the hang of using a mouse for email with Outlook Express and had never used Word, Windows or SPSS for Windows. Imagine my absolute horror (and panic!) when I opened the Pallant book and found it was about SPSS for Windows using nothing but drop-down menus and point-and-click. (It also had practically nothing on data checking, data entry, manipulation or tabulation, but that’s another story.) My first ever search of the web was for **SPSS for Windows** tutorials!.

[1] Thankfully retired: previously Lecturer, University of Birmingham, 1968-70, Senior Research Fellow, SSRC Survey Unit, 1970-76; Principal Lecturer in Sociology and Director, Survey Research Unit, Polytechnic of North London, 1976-92.

[2] www.the-sra.org.uk/mark_abrams.htm

[3] (1st edition, Open University Press, 2001: there is now a 2nd edition, 2005, with new material on log-linear regression)

[4] Enormous sparse matrix in SPSS saved file which should have been supplied by the fieldwork agency as a hierarchical file. No mainframe, just SPSS for Windows on powerful PC and a very bright young researcher, Siobhán O’Regan. No facility for recreating original data file in hierarchical format, so resorted to complex job involving repeated multiple response (me syntax, she mouse). Nearly drove us mad, but sorted after 2 days and report duly published. See Huws U and O’Regan S ***EWork in Europe: The EMERGENCE 18-Country Employer Survey*** Report No 380 Inst for Employment Studies 2001

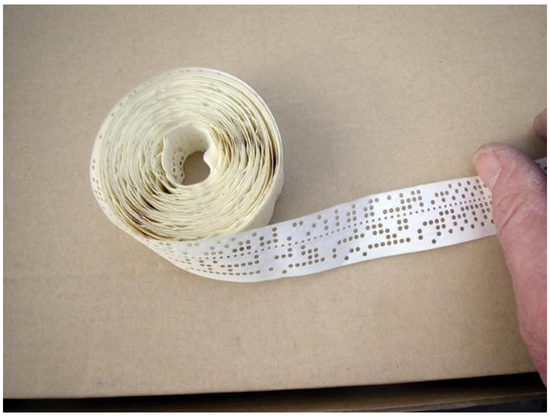
Through the web, I located a number of university staff teaching SPSS as well as dozens of course outlines and several down-loadable tutorials, the most useful of which was from computer services at Bogazici University in Turkey[5]. To help me get started, Jane Fielding (University of Surrey) kindly sent me her entire course notes (and had to explain how to open a blank syntax file!). To help with the review, SPSS France provided an evaluation version of **SPSS11 for Windows**. After much frustration (and a very steep and rapid learning curve) my first version of the Pallant review ran to 3,500 words, more than 10 times the required size, but was eventually whittled down to around 1,700 and was published in SRA News[6] in November 2002.

On the strength of this review (plus the extensive additional comments[7] trimmed from the original) and with an undertaking not to use SPSS for personal gain, I was awarded a 5-year licence to the full version. This has allowed me not only to restore several of my early surveys as SPSS portable files, but also to convert and update extensive training materials from my PNL courses for SPSS for Windows.

In return, all such training materials have been, or will be, made available to SPSS Inc.: they will also be made freely available to members of ASSESS or any other bona fide colleagues, teachers and students using SPSS. Some are already posted on the Market Research Portal site[8], but this site has problems with tabular formats, colour text and graphics and an alternative site[9] is currently under development for the more complex offerings.

1.1: Before SPSS

Between 1965 and 1968, when I was a young interviewer/researcher at Salford University, I contributed to a series of one-off computer programs[10] for data entry and tabulation, some of which were later generalised[11] for multiple sets of tables. They were written in Algol for an English Electric KDF9[12] computer (all of 64K!! RAM) which occupied a room the size of a small terraced house and took anything up to 40 minutes c.p.u. time to produce a single table: all data and programs had to be punched on, and all output produced on, and read from, 8-hole paper tape. You couldn't see what you were typing (but you got very good at reading the hole-patterns!. To get printout you had to put both program and output tapes through a special tape-reader.



8-hole paper tape

Paper tape punch

[5] Page no longer available, but recommended similar tutorials are: http://www.indiana.edu/~statmath/support/bydoc/ (getting started) and http://www.utexas.edu/cc/stat/tutorials/ (statistics analysis and more advanced help topics). A useful hub for links to other SPSS users, tutorials and topics is http://www.spsstools.net/

[6] www.the-**sra**.org.uk/documents/pdfs/**sra**_nov_2002.pdf (SRA News: see pages 10-11)

[7] yet to be posted: copies are available from the author on johnfhall@orange.fr

[8] http://www.marketresearchworld.net/index.php?option=com_content&task=category&sectionid=15&id=74&Itemid=60

[9] www.aronique.com (click button to enter, then on **web design**, then JH at bottom right)

[10] by John Whittaker and John Hall 1965-68, initially one table at a time. Conforming to the 3 E's (Economy, Efficiency, Elegance) writing in Algol was, to the author, just like writing in Latin prose: wonderful.

[11] See: Hall J F **A General Data-processing Package** BSA Maths and Computing Applications Group, University of East Anglia, April 1968

[12] Legend has it that the KDF9 was developed as project KD9 (Kidsgrove Development 9) and that the 'F' in its designation was contributed by the then Chairman. After a long and tedious discussion on what to name the machine at launch, he declared, "I don't care if you call it the F.......".

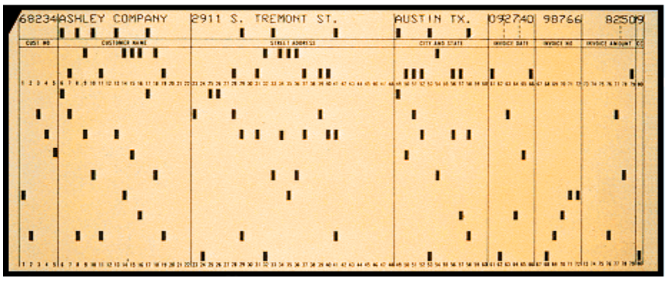
Basically these programs were for tabulation only, but with some extra programs for specific tasks (e.g. error checking, data transformation, percentages, chi-square calculation at both table and cell level), but little else apart from table titles and no row or column labelling at all. At the University of Birmingham[13] I modified them for the departmental PDP11 (everything still on paper tape). At Salford, they survived as the **Salford Survey Suite** until 1979 when I advised maintenance be discontinued in view of the widespread availability of SPSS. The experience of writing this suite gave me a unique programming insight into the handling of arrays and other computing processes which was later invaluable in assessing the capabilities and facilities (or lack thereof) in other software. It also gave me a lifelong belief in the value of cultivating good working relations with computer staff.

In 1966, the only other piece of technology available was a manual typewriter with a limited character set, but at least it was portable and had a wide carriage to take foolscap paper sideways. In those pre-Tippex days, all corrections and amendments to questionnaires and reports had to be retyped and copies made using carbon paper. This machine still exists, but was last used in February 1973 (by me) to type the questionnaire for a "quickie" survey of attitudes of senior pupils[14] in a girls' public school.

Manual (portable) typewriter

80-column Hollerith card

From Computer Desktop Encyclopedia
© 2000 The Computer Language Co. Inc.



Elsewhere, data from questionnaire surveys were typically transferred to 80-column Hollerith cards and then counted on card-sorters or tabulators or processed by computer using proprietary software (eg Donovan Data Systems) or by software packages for statistical analysis (most of which were written in Fortran on 80-column cards) which were complex and difficult to use, especially by social scientists, even if there was a manual. Boxes of cards were very heavy and cards could easily be damaged by repeated processing or by being dropped.

IBM 026 Card punch

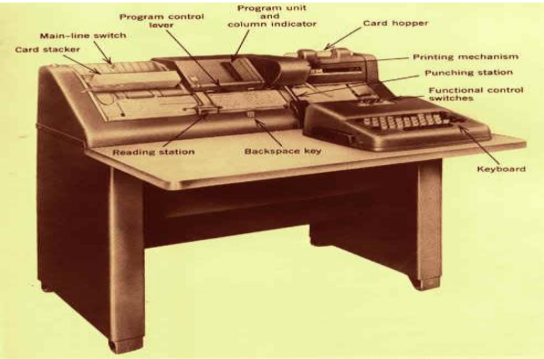
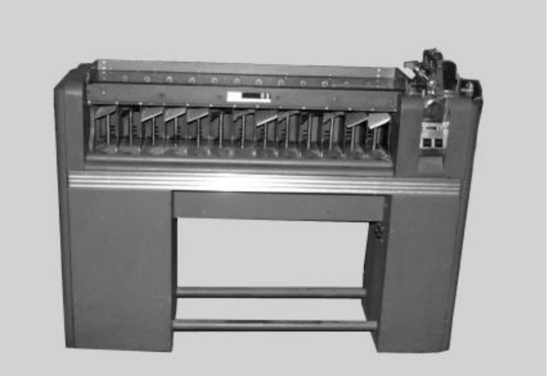


FIGURE 4.11 IBM 026 Keypunch (Courtesy IBM).

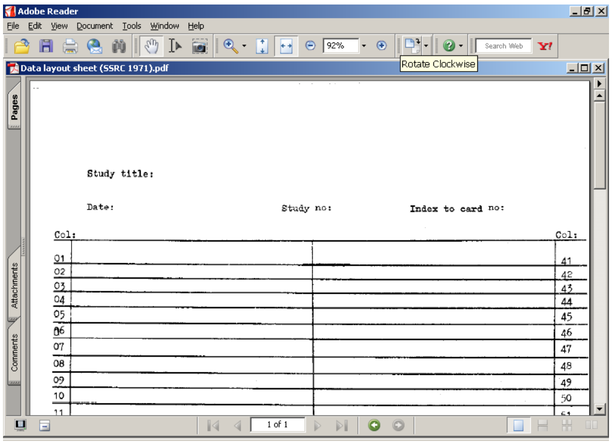
Card-sorter



[13] Lecturer in Social Studies, Dept of Transportation and Environment Planning, University of Birmingham, 1968-70.

[14] See: Mark Abrams and John Hall **Attitudes and Opinions of Girls in Senior Forms**, SSRC Survey Unit, March 1973 (mimeo 20pp) ; Sarah Abrams, **The Survey**, short article (5pp) in *Folio*, St. Paul's School magazine 1973

You also needed some kind of printed guide to data layout, such as the blank sheets used by the SSRC Survey Unit, indicating which data were stored in each of the 80 columns.

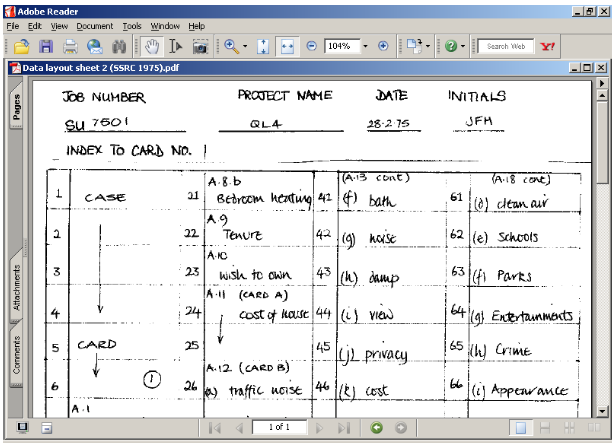

Adobe Reader
File Edit View Document Tools Window Help
92%
Search Web
Rotate Clockwise
Data layout sheet (SSRC 1971).pdf
Pages
Attachments
Comments
1 of 1

Study title:

Date: Study no: Index to card no:

Col:			Col:
01			41
02			42
03			43
04			44
05			45
06			46
07			47
08			48
09			49
10			50
11			

SSRC Survey Unit 1971 (2 x 40)


Adobe Reader
File Edit View Document Tools Window Help
104%
Search Web
Data layout sheet 2 (SSRC 1975).pdf
Pages
Attachments
Comments
1 of 1

JOB NUMBER SU 7501 PROJECT NAME QL4 DATE 28·2·75 INITIALS JFH

INDEX TO CARD NO. 1

1	CASE	21	A.8.b Bedroom heating	41	(A.13 cont) (f) bath	61	(A.18 cont) (d) clean air
2	↓	22	A.9 Tenure	42	(g) noise	62	(e) Schools
3		23	A.10 wish to own	43	(h) damp	63	(f) Parks
4	↓	24	A.11 (CARD A) cost of house	44	(i) view	64	(g) Entertainments
5	CARD	25	↓	45	(j) privacy	65	(h) Crime
6	↓ ①	26	A.12 (CARD B) (a) traffic noise	46	(k) cost	66	(i) Appearance
	A.1						

SSRC Survey Unit 1975 (4 x 20)

This early technology placed severe limits on the design and conduct of survey research, although it made for high-powered thinking and very careful work! That's why, in card-based systems, much early statistics was developed on 2 x 2 tables. It also affected what was possible, in the time available, for the provision of practical training and courses.

1.2: The origins of SPSS

SPSS first appeared in 1968, and was written in Fortran IV for an IBM 360 by three postgraduate students[15]. It came from Chicago to Edinburgh via Tony Coxon and was implemented in 1970[16] at Edinburgh Regional Computer Centre (one of the few places with an IBM) by David Muxworthy and the late Marjorie Barritt, thereby scotching university plans to commission a survey processing facility at great expense from scratch. When first installed it was reputedly called more times than the Fortran compiler, but that could well have because of all the user-errors!

Tony Coxon

David Muxworthy

Conversions to ICL followed later, but those with CDC and DEC machines got SPSS sooner.

[15] Norman Nie, Dale Bent and Tex Hull: See Appendix 1 for full details

[16] See Appendix 1 for correspondence with Tony Coxon and David Muxworthy plus relevant committee minutes.

In next to no time SPSS was everywhere. Why? Because it had an easily understood manual[17] you could buy in any university bookshop, was relatively straightforward to use (with English-language like instructions), had amazing file manipulation facilities, could print labels for variables and values, and was implemented UK-wide in many, if not most, academic and local government environments so that users moving between sectors could rapidly adjust with only minor adjustments to job control language (JCL) for different machines. It became so successful and made so much money that it threatened the charitable status of the National Opinion Research Center (NORC) at the University of Chicago and was hived off as SPSS Inc.

David Muxworthy has kindly provided the following account[18]:

> *Norman Nie and Dale Bent were political science postgrads at Stanford in the late 1960s and, fed up with the 'put a 1 in column 72' type command language of the programs at the time, they devised a language that a political scientist would want to write to specify an analysis. They scraped together some funds and hired Tex Hull to help with coding the program, which was in Fortran IV for the 360…. Norman and Tex both moved to Chicago, Norman to the National Opinion Research Center, Tex to the Computing Center. Dale went back to Alberta and, apart from having his name on some of the manuals, dropped out of SPSS.*
>
> *People got to hear about the program, which was superior in user interface to much that was available at the time, and requested copies. This led to Patrick Bova, a librarian at NORC, being hired 25% of his time to act as distribution agent and Karin Steinbrenner being hired as full time programmer. When I visited Chicago in the summer of 1972 this was the total staff. I thought I was going to a large software house. It was surprising to find it not much bigger than a one man and a dog in a bedroom outfit (at that time at least). Tex acted largely as advisor but was busy as associate director of the computing center.*
>
> *As I remember it, Jean Jenkins was hired as programmer later in 1972 or in 1973. She was certainly around at the SCSS planning meetings in the summer of 1973. The program was so successful that NORC became wary of losing their non-profit status and strongly encouraged Norman to form a company and move out. This happened sometime between 1974 (when I worked with them at NORC) and 1977 (when they had moved to an office block in downtown Chicago).*
>
> *In Edinburgh the program grew to be so popular there were demands to move it to the ICL system 4 and later the ICL 2980, the IBM having been removed by higher authority. This led to PLU organising conversions to some other platforms in UK universities, notably the ICL 1900. SPSS themselves arranged conversions to other series, notably the CDC 6600 at Northwestern University, just up the road from Chicago.*

Amongst other things SPSS was also used for the Annual Voter Registration survey in Cheshire (it even printed the address labels), applications in concrete technology (!!) and at the Greater London Council it was faster on the Census than specially commissioned software. Central government, constrained by a requirement to buy British computers, took much longer, but eventually succumbed.

From its very first days, social scientists loved SPSS, but it was received less than enthusiastically by some programmers (inefficient coding) and some statisticians (some errors, but mostly because it let people like me bypass them and do their own analysis). Anyway, where were they all when we were struggling on our own and wanted time on their precious computers or advice on statistical procedures?

1.3: Quantitative Sociology Group

A parallel development in the late 1960's was the setting up of the British Sociological Association (BSA) **Maths and Computing Applications Group**, the inaugural colloquium of which was called by Tony Coxon et al. and held at the University of East Anglia in April 1968. Delegates to this colloquium (many now retired, or close to retirement) were later to figure prominently in computing and statistical applications in the social sciences in the UK. The group later became the **BSA Mathematics, Statistics and Computing Applications Group** and eventually (including Survey Research) the **Quantitative Sociology Group**[19] which published **Quantitative Sociology Newsletter** throughout the 1970's and the

[17] Nie, Hull and Jenkins (the blue one). Listed procedures in alphabetical order, so was a bit awkward to use.

[18] See Appendix 1 for full details

[19] The group was never officially recognised by the BSA and never received any funds, hence the eventual name change.

early 1980's. I was editor 1973-74 and from 1975 until it ceased publication in 1984 for lack of material to publish. The residual funds were invested and used later to set up the Mark Abrams Prize in 1986.

.

1.4: SSRC Survey Unit (1970-1976)

In April 1970 the then Social Science Research Council (SSRC) set up its Survey Unit with the late Dr Mark Abrams as (part-time) Director. The Unit was *"attached to, but not of, the LSE*" for computing and sundry other purposes. Its brief was to offer advice and assistance in survey methods to academics and others doing surveys on public funds. I was appointed Senior Research Fellow (University Reader level: the unit's first full time post) in October 1970 and, in addition to general research and advisory duties, was given overall responsibility for Unit computing. When the computing load grew too much for one person, Jim Ring was appointed in 1972 (on graduation from his MSc in Operational Research at LSE) to be responsible for the technical side of all Unit computing. Jim contributed various specialist programs for processes and analyses which were not then available in SPSS. Indeed, we together found ways to make SPSS do things that weren't in the manual such as handling and analysing multiple response, hierarchical files and longitudinal data.

My first experience of SPSS was at the month-long SSRC Summer School on Survey Methods, held at St Edmund's College, Oxford, in 1972. I was working with the students on a live survey of "Quality of Life in Oxford" which had been designed from scratch (a real scissors and Sellotape job). The students did all their own typing, copying, printing, collating, sampling, interviewing and coding: Clive Payne (Nuffield College) arranged data preparation and analysis (with newly arrived SPSS). Some initial processing and analysis was done, but for students' own reports there were various delays and, on the final evening, when Clive, who had often worked late into the night to get jobs run, had to leave for a pressing social engagement, leaving me with the (only) SPSS manual, but no training, and no further results. I was thus left alone trying unsuccessfully for several hours to get SPSS to work until, just after midnight, I was finally ejected from the computer centre when the operator, who had been extremely patient and helpful all evening, had completed his shift and wanted to go home. (Remember this was in the days of punched cards, but at least there was a line-printer on site for all the error messages, none of which made sense to me or to the operator!) Subsequent Summer Schools ran without a problem, but for a shorter period of three weeks (at St Catherine's College, Oxford) with myself teaching SPSS, and without the live survey.

My memory is a bit vague here as to what machines (ICL1900 and 2900 series, CDC2000) were at ULCC, and when, but until SPSS arrived we used Peter Wakeford's (LSE) survey data tabulation program SDTAB for analysis and a utility program MUTOS for spreading out multi-punched data. As surveys began to come in thick and fast, and since we wouldn't be the only ones working on them, especially if they were by academics and others needing advice and assistance on their own surveys, and since there would be several of us working on the same survey at the same time, we developed various standard conventions (naming, labelling, file names and structures) for use with SPSS together with appropriate documentation.

In 1970 the University of London Computing Centre (ULCC) had a small **Survey Analysis Working Party** (SAWP) comprising Beverley Rowe (ULCC, chair), Andrew Westlake (LSHTM), myself (SSRC) and Peter Wakeford (LSE) which, with the addition of Tony Cowling (Taylor Nelson) and Nona Newman (Newcastle Univ.), became the **Study Group on Computers in Survey Analysis** (SGCSA), which in turn became the current **Association for Survey Computing**. By 1973, working to a brief from SGCSA, and (justifying the use of my time as survey methodology) I had compiled, and persuaded SSRC to publish, the first UK register of software for survey analysis[20] and also co-authored a report to the SSRC Computer Panel on their survey of computer use in social science departments in UK universities[21] (the questionnaire for which was based heavily on my previous technical and professional experiences).

Another SSRC Survey Unit initiative[22] was an exercise in SPSS error-trapping. The printouts from all failed SPSS runs at ULCC were filtered out and a questionnaire attached for users to list the reported

[20] See: Hall J F, **Computer Software for Survey Analysis**. Special Supplement to SSRC Newsletter 20, October 1973

[21] See: Utting J E G and Hall J F, **The Use of Computers in University Social Science Departments**, Occasional Paper 3, SSRC Survey Unit, 1973

[22] Suggested by the late Cathie Marsh, whom I had appointed as a trainee researcher, her first post, direct from our 1974 Summer School and halfway through her SSRC funded PhD studentship at Cambridge

and actual cause(s) of errors they had made and to check whether SPSS had correctly identified them: a short article[23] was published in the ULCC Newsletter.

1.5: Polytechnic of North London (1976-1992)

In September 1976, amid great controversy, the SSRC closed the Survey Unit and all staff were made redundant, but (in reverse order of age) all found new posts[24] before then. I ended up at the Polytechnic of North London (PNL) taking with me Jim Ring (one day a week) and a grant to develop a computer program for **Interactive Path Analysis** compatible with the **SPSS Conversational Statistical System** (SCSS). In 1978, in view of the research funds I was attracting, the Polytechnic agreed to set up a **Survey Research Unit** with myself as Unit Director and with a brief closely based on that of the old SSRC Survey Unit. Through the new Unit, I used SPSS to process and analyse dozens of surveys. On the training side I simply split the SSRC Summer School course into two part-time postgraduate evening courses, **Survey Analysis Workshop**[25] (taught initially by myself and colleagues, latterly by myself alone) followed by **Survey Research Practice** (taught entirely by senior practitioners from research institutes and agencies outside PNL). Both courses ran with great success from 1976 until 1992, but closed after I retired. There was no-one else to teach my course and, given the circumstances of my retirement, my professional colleagues outside PNL refused to continue teaching the other course.

In 1976 there were no on-site computing facilities in Ladbroke House (the base for the Faculty of Social Studies). Indeed, since it was the very first course to be held in the evenings, there were no facilities for anything else either. Each week, coffee, biscuits and plastic cups were brought in by teaching staff, a couple of kettles commandeered, and on the first evening a case of Bulgarian Cabernet Sauvignon (from the first Majestic Wine Warehouse in Tottenham) served to stimulate student exchanges at the end of the introductory session.

For the first few years, student exercises in SPSS had to be written on coding sheets and sent to PNL Computer Services in Holloway Road to be punched on 80-column Hollerith cards, then run without being checked. Results were handed out the following week. The error rate was high and resulted in frustrating losses of time and consequently of motivation. When Computer Services later supplied a surplus ICL card-punch, it was possible (for me) to correct and resubmit student jobs in time for the following week. With the installation at Ladbroke House of a computer laboratory equipped with 16 VDU terminals, 4 servers and a fast link to the mainframe on the Holloway campus, and with Jim Ring's highly user-friendly SPSS front-end (as Jim put it, **"Hall-proof = idiot-proof!"**) which made it much easier to use the Vax, edit, run, correct and resubmit SPSS setup files (and avoid exceeding disk quotas!) and to print up results, the whole course was transformed. Moreover, with the advent of personal computers, there was a remarkable improvement in the previous skills of students, particularly at undergraduate level, and this made the courses even more productive and enjoyable.

Much less time needed to be spent on basic technology and keyboard skills, and much more could be devoted to hands-on data management and analysis, with a consequent benefit to work-rates and motivation. Errors could be corrected and results checked on-screen, but printout was not delivered by courier until the next day, which meant students got them only a week later. When the courier service was supplemented by two fast line-printers, no student left empty handed even from the first session.

From 1976 to 1981 the course comprised a formal statistics element taught by the late John Utting (then Deputy Director (Research) at the National Children's Bureau, previously Deputy Director of the SSRC Survey Unit) and an SPSS element taught by myself. For the first two years, Maureen Ashman (Senior Programmer with special reponsibility for SPSS) provided technical liaison with Computer Services (including limited correction and resubmission of SPSS jobs). John Utting retired in 1981 and Jim Ring took over the statistics teaching. When Jim was unable to continue, he provided an early draft of his statistical notes for distribution to students and I revamped the evenings into one hour of SPSS presentation (covering the statistical elements as and when appropriate, but not to the same depth) followed by a two hour hands-on session in the computer laboratory and ending with discussion of

[23] See: Cathie Marsh and John Utting, article in ULCC Newsletter (1975)

[24] Colin Brown (Consumer Association), Cathie Marsh (SPS, Cambridge), Norman Perry (West Midlands Regional Planning), Jim Ring (Personal Social Services Council), Alan Marsh (Government Social Survey) John Hall (Polytechnic of North London), John Utting (National Children's Bureau)

[25] See Appendix 2 for syllabus and sample assessment

results. From 1990, as **SR501: Survey Analysis Workshop** under the Post-Qualifying Scheme, students could gain 15 points towards a CNAA Masters' degree, provided they took the assessment.

An undergraduate version of the course was originally taught (from 1980 onwards) to 4th year students on the Social Research and Planning option of the 4-year BA Hons Applied Social Studies, but was moved to the second semester of the second year of the revalidated course and continued into the modular degree scheme for BSc (Hons) Social Science as **SR206: Data Management and Analysis**. It was compulsory for the Social Research pathway and strongly recommended for Sociology. Each year, to the dismay of their Course Tutor, a handful of Sociology students opted to switch to Social Research.

For BSc Social Research students this not only complemented their Statistics course (some claimed greater understanding of statistics from the SPSS course than from the official one!) but also prepared them for effective professional placements in their 3rd year. Many of them were involved in survey based projects and, as registered students, could use PNL facilities without charge: this arrangement resulted in several publications with our students as sole or joint authors. On graduation, many obtained appointments as researchers with their placement agencies or won funded post-graduate studentships, often, in the former case, ahead of candidates with higher degrees from other universities. Several previous staff and students of PNL/SRU are now to be found in senior positions in UK social research.

Between 1976 and 1992, around 700 students were trained in the use of SPSS to process and analyse data from their own survey-based projects and/or from major questionnaire surveys in the public domain. The courses were great fun: the students enjoyed them. They gained a basic understanding of statistics, some practicalities and politics of social research, and acquired technical skills in handling and analysing survey data. The courses closed after 1992 when no-one could be found to continue the teaching. Courses with training in SPSS were later offered elsewhere, and still are (e.g. Surrey, Essex, Lancaster)

1.6: Training materials

The original course handouts were mostly written direct to a mainframe computer or in WordStar4 on a PC (Amstrad 8256 or 8512 and Euromicro) and refer to successive releases of SPSS installed on a range of machines, from the ICL2900 and CDC2000 at ULCC (via LSE) through ICL1900S, DEC-10 and DEC-KL20 and finally the Vax cluster at PNL. My current undertaking is to convert and update the course leaflets and handouts from WS4 to MS-Word and to modify the examples of setup, output and saved files for use with SPSS for Windows. Because conversion of tabular output from Vax lineprinter format and WS4 to MSWord is tedious and complicated, and also because some original data has been irretrievably lost (and thus not available to regenerate tables in Windows format), some files retain tables and figures in the original lineprinter format of output from earlier versions of SPSS. The following documents (only the naming one is specific to SPSS) are already posted on:

http://www.marketresearchworld.net/index.php?option=com_content&task=category&sectionid=15&id=74&Itemid=60

Introduction to Survey Analysis (96kb)
Introduction to Tabulation (34kb)
Conventions for Naming Variables in SPSS (56kb)

This site cannot handle colour text or graphics, and has problems with tabular formats: an alternative is being developed[26]. Meanwhile the original versions with colour-coded text, clearer tables and full colour graphics can be obtained from the author on johnfhall@orange.fr . Also available from the author are:

Introduction to the use of computers in survey analysis (1981: annotated 2006, 83kb)
[General notes not specific to SPSS, but the tables are from various releases of SPSS]

Survey Analysis Workshop - Syllabus & specimen assessment (Final version 1991-92, 65kb)
Survey Analysis Workshop - Stage 1: From questionnaire to data file (2006, 364kb)
Survey Analysis Workshop - Stage 2: Completing your data dictionary (2006, 167kb)
[Newly written for SPSS for Windows: step-by-step process from completing a fun questionnaire, coding on to a data sheet, entering data and setting up first (**data list**) and second (**variable labels**, **value labels**, **missing values**) editions of saved files, with screen dumps of each stage]

[26] See: www.aronique.co.uk click **button icon**, then **Web Design**, scroll to and click on **JH** in bottom right (very experimental)

Survey Analysis Workshop - Derived variables 1a – COUNT (2006, 355kb)

[Part 1 of a section on the use of **count** and **compute** to generate scores, analysis of scores using **frequencies** with associated statistics and graphics, **crosstabs** and **graph**: example uses items measuring attitudes to women from a survey of fifth formers]

Survey Analysis Workshop - Derived variables 1b – COMPUTE

[not quite ready, but in an advanced stage of preparation]

Survey Analysis Workshop - Statistical notes 1-7 (1992, 288kb)
Survey Analysis Workshop - Statistical notes 8-13 (1992, 108kb)

[Original documents as distributed to students: specially written by Jim Ring, and edited by myself, as a supplement to the standard texts referred to in the notes.]

Analysing multiple response with SPSS – an introduction (1992, updated 2006, 76kb)

Pallant 2001 review[27] (1st edition 2001, full original review, 2002, 3557 words, 51kb)
Pallant 2005 review[28] (2nd edition 2005, short review, 2007, 838 words, 32kb)

[Reviews of Julie Pallant, **SPSS Survival Manual**. This book is heavily biased towards inferential statistics for budding psychometricians and statisticians. It is not suitable for students in sociology and related subjects, who tend to use percentage tables. Short on file construction, data checking and completely lacking tabulation, it exclusively uses the drop-down menus and drop-down menus throughout, which is pointless for me, an ardent syntax fan, and incidentally explains why some useful commands are not mentioned at all, since they are not available in the drop-down menus. Apart from a new chapter on Loglinear Regression and a couple of paragraphs on Visual Bander, the second edition is virtually identical to the first, hence the short review]

Remaining materials to process include detailed introductions to, and explanations of, SPSS commands and syntax for data transformation and analysis, including tricks of the trade, worked examples, class exercises and homework exercises. They will eventually constitute a self-contained Teach-Yourself package and form a logical progression with integrated and repeated use of data from the fun questionnaire completed by students at the beginning of the course, from the 1989 British Social Attitudes survey and from a 1981 survey of fifth-formers at a comprehensive school in North London. Exercises using data from later surveys (eg European Social Survey) are also likely to be added. For training purposes everything needs breaking up into bite-sized chunks, with plenty of repetition for common tasks. Nowadays speed and quality of surveys and courses have greatly improved, but one other constraint persists, even in this paper: the constant need to fit everything on one side of A4.

[29]

One side of A4!

Photograph © 2001 by Len Schwer

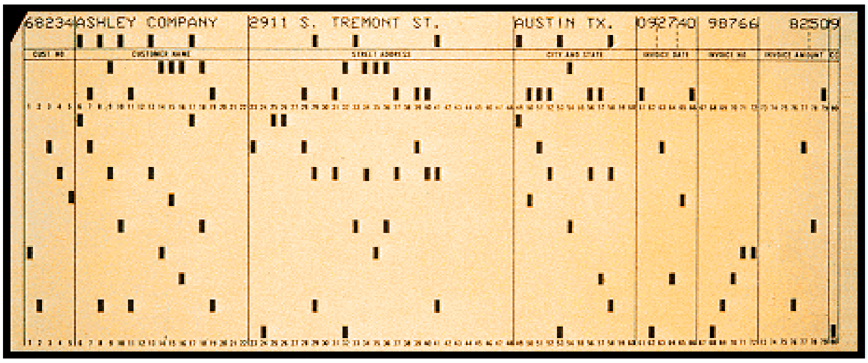
[27] The review appeared (shortened to 1206 words) in SRA News, Nov 2002 (http://66.102.9.104/search?q=cache:lXQlm9cTcn8J:www.the-sra.org.uk/documents/pdfs/sra_nov_2002.pdf+SRA+%2B+Pallant+%2B+Hall&hl=en&ct=clnk&cd=2 pp 11-12), but the additional notes have yet to be posted on their website. They are available from the author on hallj@wanadoo.fr

[28] Forthcoming in SRA News, 2007.

[29] For the uninitiated, the A4 class was a Gresley designed 4-6-2 Pacific used by LNER to haul the King's Cross-Edinburgh express, known to trainspotters as a "streak". This is Mallard, which holds the world speed record record (for a steam-hauled train) of 126mph (1938 between Thirsk and York). You can see the other side in York at the National Railway Museum. Photo obtained from http://www.schwer.net/Britain2001/York/index.htm covering a visit to York in 2001.

2: Layout, usage and changes

Throughout the 1970s and 1980s survey data was normally supplied by fieldwork agencies on 80-column cards or, with later computer developments, as card images on magnetic tape.


From Computer Desktop Encyclopedia
© 2000 The Computer Language Co. Inc.
68234 ASHLEY COMPANY
2911 S. TREMONT ST.
AUSTIN TX.
092740
98766
82509

Each column had 12 punching positions. At the top were two "zones" and below them digits 0 to 9. There could be up to three holes punched in each column. Digits 0 through 9 used a single punch, letters of the alphabet were indicated by two holes (combinations of the upper zone with 1 to 9 for A to I, lower zone with 1 to 9 for J to R and 0 with 2 to 9 for S to Z). Special characters were represented by 3-hole combinations of upper, lower or 0 with 3,4 or 8. The upper and lower zones (known in the trade as 12 and 11, Y and X, or + and -) could also be single-punched and were often used for "Don't know" or "Not applicable" responses.

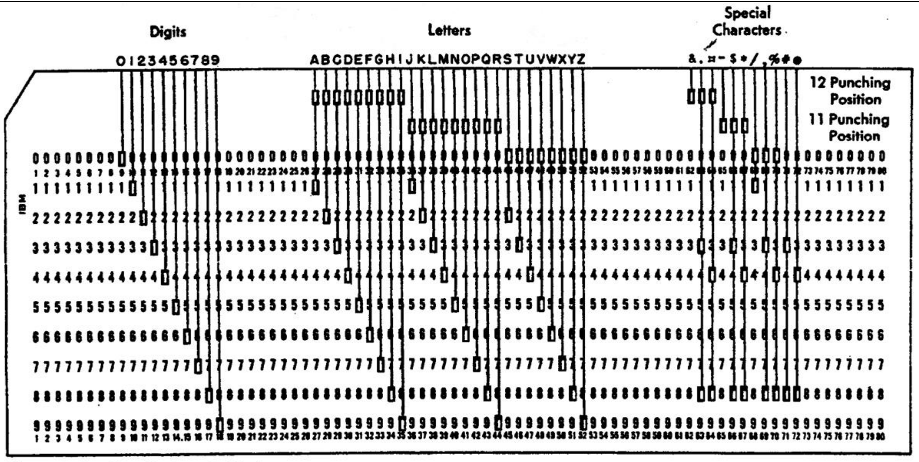


Figure 2. Punching Positions in Card

Unless otherwise specified, the cards usually contained at least some data in multi-punched format (column binary) either to code multiple responses to the same question, or to get more than one variable, in the same column). For example, the raw data from the first pilot of the SSRC Survey Unit Quality of Life[30] survey in 1971 was supplied on 80-column cards (two cards per case): the listing (printout) of the data looked like this (first three cases only: multipunches highlighted as **red $** signs).

```
001110204+57462235696172244322232422- 2O- 322K2-   3$62$$5 05902-- 89564$-147321
0012$$$%  1 23        0 19$0$78$$6110$Q31111010 23463110       4113+2211207637321
002119051-44689428858-45242524431442324T31$3823+84$8354$77 158-5-7M$6$O6$$417321
0022$$$$    2 1 3     1$1$$$$22F$11222-41010011022113100      310002220107637321
003114202+355-953273--3324454341415591+N91238-2+8257$$55+-  $- 4-7$$5$$5$2137321
0032$$$$  1  32       0 12$$$26N$11222$51111011012122010      310122215127637321
```

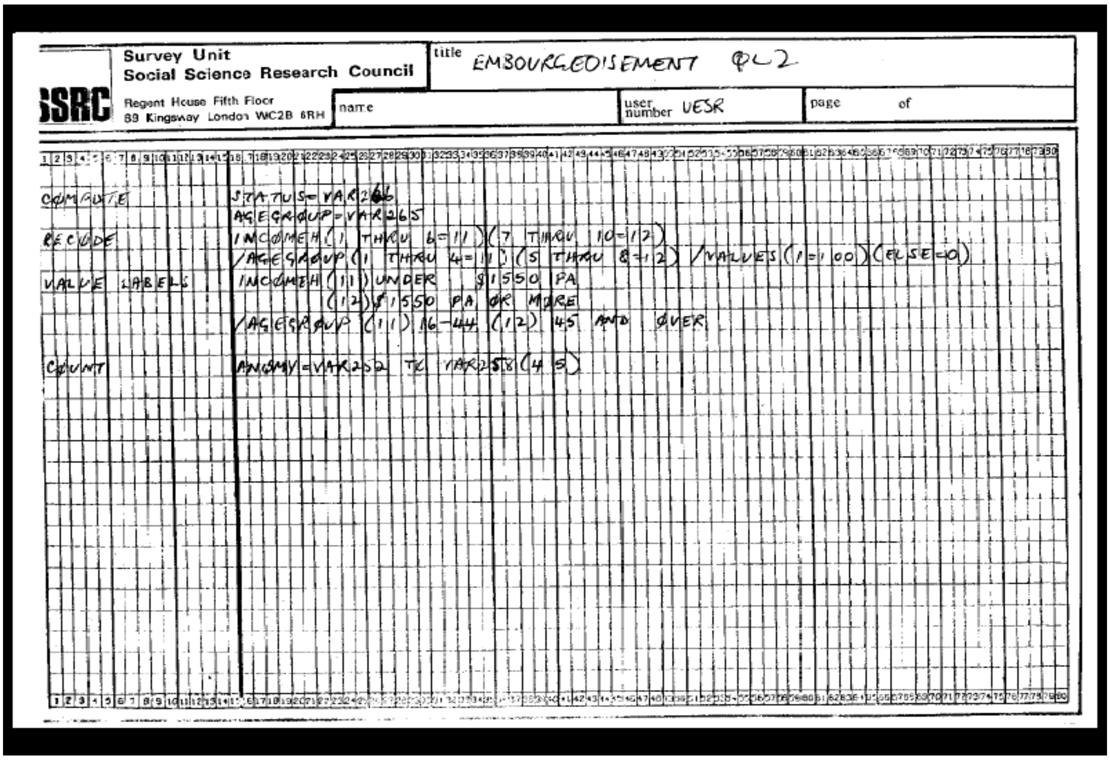
[30] UK Data Archive study 247. The survey was conducted March – May 1971, but SPSS files were not created until 1973.

Although some multipunching could be interpreted as alphabetic or special characters, much of it could not, and was printed up as a dollar sign **$**. For this reason, at SSRC, we always ran SPSS on new data using **alphanumeric** format, partly because some data would be multipunched and partly because we were never quite sure what the raw data might contain (some numeric data was occasionally mis-punched as alphabetic, even after verification). Attempting to read such data as numeric would result in an error. Multipunched data were later spread out (using the LSE program **MUTOS**), in this case on to four additional cards to yield a final raw data set with 6 cards per case e.g. (first case only):

```
001110204+57462235696172244322232422- 2O- 322K2-   3$62$$5 05902-- 89564$-147321
0012$$$%  1 23        0 19$0$78$$6110$Q31111010 23463110      4113+2211207637321
00139000000101000090000001001000101000110100009000000000900000000101001000100 7321
00140001001011000001010001000000000000100001000000000000101000000900000000000 17321
00150000000100100001000100000100100110010000011112222222222222221290000000000 17321
0016000000001011100000010000000101                                          7321
```

The original layout of the SPSS language was determined by the use of 80-column Hollerith cards on which columns 1-15 were reserved for commands, columns 16-72 for sub-commands and specifications, and columns 73-80 for numbering of the cards (a necesary precaution when decks could contain hundreds, if not thousands, of cards: imagine dropping a trayful!). We never needed numbering for raw data since case and card numbers were always punched at the beginning of each card and the survey identification code at the end. Since SPSS was easy to read and the card contents were printed across the top of each card, we very rarely needed to number SPSS setup jobs either. What we did use was a standard 80-column data-preparation sheet, modified for SPSS. Each sheet had 25 rows, allowing up to 25 lines of code to be written, and had a heavy line drawn between cols 15 and 16 as a visual aid to keeping subcommands and specifications out of the field reserved for commands.

SSRC Survey Unit coding sheet for 80-column Hollerith cards (c.1973) for use with SPSS



SSRC Survey Unit
Social Science Research Council
Regent House Fifth Floor
89 Kingsway London WC2B 6RH

title EMBOURGEOISEMENT ØL2
name
user number VESR
page of

```
CØMPUTE         STATUS=VAR266
                AGEGRØUP=VAR265
RECØDE          INCØMEH(1 THRU 6=11)(7 THRU 10=12)
                /AGEGRØUP(1 THRU 4=11)(5 THRU 8=12) /VALUES(1=100)(ELSE=0)
VALUE LABELS    INCØMEH(11)UNDER £1550 PA
                       (12)£1550 PA ØR MØRE
                /AGEGRØUP(11)16-44(12)45 AND ØVER

CØUNT           ANGMY=VAR252 TO VAR258(4 5)
```

Cards were later replaced by card-images on magnetic tape or disk (and by lines on VDU screens). As SPSS has become more “intelligent”, layout restrictions have gradually been lifted, but commands still have to start in column 1, must be followed by at least one space and there must be at least one space at the beginning of any continuation lines. However, for teaching purposes (and even for experienced researchers) it is still useful to start with a visually distinct layout as it makes the files easier to read and the logic easier to follow. Hence the extensive use, in PNL training and research materials, of tabs to separate SPSS commands from sub-commands and specifications.

2.1: Evolution of SPSS syntax

Since 1972, when I was first exposed to SPSS, there have been many subsequent releases and updates, not just mainframe versions, but also **SPSS PC+** and more recently **SPSS for Windows**. **SPSS11.0** (the version made available to me) has most, but not all, facilities of mainframe release 4 of **SPSS-X**. Examples of changes over the years include:

VARxxx TO VARyyy	≡	**Vx to Vy (Qx TO Qy)**
UPPER CASE only in labels	≡	**Any printing characters in primes**
Limits to characters in labels **40 for variables** **20 for values**	≡	**Removed, theoretically 255,** **but printout constraints apply**
VARIABLE LIST **INPUT FORMAT** **(Fortran format statement)** **INPUT MEDIUM**	≡	**DATA LIST** **FILE =** **RECORDS =**
BREAKDOWN	≡	**MEANS**

Because of these changes, many setup jobs from the 1970s and 1980s will no longer work.

For example, the original SPSS setup file for the 2-card **Quality of Life** data set (1st pilot survey, 1971) included the following data definition, following the **positional** variable naming convention developed at the SSRC Survey Unit and reading most of the data as **alphanumeric** in the Fortran format statement:

File definition 1973: (Quality of Life: 1st pilot survey 1971, SSRC Survey Unit)

```
RUN NAME        QL1UK1  - PILOT 1 FIRST SYSTEM FILE
FILE NAME       QL1UK1 QUALITY OF LIFE PILOT I UK
VARIABLE LIST   VAR101 VAR105 VAR109 TO VAR137
                VAR141,VAR144,VAR145,VAR148
                VAR149 VAR152 VAR155 VAR158 VAR159
                VAR162 VAR165 VAR166 VAR169 VAR172
                VAR175,VAR176,
                VAR209 TO VAR223 VAR225
                VAR230 VAR234 TO VAR237 VAR240 TO VAR256
                VAR263 VAR264 VAR266 TO VAR268 VAR270
INPUT MEDIUM    INDATA
INPUT FORMAT    FIXED
                (F3.0,1X,A4,F1.0,13A1,
                14F1.0,A1,3X,A1,2X,F1.0,A1,
                2X,2A1,2X,A1,2X,A1,2X,2A1,2X,A1,
                2X,2A1,2X,A1,2X,A1,2X,2A1,4X/
                8X,15A1,1X,1A1,4X,A2,2X,A1,
                F1.0,2A1,2X,17A1,6X,A1,A2,2A1,A2,A4)
NO. OF CASES    213
```

In the original version of SPSS it was possible to define variables in alphanumeric format and then **recode** the alpha values to numeric using **(convert)** to keep the **same variable names**. Recodes for values other than the digits 0-9 and the two zone punches had to be defined separately.

Variables read in as alpha from card 1 were later recoded to numeric with:

```
RECODE       VAR105 ('++++'=9999)
            (CONVERT)/
            VAR110 ('+'=2)('-'=1)('0'=88)
            (CONVERT)/
            VAR111 TO VAR122 VAR137 VAR141 VAR145
            VAR149 VAR152 VAR155 VAR158 VAR162
            VAR166 VAR169 VAR172 ('-'=10)('+'=99)
            (CONVERT)/
            VAR144 (1=2)/
            VAR148 VAR165 ('+'=1) ('-'=2)
            (CONVERT)/
            VAR159 (' '=1) ('-'=0)
            (CONVERT)/
            VAR175 ('+' ' '=88) ('4'=3)
            (CONVERT)/
            VAR176 (' ','+'=99)
            (CONVERT)
```

*[**NB:** Remember this was all on 80-column cards. Keeping the* **(CONVERT)/** *on a separate line was the result of making several copies of that card and inserting them as appropriate. It may seem wasteful of cards, but it saved a lot of keypunching. The same applies to typing in lines in syntax today.]*

However, **string** variables (as they are now called) can only be converted **into new variables**. In order to recreate the file it was necessary, not only to use **data list**, but also to use **dummy variable names**. These were later recoded into the original names (to tally with the user manual) and dropped when the file was saved. Thus, reading from the original data set, but with dummy variable names:

```
data  list
     file 'f:qluk1.dat'  records 6
     /1  serial 1-3
         v105 to v180 5-80 (a)
     /2  v209 to v280 9-80 (a).
```

to produce an intermediate display:

```
Data List will read 6 records from the command file

Variable    Rec    Start    End         Format

SERIAL        1        1      3         F3.0
V105          1        5      5         A1
V106          1        6      6         A1
V107          1        7      7         A1
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~
V278          2       78     78         A1
V279          2       79     79         A1
V280          2       80     80         A1
```

*[**NB:** This example illustrates the advantage of the **positional** convention for variable names: they can be checked visually against the record number and start column.]*

The conversion from dummy variables with alpha values back into the original (user-manual) variables with numeric values can then be effected by:

```
RECODE     V209 TO V222  (' ','+','-'=0) ('1'=2) ('2'=1) ('3'=-2) (CONVERT)
              into VAR209 TO var218 xvar219 var220 to VAR222 /
           V223 V225 V234 ('+'=88) (CONVERT)
              into VAR223 VAR225 VAR234 /
           V230 ('99'=98) ('++'=99) (CONVERT) into var230/
           V236 V237 ('+'=99) (CONVERT) into var236 var237/
           V240 ('+'=88) ('1' '2'=3)(CONVERT) into xvar240/
           V241 TO V252(' '=88)(CONVERT) into VAR241 TO VAR252/
           V253 (' '=88) ('4' '5'=3)(CONVERT) into var253/
           V254 TO V256(' '=88)(CONVERT) into VAR254 TO VAR256.
RECODE      V263 ('+'=88) (CONVERT) into var263
           / V264 ('++'=88) (CONVERT) into var264
           / V266 V267 ('+'=1) ('-'=2) ('0'=3) ('1'=4) ('2'=5)  ('3'=6) ('4'=7)
             ('5'=8) ('6'=9) ('7'=99) (' '=99)(CONVERT) into var266 var267
           /V268 (CONVERT) into var268
           / V270 ('++++'=88) (CONVERT) into var270.

SAVE OUT  ‘F:QL1.SAV’
           /KEEP VAR209 TO VAR270.
```

*[**NB:** There’s a switch mid-job to putting slashes before the variable names! Much easier to check.]*

Modifications were subsequently needed for the variable labels and value labels to get rid of apostrophes and full stops (which SPSS interpreted as beginning a label or ending a command). These were tedious rather than complicated and took several runs as they were quite difficult to spot, but with the sheer speed of SPSS it was quicker to run jobs, check the error reports, amend the setup syntax and then delete the output file without saving it.

2.2: Other developments in SPSS

Blue (later maroon) manual **(in A-Z order of commands)**	≡	**Norusis (1988)** **(in user-friendly** **research process order)** **..but for SPSS13 ???**
Batch only **(on 80-column cards)**	≡	**Interactive** **(via VDU keyboard)**
Output on line-printer only	≡	**VDU with on-screen scrolling**
Mainframe only	≡	**SPSS PC+** **SPSS for Windows**
UPPER CASE only	≡	**lower case**
Full syntax only	≡	**Abbreviated syntax** **Drop-down menus**

*[**NB:** For some purposes, the switch from syntax to drop-down menus may well be a retrograde step]*

2.3: Variable Names

Now let's have a look at some examples from surveys conducted and processed by other people, and using conventions derived direct from the original SPSS manuals, but modified as restrictions on layout were lifted. The first is an extract from the SPSS setup file written by **John Curtice** and **Andrew Shaw** at Liverpool University for the 1987 survey of **British Social Attitudes** conducted by Roger Jowell and colleagues at **Social and Community Planning Research** (SCPR, now **NatCen**). The questionnaire consisted of a main section (interviewer administered) plus one of two alternative sub-sections A or B (also interviewer administered) and its related self-completion section.

Variable names can be up to 8 characters in length and must start with a letter of the alphabet. Curtice and Shaw used **mnemonic** names, which were supposed to look something like the variables they represented and therefore be meaningful and easier to understand and remember. We shall see!

An alternative convention, developed at the SSRC Survey Unit and derived from the LSE Survey Data Tabulation program **SDTAB**, makes it easier to work direct from a questionnaire (provided it has data layout printed on it). This uses **positional** variable names of the form **Vddd** (or **Vdddd** if there are 10 or more cards) in which the last pair of digits indicates the start column of a field and the first digit(s) the card/record number. Data from record 2 (of 23) were specified in mnemonic form on the data list by:

(British Social Attitudes 1987)

```
/2 VERSION 8
   READPAP 9 WHPAPER 10-11 SUPPARTY CLOSEPTY 12-13 PARTYID1 14-15
   IDSTRNG CNTLCNCL RATES RENTS EEC NATO NATION USANUKE OWNNUKE   UKNUCPOL
   DEFPARTY PEACE NIRELAND TROOPOUT 16-29
   PRICES UNEMP UNEMPINF CONCERN INDUSTRY ECOHELP1 TO ECOHELP9 STATEOWN
   TAXAVOID TAXSNEAK INCOMGAP TAXHI TAXMID TAXLOW SRINC
   HINCDIFF HINCPAST HINCXPCT 31-55
   RECONACT 56-57
   RFTEDUC RTRAING RPAIDWRK RWAITWRK RREGISTD RSEEKWRK RNTLOOK RSICK
   RRETIRD RATHOME RELSE REMPLOYE 58-69 EJBHOURS 70-71
   EJBHRCAT WAGENOW PAYGAP WAGEXPCT NUMEMP EMSMEWRK EMSEXWRK
   EMWOMCLD EMWOMWLD 72-80
```

…which is more easily written using the positional convention as:

```
/2   version 8
     v209 9 v210 10-11 v212 v213 12-13 v214 14-15 v216 to v229 16-29 v231 to v255 31-55
     v256 56-57 v258 to v269 58-69 v270 70-71 v272 to v280 72-80
```

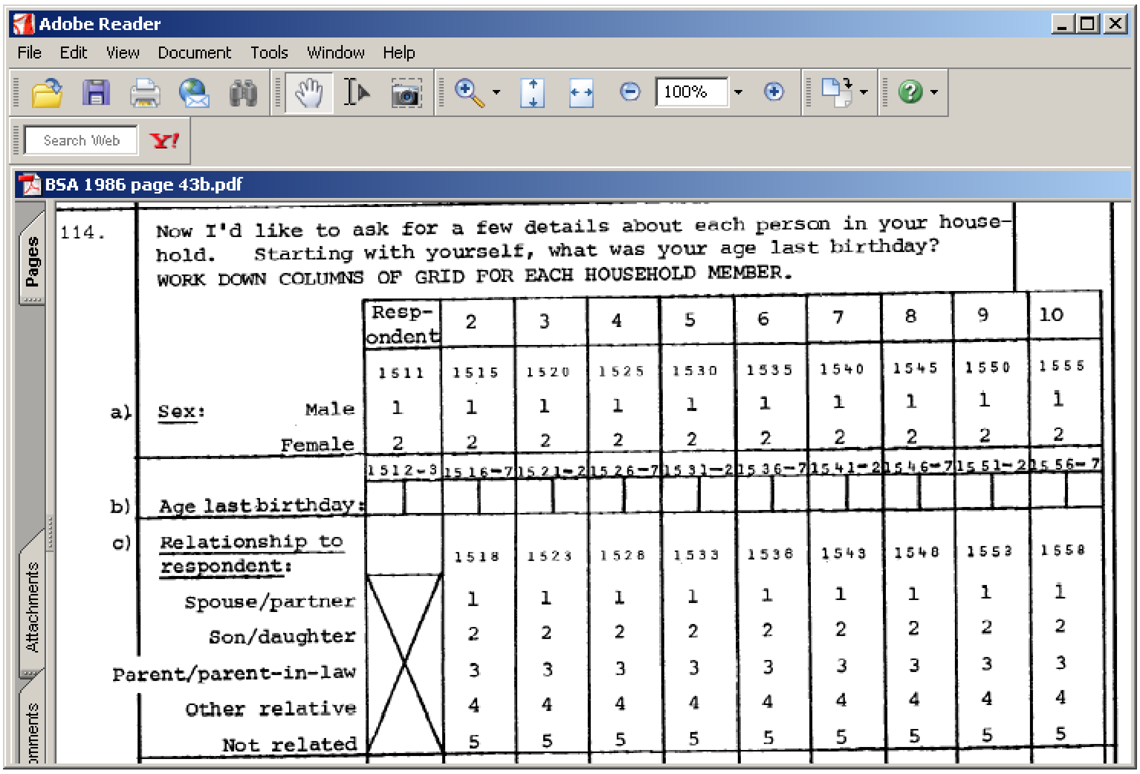
…and has the advantage, when SPSS is run, of enabling a visual check on whether each variable has been read from the correct field (the name should match the record and start column for each variable).

```
Data List will read 23 records from F:\bsa87.dat

Variable   Rec   Start    End         Format

VERSION      2       8      8         F1.0
V209         2       9      9         F1.0
V210         2      10     11         F2.0
V212         2      12     12         F1.0
~ ~ ~       ~ ~     ~ ~    ~ ~       ~ ~ ~
V278         2      78     78         F1.0
V279         2      79     79         F1.0
V280         2      80     80         F1.0
```

This convention is especially useful for sections of the questionnaire which have no question numbers. For example household information collected by means of a grid (as in the facsimile below of part of the questionnaire for the British Social Attutudes survey 1986).



114. Now I'd like to ask for a few details about each person in your household. Starting with yourself, what was your age last birthday?
WORK DOWN COLUMNS OF GRID FOR EACH HOUSEHOLD MEMBER.

		Resp-ondent	2	3	4	5	6	7	8	9	10
		1511	1515	1520	1525	1530	1535	1540	1545	1550	1555
a)	Sex: Male	1	1	1	1	1	1	1	1	1	1
	Female	2	2	2	2	2	2	2	2	2	2
		1512-3	1516-7	1521-2	1526-7	1531-2	1536-7	1541-2	1546-7	1551-2	1556-7
b)	Age last birthday:										
c)	Relationship to respondent:		1518	1523	1528	1533	1538	1543	1548	1553	1558
	Spouse/partner		1	1	1	1	1	1	1	1	1
	Son/daughter		2	2	2	2	2	2	2	2	2
	Parent/parent-in-law		3	3	3	3	3	3	3	3	3
	Other relative		4	4	4	4	4	4	4	4	4
	Not related		5	5	5	5	5	5	5	5	5

…for which instead of using:

```
/15    sexpers1 11 agepers1 12-13
       sexpers2 15 agepers2 16-17
       ~ ~ ~
       ~ ~ ~
       sexpers0 55 agepers0 56-57
```

[0, not 10, to keep to 8 characters]

. . . (which will later be difficult to find in the data file) it's much better to use:

```
/15    v1511 11 v1512 12-13
       v1515 15 v1516 16-17
       ~ ~ ~
       ~ ~ ~
       v1556 56-57
```

. . .which relate directly to the questionnaire and will be easier to find in the file. Properly written variable labels (with question number at beginning where appropriate) will indicate the nature of the variable.

```
V1511 'Q114: Sex of respondent'
V1512 'Q114: Age of respondent'
V1515 'Q114: Sex of 2nd person'
V1516 'Q114: Age of 2nd person'
```

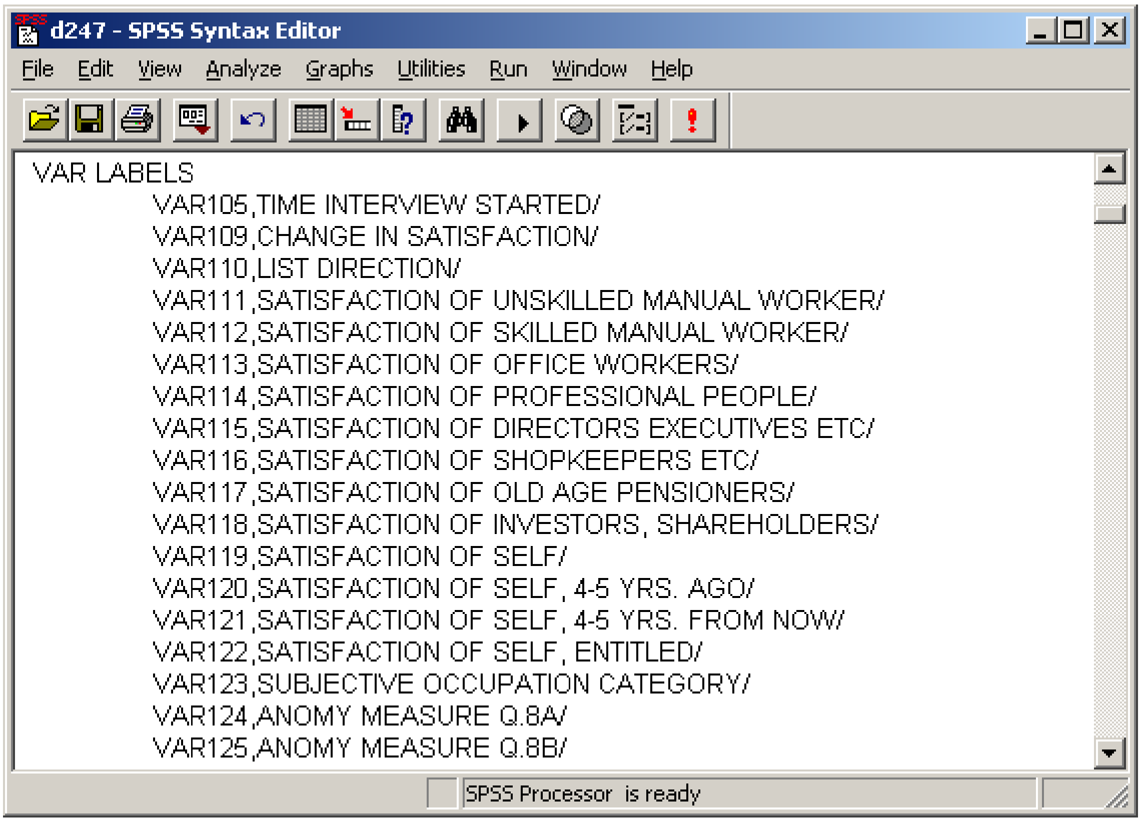
V1511 and **V1512** can be changed later to **SEX** and **AGE** [**rename (v1511 v1512 = sex age)**], but for demographic variables used frequently for analysis it is preferable to create new variables in another section of the file to make later analysis quicker (eg: **cros sex to incgroup by v213 to v227 /cel row.**)

```
compute sex = v1511.
compute age = v1512.
```

2.4: Variable Labels

In 1973, we had to write variable labels in **UPPER CASE** only (40 characters maximum), use **commas** to separate names from labels, and keep label specifications separated from each other by **slashes** (at the end of the line as per the manual). We also kept each variable on a separate line.

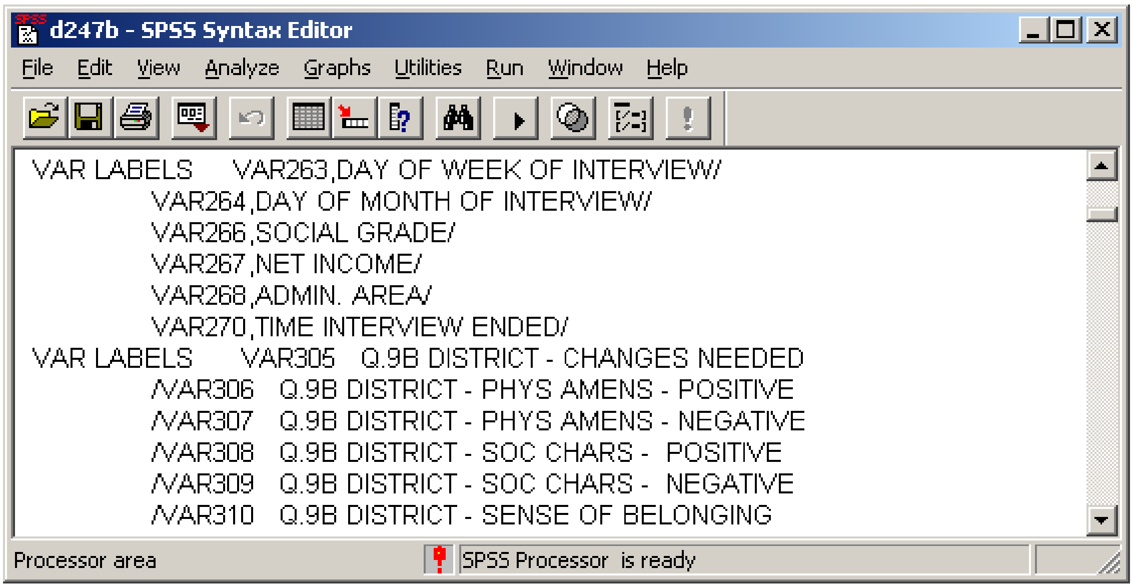
1973a: (SSRC Quality of Life Survey 1st Pilot 1971)


d247 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
SPSS Processor is ready

```
VAR LABELS
          VAR105,TIME INTERVIEW STARTED/
          VAR109,CHANGE IN SATISFACTION/
          VAR110,LIST DIRECTION/
          VAR111,SATISFACTION OF UNSKILLED MANUAL WORKER/
          VAR112,SATISFACTION OF SKILLED MANUAL WORKER/
          VAR113,SATISFACTION OF OFFICE WORKERS/
          VAR114,SATISFACTION OF PROFESSIONAL PEOPLE/
          VAR115,SATISFACTION OF DIRECTORS EXECUTIVES ETC/
          VAR116,SATISFACTION OF SHOPKEEPERS ETC/
          VAR117,SATISFACTION OF OLD AGE PENSIONERS/
          VAR118,SATISFACTION OF INVESTORS, SHAREHOLDERS/
          VAR119,SATISFACTION OF SELF/
          VAR120,SATISFACTION OF SELF, 4-5 YRS. AGO/
          VAR121,SATISFACTION OF SELF, 4-5 YRS. FROM NOW/
          VAR122,SATISFACTION OF SELF, ENTITLED/
          VAR123,SUBJECTIVE OCCUPATION CATEGORY/
          VAR124,ANOMY MEASURE Q.8A/
          VAR125,ANOMY MEASURE Q.8B/
```

However, it was easy to forget the slash at the end of the line (a common cause of error messages) so we moved it to the beginning of the next variable (and SPSS still worked!).

1973b: (SSRC Quality of Life Survey 1973)

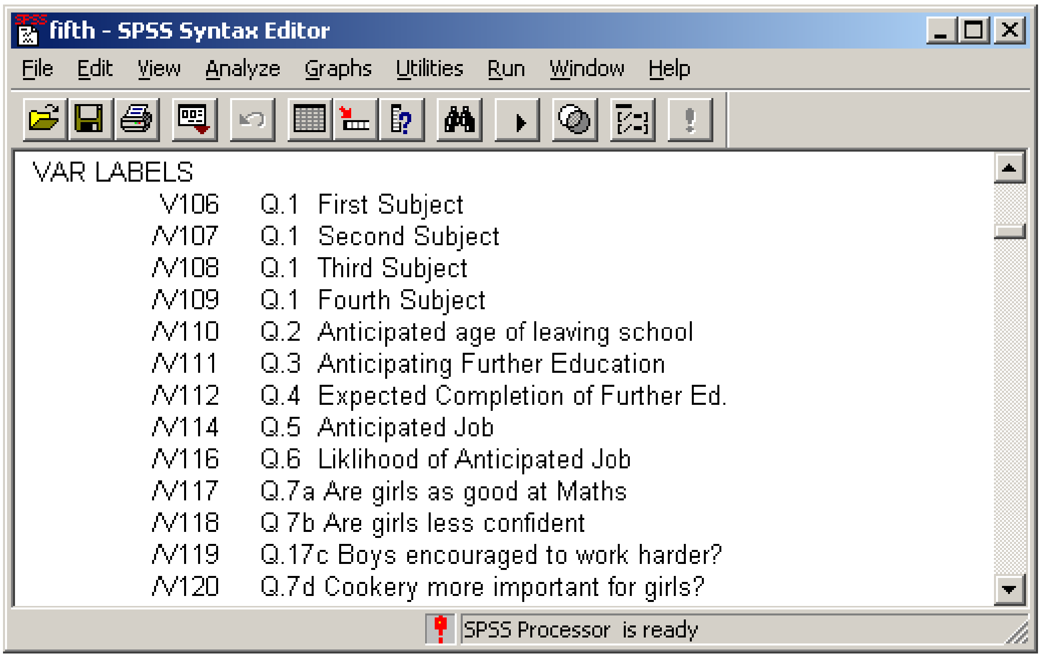

d247b - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
Processor area
SPSS Processor is ready

```
VAR LABELS   VAR263,DAY OF WEEK OF INTERVIEW/
          VAR264,DAY OF MONTH OF INTERVIEW/
          VAR266,SOCIAL GRADE/
          VAR267,NET INCOME/
          VAR268,ADMIN. AREA/
          VAR270,TIME INTERVIEW ENDED/
VAR LABELS    VAR305  Q.9B DISTRICT - CHANGES NEEDED
          /VAR306  Q.9B DISTRICT - PHYS AMENS - POSITIVE
          /VAR307  Q.9B DISTRICT - PHYS AMENS - NEGATIVE
          /VAR308  Q.9B DISTRICT - SOC CHARS -  POSITIVE
          /VAR309  Q.9B DISTRICT - SOC CHARS -  NEGATIVE
          /VAR310  Q.9B DISTRICT - SENSE OF BELONGING
```

[Note change of format mid-setup!]

Later, lower case was allowed, for variable names as well as labels, but we kept upper case for variable names. The labels did not need to be enclosed in primes, and full stops could be used in the label.

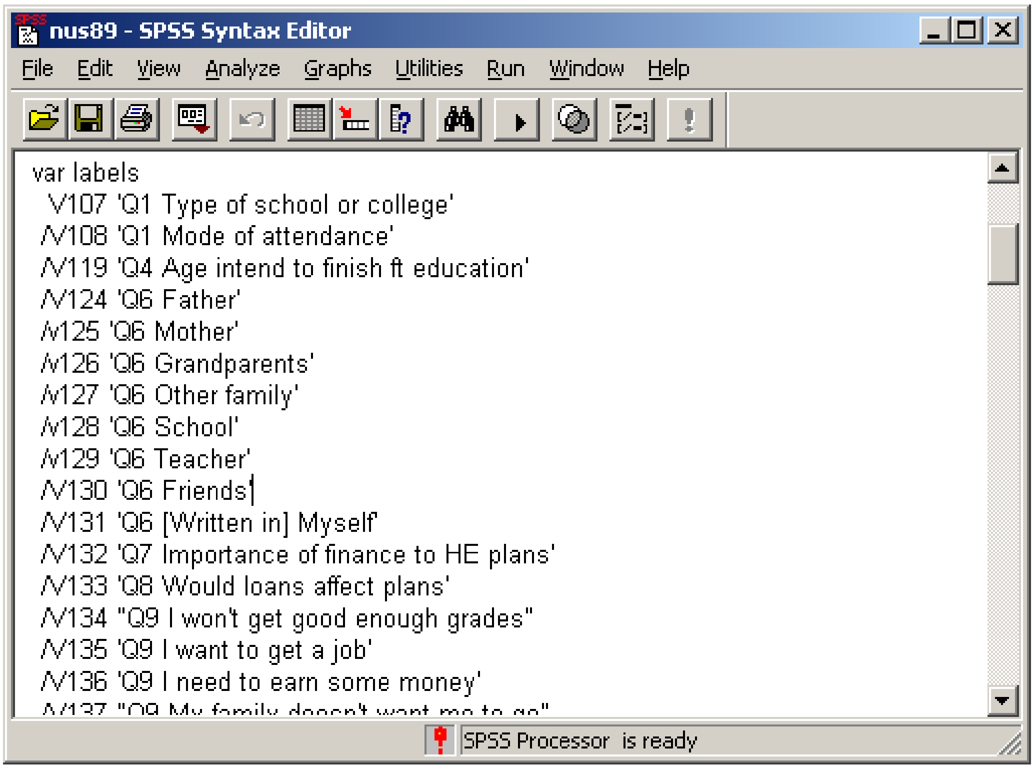
1981: (Survey of fifth form pupils in a North London comprehensive)


fifth - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
SPSS Processor is ready

```
VAR LABELS
            V106    Q.1  First Subject
            V107    Q.1  Second Subject
            V108    Q.1  Third Subject
            V109    Q.1  Fourth Subject
            V110    Q.2  Anticipated age of leaving school
            V111    Q.3  Anticipating Further Education
            V112    Q.4  Expected Completion of Further Ed.
            V114    Q.5  Anticipated Job
            V116    Q.6  Liklihood of Anticipated Job
            V117    Q.7a Are girls as good at Maths
            V118    Q 7b Are girls less confident
            V119    Q.17c Boys encouraged to work harder?
            V120    Q.7d Cookery more important for girls?
```

Later still, labels had to be enclosed in single primes (or double primes if there was an apostrophe in the label). People started using lower case, even for variable names (which appeared on output in upper case) and a single space instead of tabs on new continuation lines.

1989: (NUS Student Finance Survey[31] 1989)


nus89 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
SPSS Processor is ready

```
var labels
 V107 'Q1 Type of school or college'
 V108 'Q1 Mode of attendance'
 V119 'Q4 Age intend to finish ft education'
 V124 'Q6 Father'
 v125 'Q6 Mother'
 v126 'Q6 Grandparents'
 v127 'Q6 Other family'
 v128 'Q6 School'
 v129 'Q6 Teacher'
 V130 'Q6 Friends'
 V131 'Q6 [Written in] Myself'
 V132 'Q7 Importance of finance to HE plans'
 V133 'Q8 Would loans affect plans'
 V134 "Q9 I won't get good enough grades"
 V135 'Q9 I want to get a job'
 V136 'Q9 I need to earn some money'
 V137 "Q9 My family doesn't want me to go"
```

[31] See: Aughterson K and Foley K **Opportunity Lost: a survey of the intentions and attitudes of young people as affected by the proposed system of student loans** (National Union of Students, 1989). The example above was written by one of the NUS researchers, who had attended my **Survey Analysis Workshop** the previous year. Checking and coding of completed questionnaires was done as a vacation job by three of my students, Gill Bendelow, Kim Clarke and Malcolm Williams. All three got Firsts: Malcolm now holds the first UK Chair in Social Research Methodology at Plymouth.

Returning to the 1987 **British Social Attitudes** survey, the extract below defines the variable labels for the first seven variables in the file. Commas were no longer needed to separate names from labels because SPSS now treated the space after the variable name as a delimiter. Variable labels could by then be written in upper and lower case (enclosed in primes) yet Curtice[32] and Shaw were still using UPPER CASE for everything! They also put the slash at the end of each label specification, but it is better to put it **before** the second and subsequent variable names. It was very easy to forget the final slash, a common cause of SPSS error messages, and it was also difficult to find the culprit later.

British Social Attitudes 1987

```
VARIABLE LABELS
     VERSION  QUESTIONNAIRE VERSION ADMINISTERED/
     READPAP  Q1A R READS NEWSPAPER 3+ TIMES PER WEEK/
     WHPAPER  Q1B [IF READS 3+ TIMES] WHICH PAPER/
     SUPPARTY  Q2A POLITICAL PARTY SUPPORTER/
     CLOSEPTY  Q2C [IF NOT SUPORTR] CLOSER TO ONE PARTY/
     PARTYID1  Q2B & 2D & 2E PARTY IDENTIFICATION[FULL]/
     IDSTRNG   Q2F HOW STRONG PARTY IDENTIFICATION/
```

It is not easy to find your way around inside this setup file (although the question number helps). It would have been better to use tabs to align the labels and lower case labels to make the file easier to read. It’s a very large file to do this piecemeal, but there is a quick way of doing it. By copying the set of **variable label** commands into Word and using **Format.. Change Case** (to change everything to lower case) and then **Find.. Replace** [**/** with **‘**] [**v2** with **/v2**] and [**q** with **‘Q**] (**in that order!**), do some manual editing, then copy the result back into the SPSS syntax file we get this[33]:

```
variable labels
     version      ‘Questionnaire version administered’
    /readpap      ‘Q1a R reads newspaper 3+ times per week’
    /whpaper      ‘Q1b [if reads 3+ times] which paper’
    /supparty     ‘Q2a Political party supporter’
    /closepty     ‘Q2c [if not suportr] closer to one party’
    /partyid1     ‘Q2b & 2d & 2e party identification [full]’
     /idstrng     ‘Q2f How strong party identification’
```

This is much easier to read than the original, but the following is nicer to look at and is much easier to work with direct from the questionnaire.

```
variable labels
     version      ‘Questionnaire version administered’
     /v209        ‘Q1a R reads newspaper 3+ times per week’
     /v210        ‘Q1b [if reads 3+ times] which paper’
     /v212        ‘Q2a Political party supporter’
     /v213        ‘Q2c [if not suportr] closer to one party’
     /v214        ‘Q2b & 2d & 2e party identification [full]’
     /v216        ‘Q2f How strong party identification’
```

[32] John Curtice is now Professor of Politics at Strathclyde: his original convention for variable names in the 1983 British Social Attitudes was carried over into subsequent waves and was also adopted for the 2002 European Social Survey. Whilst this may be of use to those researchers interested in a few of the same variables over time, they are very awkward to use for everyone else. **The juggernaut rolls on!** However the ESS team has at least always used lower case labels and for subsequent waves now puts the question number at the beginning of the variable label.

[33] [**NB** **‘Q** appeared in odd places wherever there was a **q** other than at the beginning of a label. Some variables, eg from classification sections, needed to be edited manually to restore capital letters.]

How was this done? No, I didn't rewrite the entire setup file, I used the **rename** facility.

```
rename variables
     (readpap to idstrng = v209,v210,v212 to v214,v216).
```

… and if you're worried about putting things back as they were:

```
rename variables
     (v209,v210,v212 to v214,v216 =
  readpap whpap supparty closepty partyid1 idstrng).
```

Easy when you know how!

To do this with the whole file would be very tedious in SPSS syntax or Word (unless there's a way of doing it with the **Tables** format), but in **WordStar** and **EDT** you could strip out whole columns of text and paste them back on the other side of the **=** sign, another reason for using tabs in the layout. In fact it's easier these days to save both versions and keep both mnemonic and positional camps happy. Note the need in the second **rename** command to write out the list of original variable names **in full**. Using **to** would cause an error as the original names have been changed and SPSS won't be able to find them in the file. In fact it's probably easier for editing (and safer) to keep both lists in full on separate lines and a line between with the **=** sign, eg.

```
rename variables
     (v209,v210,v212 to v214,v216
     =
     readpap whpap supparty closepty partyid1 idstrng).
```

Best to play safe and save the new file with a different name.

2.5: Value Labels

In 1973 value labels were permitted in **UPPER CASE** only and, on printout, were limited to 20 characters for rows and 16 characters for column headings (in 2 blocks of 8): anything longer would either cause an error or be truncated. This made for some contorted spelling, tortuous abbreviations and additional packing spaces, especially in column headings: otherwise the output, already awful, looked even worse. As today, a list of variables could have the same value labels.

Each value had to be in round brackets followed by the value label. Variable lists were supposed to be separated by slashes at the end of the line, but (long before Norusis appeared) we always put them before the variable name on continuation lines: this way they were easier to see and less likely to be forgotten. At first **VALUE LABELS** had to be in col 1-16 and the specification in cols 16-72. The requirement to start in col 16 was later dropped and eventually lower case was introduced. By the late 1980s labels had to be in single primes (double primes if there was a prime in the label). However for ease of use and training purposes it is still best to use tabs to inset specifications and continuation lines.

1973 (Attitudes and Opinions of Senior Pupils: St Paul's School for Girls)

```
VALUE LABELS   FORM(1)LOWER  FIFTH(2)UPPER  FIFTH(3)LOWER  SIXTH (4)UPPER  SIXTH
               /YEARBORN(1)1954(2)1955(3)1956(4)1957(5)1958
               /MONTH(1)JANUARY(2)FEBRUARY(3)MARCH(4)APRIL(5)MAY(6)JUNE(7)JULY
               (8)AUGUST(9)SEPTEMBR(10)OCTOBER(11)NOVEMBER(12)DECEMBER
               /VAR111 TO VAR119(1)MOST   IMPORTNT(2)NEITHER(3)LEAST  IMPORTNT
               /JOB1 TO JOBAT25(1)ACCNTNCY,FINANCE(2)ARCHIT- ECTURE
                 (3)CIVIL  ENGINEER(4)CREATIVE ARTIST(5)DOCTOR, DENTIST
                 (6)FASHION(7)GOVNMNT,ADMIN.(8)HOUSE  -WIFE(9)INDUST. TECH.
                (10)JOURN-  ALISM(11)MILITARY SERVICE(12)NURSING
                (13)OUTDOOR,ATHLETIC(14)OWN    BUSINESS(15)PERFORM-ING ARTS
                (16)PERSONN-EL MNGMT(17)POLITICS(18)PUBLISH -ING
                (19)SALES + MARKETNG(20)SCIENCE-MATHS(21)SCIENCE-BIOLOGY
                (22)SCIENCE-SOCIAL(23)SECRET -ARY(24)SOCIAL  WORK
                (25)SOLICTR,BARRISTR(26)TEACHER-PRIMARY(27)TEACHER-SECNDARY
                (28)TOWN   PLANNING(29)TV,FILM PRODUCER(30)UNIVSTY LECTURER
                (31)LIBRAR  -IAN(32)PUBLIC  RELATNS(33)COMP-  UTERS(34)OTHER
```

This is very crowded (I didn't write it!) but cards and disk space (at least in the UK) were free, so we used to put each label on a separate line:

1973 (SSRC Quality of Life: 1st pilot survey 1971)

```
VALUE LABELS  VAR109  (1) LOT MORE
                      (2) LITTLE MORE
                       (3) SAME
                       (4) LITTLE  LESS
                       (5) LOT    LESS
              /VAR110  (1) FORWARDS
                        (2) BACKWRDS
              /VAR123  (1) UNSKILLDMAN WKRS
                        (2) SKILLDMAN WKRS
                        (3) OFFICE WORKERS
                        (4) PROFES-   SIONAL
                        (5) COMPANY DIRECTRS
                        (6) SHOP  KPRS ETC
                        (7) OAP'S
                        (8) INVESTRS  ETC
                        (9) NOT    KNOWN
```

Applying this to the St Paul's survey, by inserting carriage returns and tabs, and spaces after the brackets we can improve it, but it's even better with lower case, although finicky to edit:

1973 (Attitudes and Opinions of Senior Pupils: St Paul's Girls' School)

Before

```
VALUE LABELS    FORM
                    (1) LOWER   FIFTH
                    (2) UPPER   FIFTH
                    (3) LOWER   SIXTH
                    (4) UPPER   SIXTH
                /YEARBORN
                    (1) 1954
                    (2) 1955
                    (3) 1956
                    (4) 1957
                    (5) 1958
                /MONTH
                    (1) JANUARY
                    (2) FEBRUARY
                    (3) MARCH
                    (4) APRIL
                    (5)MAY
                    (6)JUNE
                    (7)JULY
                    (8)AUGUST
                    (9) SEPTEMBR
                    (10) OCTOBER
                    (11) NOVEMBER
                    (12) DECEMBER
                /VAR111 TO VAR119
                    (1) MOST   IMPORTNT
                    (2) NEITHER
                    (3) LEAST  IMPORTNT
                /JOB1 TO JOBAT25
                (1) ACCNTNCY,FINANCE
                (2) ARCHIT- ECTURE
                (3) CIVIL   ENGINEER
                (4) CREATIVE ARTIST
                (5) DOCTOR, DENTIST
                (6) FASHION
                (7) GOVNMNT,ADMIN.
                (8) HOUSE   -WIFE
                (9) INDUST. TECH.
                (10) JOURN-  ALISM
                (11) MILITARY SERVICE
                (12) NURSING
                (13) OUTDOOR,ATHLETIC
                (14) OWN     BUSINESS
                (15) PERFORM-ING ARTS
                (16) PERSONN-EL MNGMT
                (17) POLITICS
                (18) PUBLISH -ING
                (19) SALES + MARKETNG
                (20) SCIENCE-MATHS
                (21) SCIENCE-BIOLOGY
                (22) SCIENCE-SOCIAL
                (23) SECRET  -ARY
                (24) SOCIAL  WORK
                (25) SOLICTR,BARRISTR
                (26) TEACHER-PRIMARY
                (27) TEACHER-SECNDARY
                (28) TOWN    PLANNING
                (29) TV,FILM PRODUCER
                (30) UNIVSTY LECTURER
                (31) LIBRAR  -IAN
                (32) PUBLIC  RELATNS
                (33) COMP-   UTERS
                (34) OTHER
```

After

```
val lab   form
              (1) Lower   fifth
              (2) Upper   fifth
              (3) Lower   sixth
              (4) Upper   sixth
          /yearborn
              (1) 1954
              (2) 1955
              (3) 1956
              (4) 1957
              (5) 1958
          /month
              (1) January
              (2) February
              (3) March
              (4) April
              (5) May
              (6) June
              (7) July
              (8) August
              (9) Septembr
              (10) October
              (11) November
              (12) December
          /var111 to var119
              (1) Most   importnt
              (2) neither
              (3) Least  importnt
          /job1 to jobat25
              (1) Accntncy,finance
              (2) Archit- ecture
              (3) Civil   engineer
              (4) Creative artist
              (5) Doctor, dentist
              (6) Fashion
              (7) Govnmnt,admin.
              (8) House   -wife
              (9) Indust. tech.
              (10) Journ-  alism
              (11) Military service
              (12) Nursing
              (13) Outdoor,athletic
              (14) own     business
              (15) Perform-ing arts
              (16) Personn-el mngmt
              (17) Politics
              (18) Publish -ing
              (19) Sales + marketng
              (20) Science-maths
              (21) Science-biology
              (22) Science-social
              (23) Secret  -ary
              (24) Social  work
              (25) Solictr,barristr
              (26) Teacher-primary
              (27) Teacher-secndary
              (28) Town    planning
              (29) Tv,film producer
              (30) Univsty lecturer
              (31) LIbrar  -ian
              (32) Public  relatns
              (33) Comp-   uters
              (34) Other
```

3: European Social Survey 2002

3.1: Variable names and labels

As long as questionnaires were printed with data location information, positional naming of variables was fine, but modern developments such as **CAPI** and **BLISS** have now made life more difficult for the secondary researcher. For example, the questionnaire for the 2002 **European Social Survey** has **no data information** (other than response codes) printed on it.

Extract from questionnaire: European Social Survey 2002

A1 **CARD 1** On an average weekday, how much time, in total, do you spend watching television? Please use this card to answer.

No time at all	00	**GO TO A3**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK A2**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
(Don’t know)		

A2 **STILL CARD 1** And again on an average weekday, how much of your time watching television is spent watching **news** or programmes about **politics and current affairs**[34]? Still use this card.

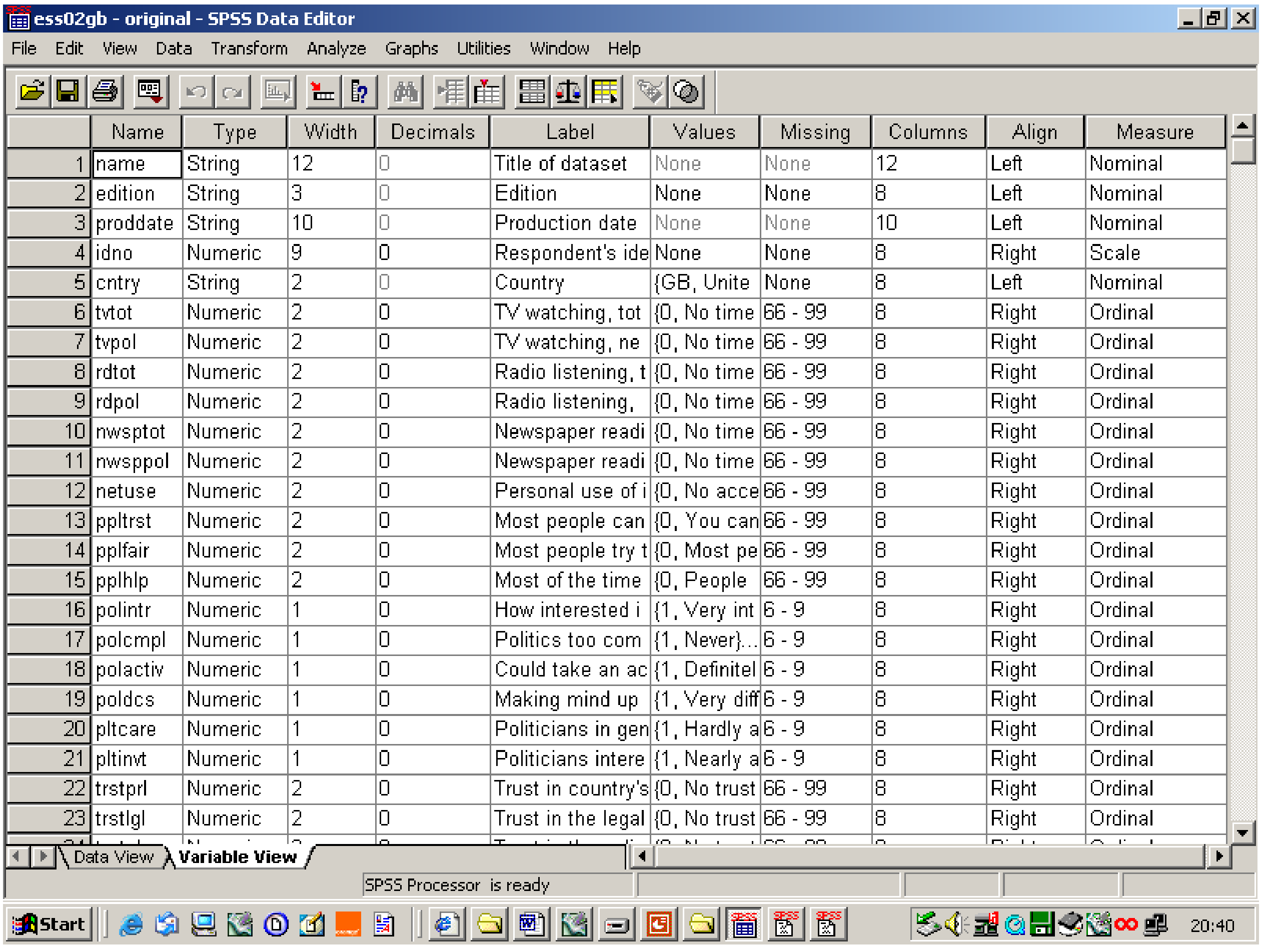
No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don’t know)	88

[NB No information on data layout or location in data set]

[34] **Prompted if necessary:** *About “politics and current affairs”: about issues to do with governance and public policy, and with the people connected with these affairs.*

European Social Survey 2002 - GB respondents only[35]

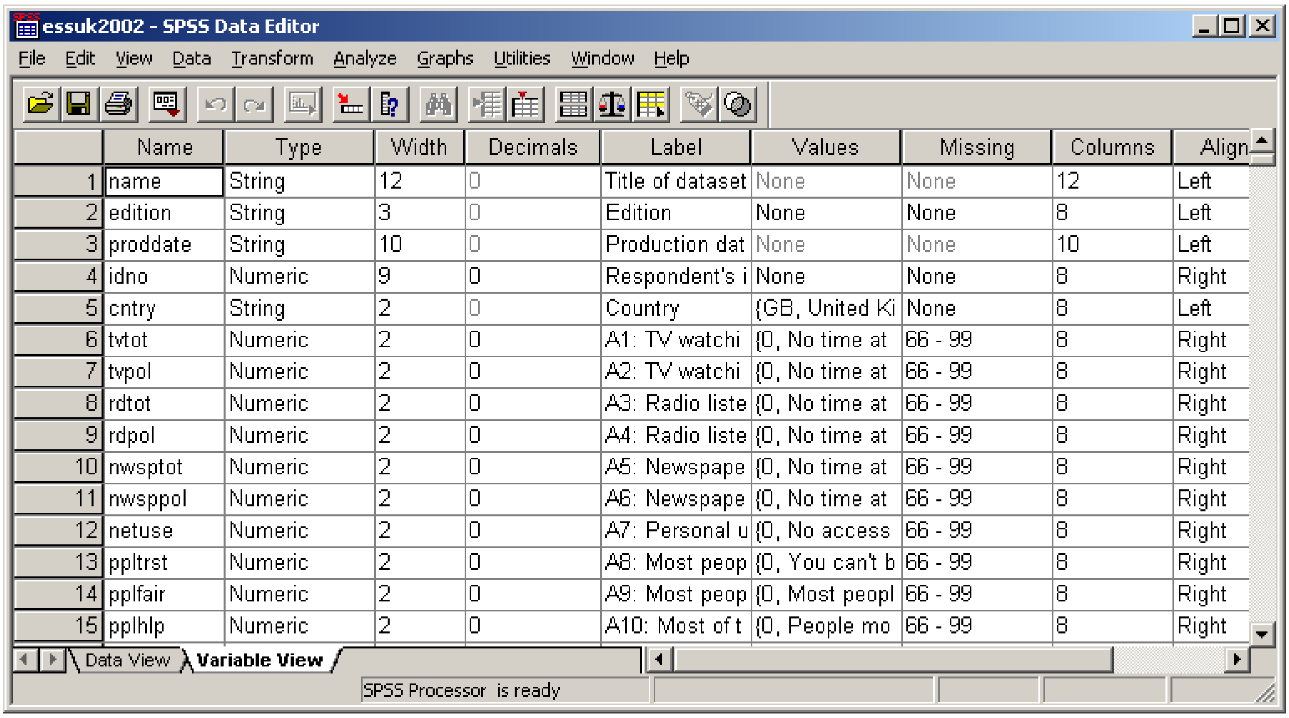
Data Editor as initialised:


ess02gb - original - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
Start
20:40

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	name	String	12	0	Title of dataset	None	None	12	Left	Nominal
2	edition	String	3	0	Edition	None	None	8	Left	Nominal
3	proddate	String	10	0	Production date	None	None	10	Left	Nominal
4	idno	Numeric	9	0	Respondent's ide	None	None	8	Right	Scale
5	cntry	String	2	0	Country	{GB, Unite	None	8	Left	Nominal
6	tvtot	Numeric	2	0	TV watching, tot	{0, No time	66 - 99	8	Right	Ordinal
7	tvpol	Numeric	2	0	TV watching, ne	{0, No time	66 - 99	8	Right	Ordinal
8	rdtot	Numeric	2	0	Radio listening, t	{0, No time	66 - 99	8	Right	Ordinal
9	rdpol	Numeric	2	0	Radio listening,	{0, No time	66 - 99	8	Right	Ordinal
10	nwsptot	Numeric	2	0	Newspaper readi	{0, No time	66 - 99	8	Right	Ordinal
11	nwsppol	Numeric	2	0	Newspaper readi	{0, No time	66 - 99	8	Right	Ordinal
12	netuse	Numeric	2	0	Personal use of i	{0, No acce	66 - 99	8	Right	Ordinal
13	ppltrst	Numeric	2	0	Most people can	{0, You can	66 - 99	8	Right	Ordinal
14	pplfair	Numeric	2	0	Most people try t	{0, Most pe	66 - 99	8	Right	Ordinal
15	pplhlp	Numeric	2	0	Most of the time	{0, People	66 - 99	8	Right	Ordinal
16	polintr	Numeric	1	0	How interested i	{1, Very int	6 - 9	8	Right	Ordinal
17	polcmpl	Numeric	1	0	Politics too com	{1, Never}...	6 - 9	8	Right	Ordinal
18	polactiv	Numeric	1	0	Could take an ac	{1, Definitel	6 - 9	8	Right	Ordinal
19	poldcs	Numeric	1	0	Making mind up	{1, Very diff	6 - 9	8	Right	Ordinal
20	pltcare	Numeric	1	0	Politicians in gen	{1, Hardly a	6 - 9	8	Right	Ordinal
21	pltinvt	Numeric	1	0	Politicians intere	{1, Nearly a	6 - 9	8	Right	Ordinal
22	trstprl	Numeric	2	0	Trust in country's	{0, No trust	66 - 99	8	Right	Ordinal
23	trstlgl	Numeric	2	0	Trust in the legal	{0, No trust	66 - 99	8	Right	Ordinal

SPSS Processor is ready

Data Editor after modification (to include question number at beginning of variable labels)


essuk2002 - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	name	String	12	0	Title of dataset	None	None	12	Left
2	edition	String	3	0	Edition	None	None	8	Left
3	proddate	String	10	0	Production dat	None	None	10	Left
4	idno	Numeric	9	0	Respondent's i	None	None	8	Right
5	cntry	String	2	0	Country	{GB, United Ki	None	8	Left
6	tvtot	Numeric	2	0	A1: TV watchi	{0, No time at	66 - 99	8	Right
7	tvpol	Numeric	2	0	A2: TV watchi	{0, No time at	66 - 99	8	Right
8	rdtot	Numeric	2	0	A3: Radio liste	{0, No time at	66 - 99	8	Right
9	rdpol	Numeric	2	0	A4: Radio liste	{0, No time at	66 - 99	8	Right
10	nwsptot	Numeric	2	0	A5: Newspape	{0, No time at	66 - 99	8	Right
11	nwsppol	Numeric	2	0	A6: Newspape	{0, No time at	66 - 99	8	Right
12	netuse	Numeric	2	0	A7: Personal u	{0, No access	66 - 99	8	Right
13	ppltrst	Numeric	2	0	A8: Most peop	{0, You can't b	66 - 99	8	Right
14	pplfair	Numeric	2	0	A9: Most peop	{0, Most peopl	66 - 99	8	Right
15	pplhlp	Numeric	2	0	A10: Most of t	{0, People mo	66 - 99	8	Right

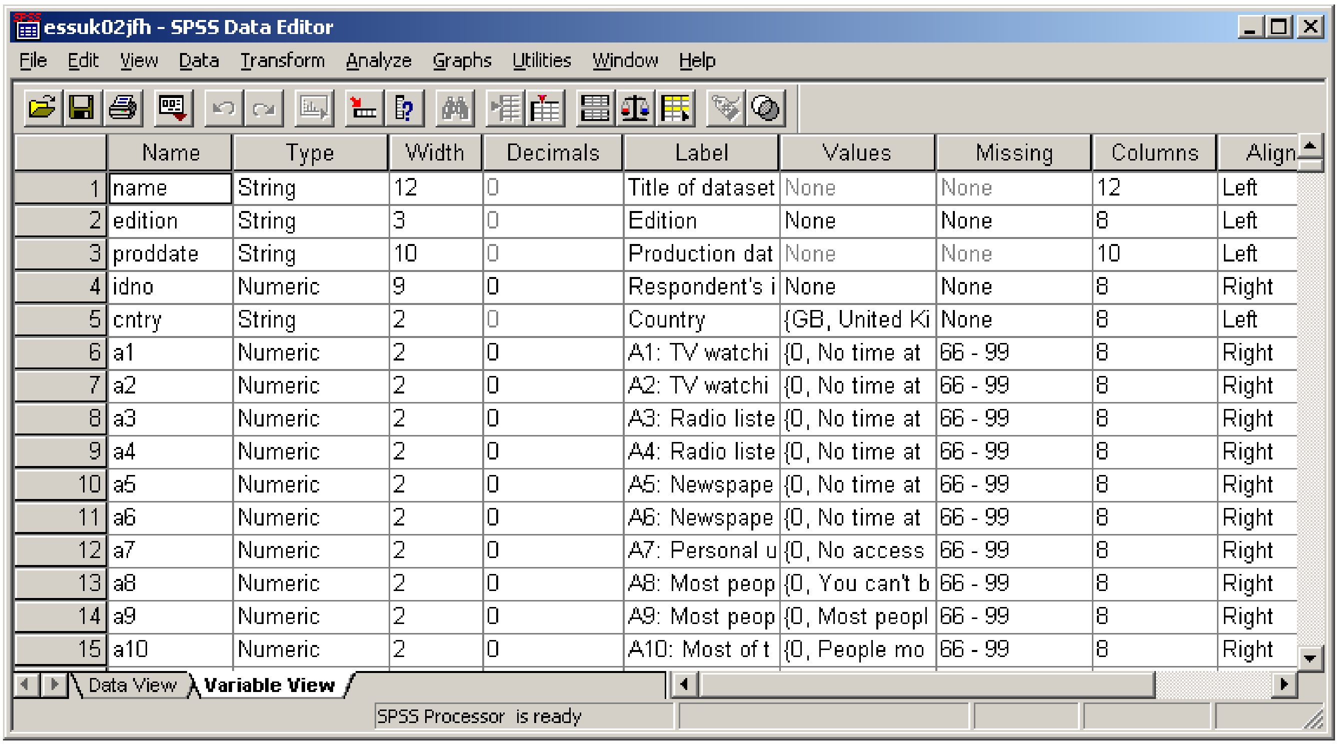
SPSS Processor is ready

[35] Extracted from downloadable SPSS saved file on http://ess.nsd.uib.no/index.jsp?module=main&country=

This version is fine for those who prefer, or need, to work with the original variable names, and the labels[36] are now a little clearer in the **Data Editor**.   However, to find our way round the file with the original questionnaire in front of us, we can make things even easier by using question numbers for some variable names. First we need to use **rename**:

```
rename variables
      (tvtot to pplhlp = a1 to a10)
      (polintr to vote prtvtgb = b1 to b13 b14gb)
      (contplt to ilglpst clsprty prtclgb = b15 to b24 b25a b25gb)
      (prtdgcl mmbprty prtmbgb = B25c b26 b27GB)
      (lrscale to scnsenv = b28 to b50)
      (happy to dscrgrp = c1 to c16)
      (dscrrce to dscroth  = c17_1 to c17_10)
      (dscrdk to dscrna =  c17_dk, c17_ref, c17_nap, c17_na)
      (ctzcntr to livecntr = c18 to c22)
      (lnghoma lnghomb = c23_1, c23_2)
      (blgetmg to mocntn = c24 to c28)
      (imgetn to imbghct = d1 to d31)
      ( ctbfsmv to stimrdt = d32 to d44)
      ( lwdscwp to blncmig = d45 to d58)
      (hlpppl to imprwct = e20 to e43)
      (hhmmb gndr yrbrn = f1, f2, f3)
      (domicil edulvl eduyrs = f5, f6, f7)
      (emplrel wkhct wkhtot = f12 f19 f20)
      (uemp3m to brwmny edulvlp = f25 to f32 f34)
      (edulvlf emprf14 occf14 occm14 = f45 f46 f51 f56)
      (marital to chldhhe = f58 to f65).
```

..to change the variable names as well. *[NB: The lines in **red** are for variables used in later examples]*


essuk02jfh - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

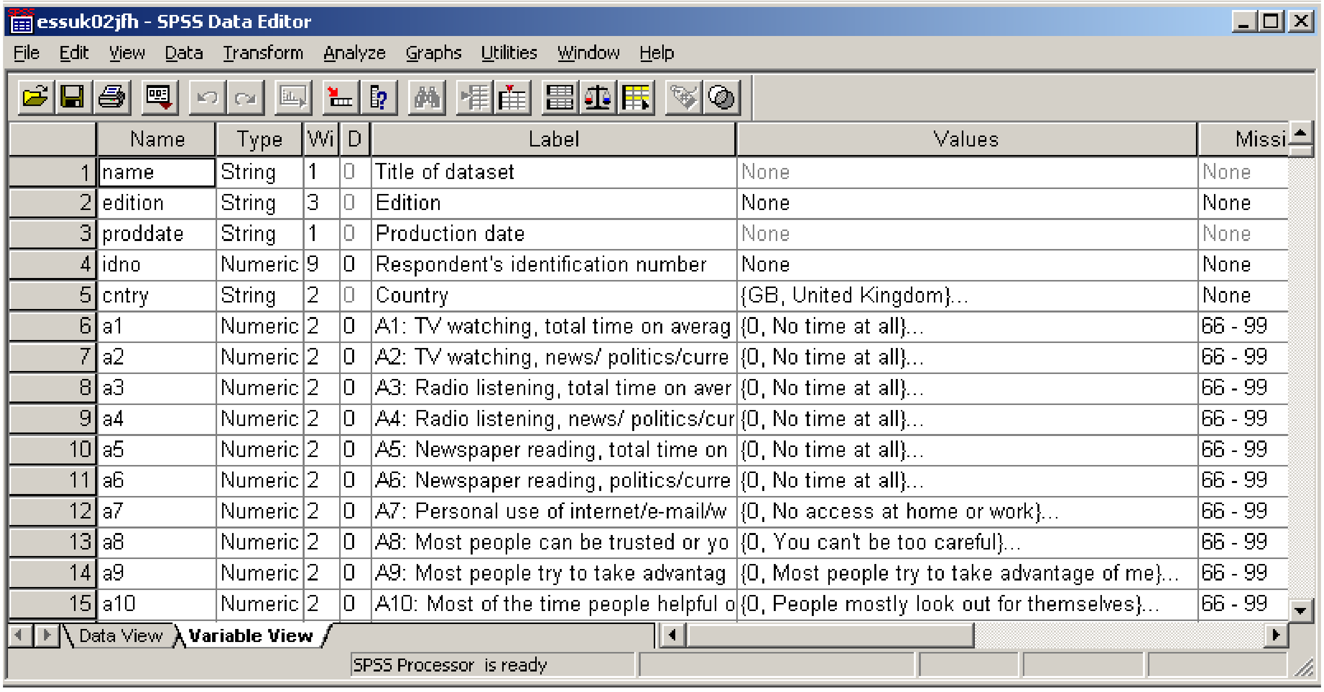
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	name	String	12	0	Title of dataset	None	None	12	Left
2	edition	String	3	0	Edition	None	None	8	Left
3	proddate	String	10	0	Production dat	None	None	10	Left
4	idno	Numeric	9	0	Respondent's i	None	None	8	Right
5	cntry	String	2	0	Country	{GB, United Ki	None	8	Left
6	a1	Numeric	2	0	A1: TV watchi	{0, No time at	66 - 99	8	Right
7	a2	Numeric	2	0	A2: TV watchi	{0, No time at	66 - 99	8	Right
8	a3	Numeric	2	0	A3: Radio liste	{0, No time at	66 - 99	8	Right
9	a4	Numeric	2	0	A4: Radio liste	{0, No time at	66 - 99	8	Right
10	a5	Numeric	2	0	A5: Newspape	{0, No time at	66 - 99	8	Right
11	a6	Numeric	2	0	A6: Newspape	{0, No time at	66 - 99	8	Right
12	a7	Numeric	2	0	A7: Personal u	{0, No access	66 - 99	8	Right
13	a8	Numeric	2	0	A8: Most peop	{0, You can't b	66 - 99	8	Right
14	a9	Numeric	2	0	A9: Most peop	{0, Most peopl	66 - 99	8	Right
15	a10	Numeric	2	0	A10: Most of t	{0, People mo	66 - 99	8	Right

SPSS Processor is ready

..on which you can't see everything properly, especially the labels.

[36] For the 2005 wave, the question number now appears at the beginning of the variable labels.

You don’t really need it all anyway, so once you’ve checked the file over, it helps to adjust the column widths to get rid of the redundant spaces in the **Type**, **Width**, and **Decimals** columns and maximise the text displayed for the variable labels (and increase it a bit for the value labels) to hide the other columns.


essuk02jfh - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready

	Name	Type	Wi	D	Label	Values	Missi
1	name	String	1	0	Title of dataset	None	None
2	edition	String	3	0	Edition	None	None
3	proddate	String	1	0	Production date	None	None
4	idno	Numeric	9	0	Respondent's identification number	None	None
5	cntry	String	2	0	Country	{GB, United Kingdom}...	None
6	a1	Numeric	2	0	A1: TV watching, total time on averag	{0, No time at all}...	66 - 99
7	a2	Numeric	2	0	A2: TV watching, news/ politics/curre	{0, No time at all}...	66 - 99
8	a3	Numeric	2	0	A3: Radio listening, total time on aver	{0, No time at all}...	66 - 99
9	a4	Numeric	2	0	A4: Radio listening, news/ politics/cur	{0, No time at all}...	66 - 99
10	a5	Numeric	2	0	A5: Newspaper reading, total time on	{0, No time at all}...	66 - 99
11	a6	Numeric	2	0	A6: Newspaper reading, politics/curre	{0, No time at all}...	66 - 99
12	a7	Numeric	2	0	A7: Personal use of internet/e-mail/w	{0, No access at home or work}...	66 - 99
13	a8	Numeric	2	0	A8: Most people can be trusted or yo	{0, You can't be too careful}...	66 - 99
14	a9	Numeric	2	0	A9: Most people try to take advantag	{0, Most people try to take advantage of me}...	66 - 99
15	a10	Numeric	2	0	A10: Most of the time people helpful o	{0, People mostly look out for themselves}...	66 - 99

3.2: An example of awkward labelling

The following section of the questionnaire seems clear enough:

ASK ALL

C16 Would you describe yourself as being a member of a group that is discriminated against in this country?

Yes	1	**ASK C17**
No	2	**GO TO C18**
(Don't know)	8	

C17 On what grounds is your group discriminated against? **PROBE:** 'What other grounds?'

CODE ALL THAT APPLY

Colour or race	01
Nationality	02
Religion	03
Language	04
Ethnic group	05
Age	06
Gender	07
Sexuality	08
Disability	09
Other (**WRITE IN**)___________________________	10
(Don't know)	88

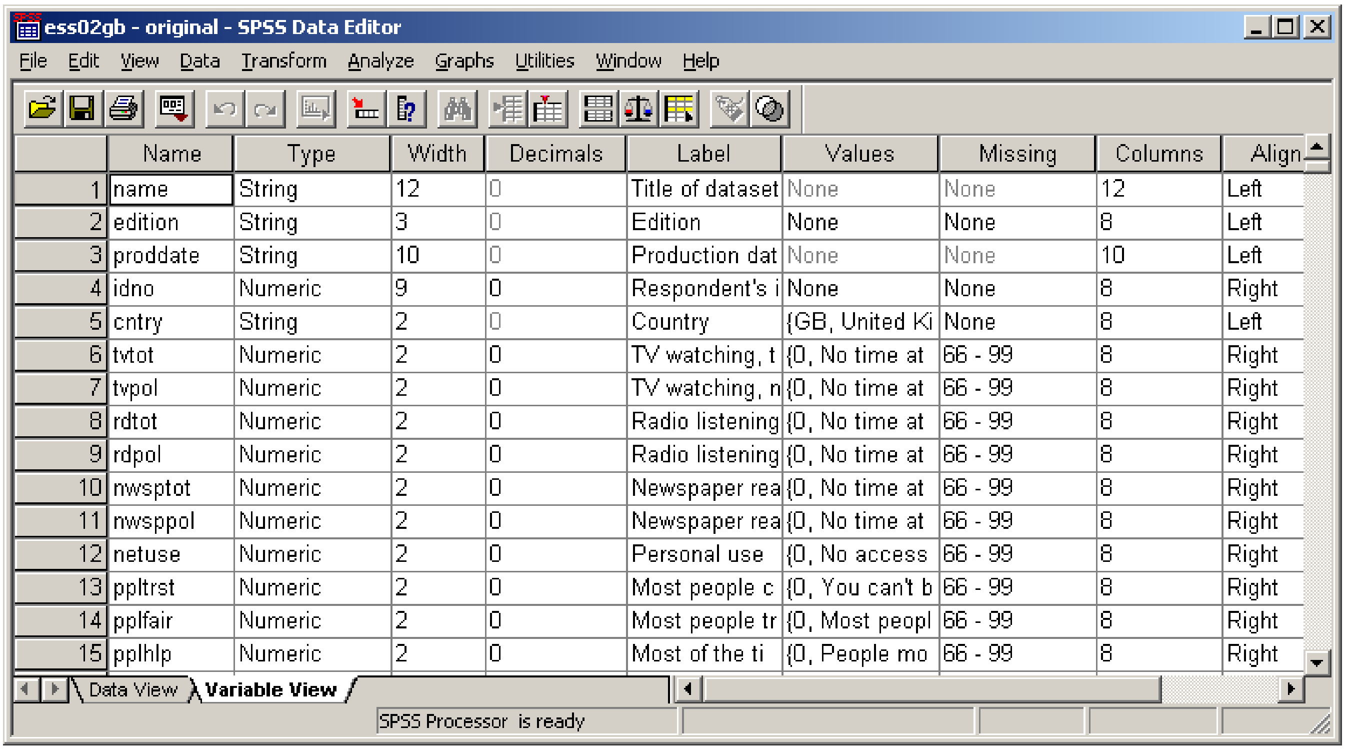
This is a simple filter question followed by a multiple response question for those who answered, "Yes". However, a secondary researcher wishing to analyse data from this question is faced with a problem. First, there is no indication of data layout; second, the variable names and labels in the original SPSS saved file distributed by ESS) illustrate the separate problems of using mnemonic variable names (so you don't know where they are in the file or to which question they relate) and long variable labels with no question number, redundant information at the beginning and all the useful information at the end, so it's masked unless you widen the **Labels** column). After I pointed this out, the European Social Survey now puts the question number at the beginning of variable labels, but has retained mnemonic variable names.

*[**NB:** The ESS data file has a separate variable for each response, but although the responses are precoded 01 to 10 on the questionnaire, the valid responses for each variable have been entered as binary (0,1) and the value labels as (0 = not marked, 1 = marked) and the missing values as 6-9. Now there's confusion for you! For some analysis, these binary values will need to be changed to sequential (you'll see why later).]*

It takes a while to find the associated variables for the above questionnaire extract, and when you do find them it's not immediately clear what they are.

Here's what I mean

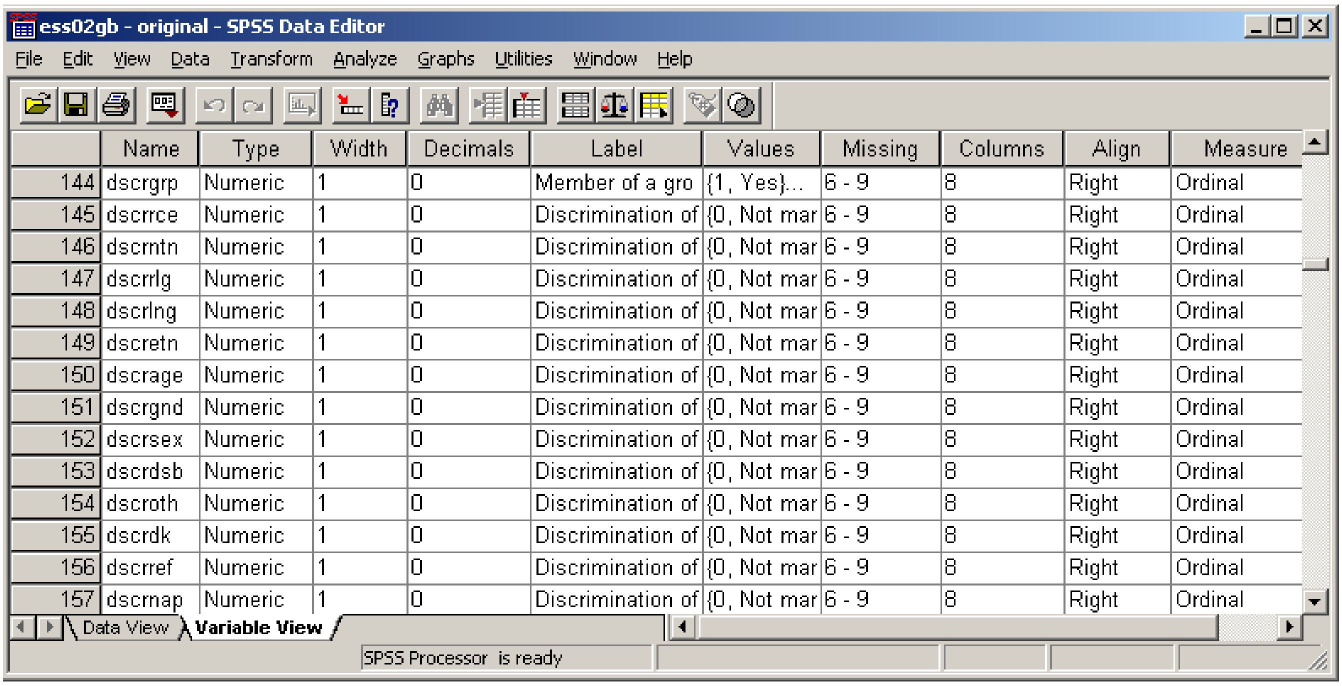
Data Editor as initialised


ess02gb - original - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	name	String	12	0	Title of dataset	None	None	12	Left
2	edition	String	3	0	Edition	None	None	8	Left
3	proddate	String	10	0	Production dat	None	None	10	Left
4	idno	Numeric	9	0	Respondent's i	None	None	8	Right
5	cntry	String	2	0	Country	{GB, United Ki	None	8	Left
6	tvtot	Numeric	2	0	TV watching, t	{0, No time at	66 - 99	8	Right
7	tvpol	Numeric	2	0	TV watching, n	{0, No time at	66 - 99	8	Right
8	rdtot	Numeric	2	0	Radio listening	{0, No time at	66 - 99	8	Right
9	rdpol	Numeric	2	0	Radio listening	{0, No time at	66 - 99	8	Right
10	nwsptot	Numeric	2	0	Newspaper rea	{0, No time at	66 - 99	8	Right
11	nwsppol	Numeric	2	0	Newspaper rea	{0, No time at	66 - 99	8	Right
12	netuse	Numeric	2	0	Personal use	{0, No access	66 - 99	8	Right
13	ppltrst	Numeric	2	0	Most people c	{0, You can't b	66 - 99	8	Right
14	pplfair	Numeric	2	0	Most people tr	{0, Most peopl	66 - 99	8	Right
15	pplhlp	Numeric	2	0	Most of the ti	{0, People mo	66 - 99	8	Right

SPSS Processor is ready

How do you find the relevant variables in this lot? Well, you can scroll down to see if there are any possible candidates, but you might not find them first time:

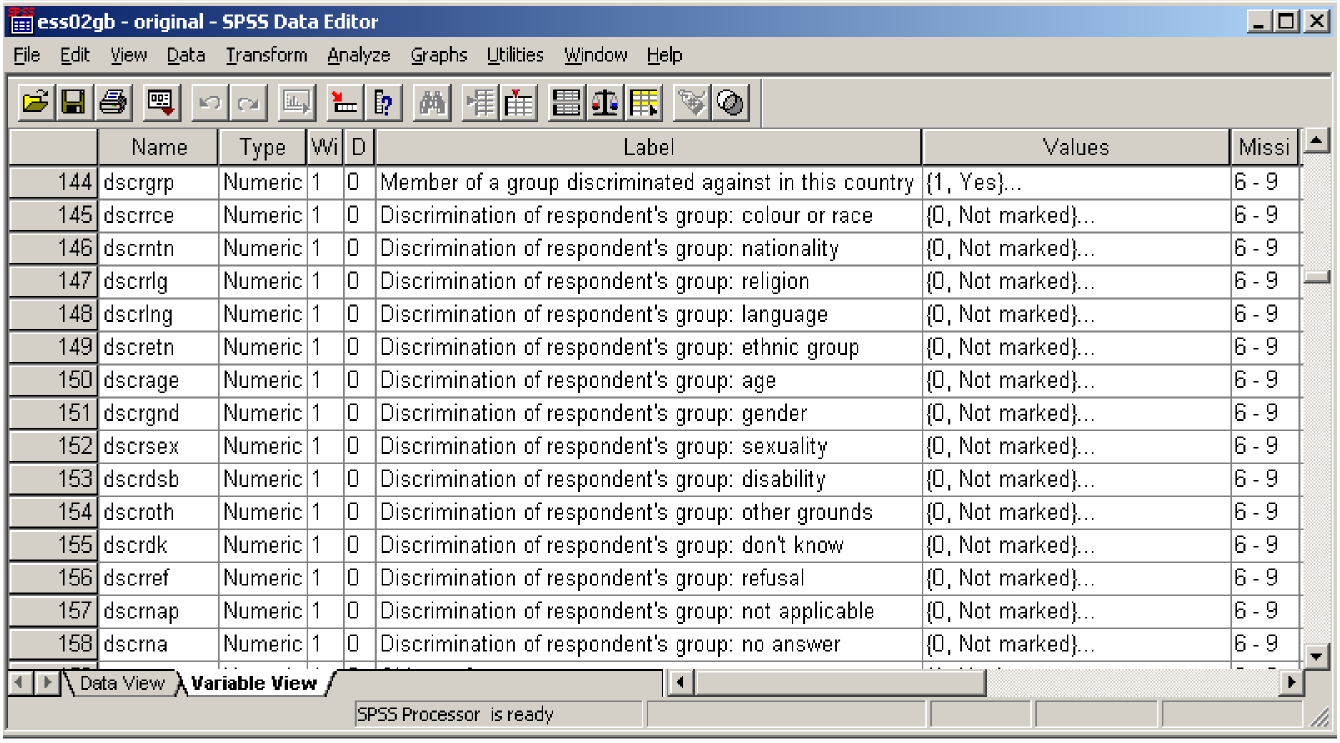

ess02gb - original - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
144	dscrgrp	Numeric	1	0	Member of a gro	{1, Yes}...	6 - 9	8	Right	Ordinal
145	dscrrce	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
146	dscrntn	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
147	dscrrlg	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
148	dscrlng	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
149	dscretn	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
150	dscrage	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
151	dscrgnd	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
152	dscrsex	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
153	dscrdsb	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
154	dscroth	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
155	dscrdk	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
156	dscrref	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal
157	dscrnap	Numeric	1	0	Discrimination of	{0, Not mar	6 - 9	8	Right	Ordinal

SPSS Processor is ready

It's better to start by adjusting the column widths as before, but this time widen the **Labels** column to reveal all the labels in full: then scroll down to search for likely variables. The first one is in row 144.

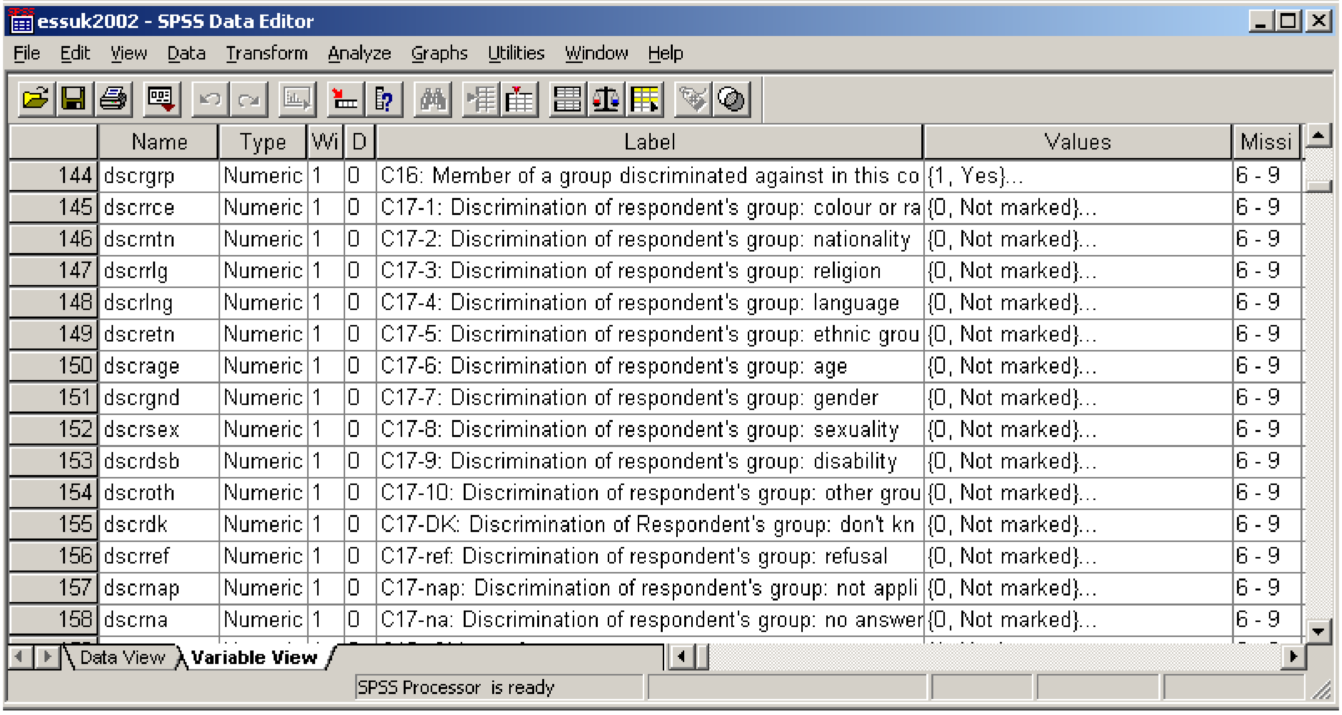
Data Editor after widening Labels column and scrolling down


ess02gb - original - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready

	Name	Type	Wi	D	Label	Values	Missi
144	dscrgrp	Numeric	1	0	Member of a group discriminated against in this country	{1, Yes}...	6 - 9
145	dscrrce	Numeric	1	0	Discrimination of respondent's group: colour or race	{0, Not marked}...	6 - 9
146	dscrntn	Numeric	1	0	Discrimination of respondent's group: nationality	{0, Not marked}...	6 - 9
147	dscrrlg	Numeric	1	0	Discrimination of respondent's group: religion	{0, Not marked}...	6 - 9
148	dscrlng	Numeric	1	0	Discrimination of respondent's group: language	{0, Not marked}...	6 - 9
149	dscretn	Numeric	1	0	Discrimination of respondent's group: ethnic group	{0, Not marked}...	6 - 9
150	dscrage	Numeric	1	0	Discrimination of respondent's group: age	{0, Not marked}...	6 - 9
151	dscrgnd	Numeric	1	0	Discrimination of respondent's group: gender	{0, Not marked}...	6 - 9
152	dscrsex	Numeric	1	0	Discrimination of respondent's group: sexuality	{0, Not marked}...	6 - 9
153	dscrdsb	Numeric	1	0	Discrimination of respondent's group: disability	{0, Not marked}...	6 - 9
154	dscroth	Numeric	1	0	Discrimination of respondent's group: other grounds	{0, Not marked}...	6 - 9
155	dscrdk	Numeric	1	0	Discrimination of respondent's group: don't know	{0, Not marked}...	6 - 9
156	dscrref	Numeric	1	0	Discrimination of respondent's group: refusal	{0, Not marked}...	6 - 9
157	dscrnap	Numeric	1	0	Discrimination of respondent's group: not applicable	{0, Not marked}...	6 - 9
158	dscrna	Numeric	1	0	Discrimination of respondent's group: no answer	{0, Not marked}...	6 - 9

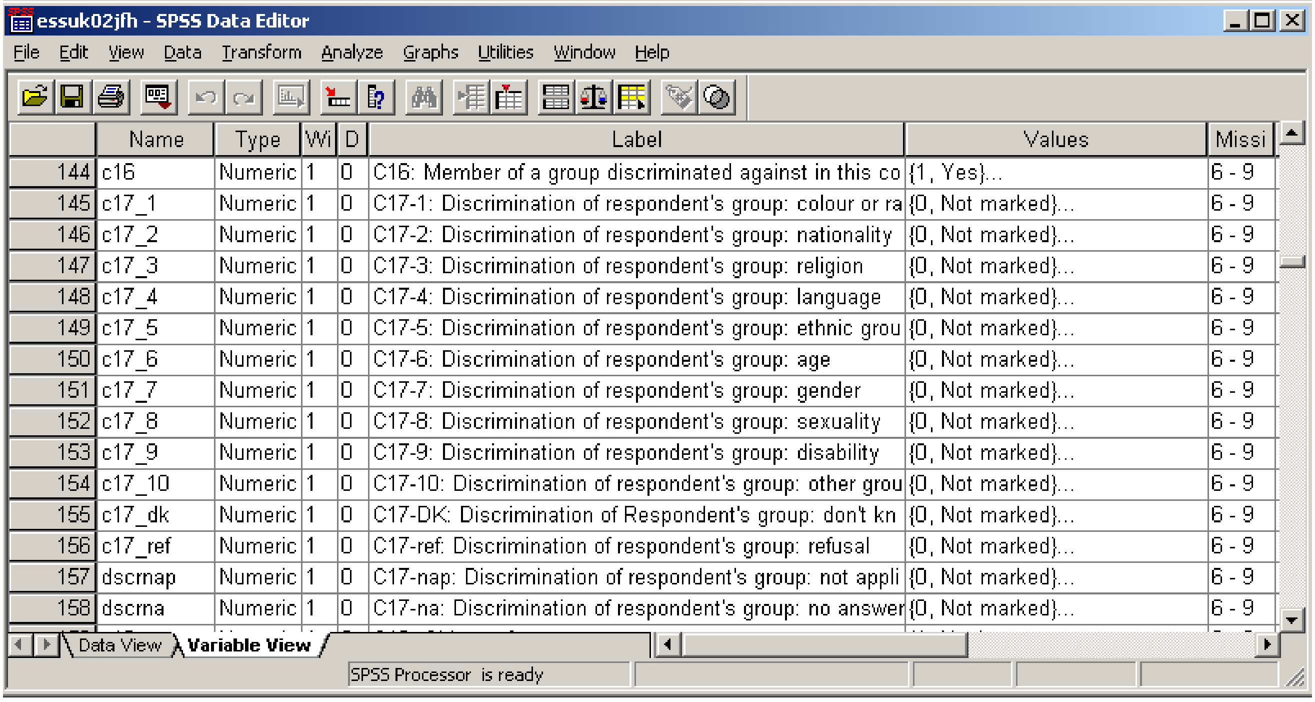
You can make things a little easier by inserting the question number and response code at the beginning of the label. It looks a bit messy, but makes the variables easier to find.

Data Editor after adding question number and response code to the beginning of the label, but **still with mnemonic variable names:**


essuk2002 - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready

	Name	Type	Wi	D	Label	Values	Missi
144	dscrgrp	Numeric	1	0	C16: Member of a group discriminated against in this co	{1, Yes}...	6 - 9
145	dscrrce	Numeric	1	0	C17-1: Discrimination of respondent's group: colour or ra	{0, Not marked}...	6 - 9
146	dscrntn	Numeric	1	0	C17-2: Discrimination of respondent's group: nationality	{0, Not marked}...	6 - 9
147	dscrrlg	Numeric	1	0	C17-3: Discrimination of respondent's group: religion	{0, Not marked}...	6 - 9
148	dscrlng	Numeric	1	0	C17-4: Discrimination of respondent's group: language	{0, Not marked}...	6 - 9
149	dscretn	Numeric	1	0	C17-5: Discrimination of respondent's group: ethnic grou	{0, Not marked}...	6 - 9
150	dscrage	Numeric	1	0	C17-6: Discrimination of respondent's group: age	{0, Not marked}...	6 - 9
151	dscrgnd	Numeric	1	0	C17-7: Discrimination of respondent's group: gender	{0, Not marked}...	6 - 9
152	dscrsex	Numeric	1	0	C17-8: Discrimination of respondent's group: sexuality	{0, Not marked}...	6 - 9
153	dscrdsb	Numeric	1	0	C17-9: Discrimination of respondent's group: disability	{0, Not marked}...	6 - 9
154	dscroth	Numeric	1	0	C17-10: Discrimination of respondent's group: other grou	{0, Not marked}...	6 - 9
155	dscrdk	Numeric	1	0	C17-DK: Discrimination of Respondent's group: don't kn	{0, Not marked}...	6 - 9
156	dscrref	Numeric	1	0	C17-ref: Discrimination of respondent's group: refusal	{0, Not marked}...	6 - 9
157	dscrnap	Numeric	1	0	C17-nap: Discrimination of respondent's group: not appli	{0, Not marked}...	6 - 9
158	dscrna	Numeric	1	0	C17-na: Discrimination of respondent's group: no answer	{0, Not marked}...	6 - 9

You can use **rename variables** to change the variable names to match the question numbers, but there's still far too much redundant information at the beginning of the variable labels.

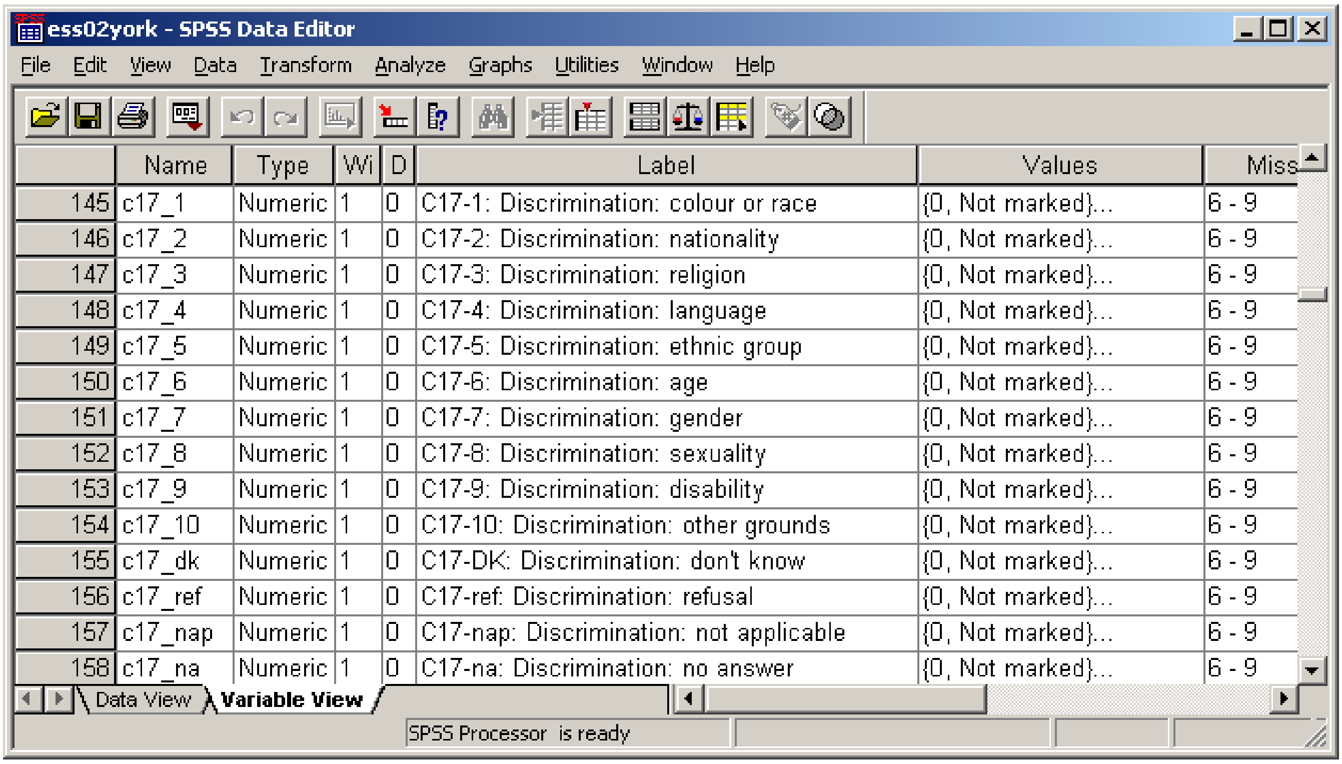

essuk02jfh - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Wi	D	Label	Values	Missi
144	c16	Numeric	1	0	C16: Member of a group discriminated against in this co	{1, Yes}...	6 - 9
145	c17_1	Numeric	1	0	C17-1: Discrimination of respondent's group: colour or ra	{0, Not marked}...	6 - 9
146	c17_2	Numeric	1	0	C17-2: Discrimination of respondent's group: nationality	{0, Not marked}...	6 - 9
147	c17_3	Numeric	1	0	C17-3: Discrimination of respondent's group: religion	{0, Not marked}...	6 - 9
148	c17_4	Numeric	1	0	C17-4: Discrimination of respondent's group: language	{0, Not marked}...	6 - 9
149	c17_5	Numeric	1	0	C17-5: Discrimination of respondent's group: ethnic grou	{0, Not marked}...	6 - 9
150	c17_6	Numeric	1	0	C17-6: Discrimination of respondent's group: age	{0, Not marked}...	6 - 9
151	c17_7	Numeric	1	0	C17-7: Discrimination of respondent's group: gender	{0, Not marked}...	6 - 9
152	c17_8	Numeric	1	0	C17-8: Discrimination of respondent's group: sexuality	{0, Not marked}...	6 - 9
153	c17_9	Numeric	1	0	C17-9: Discrimination of respondent's group: disability	{0, Not marked}...	6 - 9
154	c17_10	Numeric	1	0	C17-10: Discrimination of respondent's group: other grou	{0, Not marked}...	6 - 9
155	c17_dk	Numeric	1	0	C17-DK: Discrimination of Respondent's group: don't kn	{0, Not marked}...	6 - 9
156	c17_ref	Numeric	1	0	C17-ref: Discrimination of respondent's group: refusal	{0, Not marked}...	6 - 9
157	dscrnap	Numeric	1	0	C17-nap: Discrimination of respondent's group: not appli	{0, Not marked}...	6 - 9
158	dscrna	Numeric	1	0	C17-na: Discrimination of respondent's group: no answer	{0, Not marked}...	6 - 9

SPSS Processor is ready

An alternative solution[37] (which would match the positions of variables in the file) would be to use variable names derived from the row they are on ie **V144** to **V158** The row numbers will change if you start selecting variables to save in another file, but at least they'll be in same order and easier to find.

Now chop out redundant information at the beginning of the variable labels to yield:


ess02york - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Wi	D	Label	Values	Miss
145	c17_1	Numeric	1	0	C17-1: Discrimination: colour or race	{0, Not marked}...	6 - 9
146	c17_2	Numeric	1	0	C17-2: Discrimination: nationality	{0, Not marked}...	6 - 9
147	c17_3	Numeric	1	0	C17-3: Discrimination: religion	{0, Not marked}...	6 - 9
148	c17_4	Numeric	1	0	C17-4: Discrimination: language	{0, Not marked}...	6 - 9
149	c17_5	Numeric	1	0	C17-5: Discrimination: ethnic group	{0, Not marked}...	6 - 9
150	c17_6	Numeric	1	0	C17-6: Discrimination: age	{0, Not marked}...	6 - 9
151	c17_7	Numeric	1	0	C17-7: Discrimination: gender	{0, Not marked}...	6 - 9
152	c17_8	Numeric	1	0	C17-8: Discrimination: sexuality	{0, Not marked}...	6 - 9
153	c17_9	Numeric	1	0	C17-9: Discrimination: disability	{0, Not marked}...	6 - 9
154	c17_10	Numeric	1	0	C17-10: Discrimination: other grounds	{0, Not marked}...	6 - 9
155	c17_dk	Numeric	1	0	C17-DK: Discrimination: don't know	{0, Not marked}...	6 - 9
156	c17_ref	Numeric	1	0	C17-ref: Discrimination: refusal	{0, Not marked}...	6 - 9
157	c17_nap	Numeric	1	0	C17-nap: Discrimination: not applicable	{0, Not marked}...	6 - 9
158	c17_na	Numeric	1	0	C17-na: Discrimination: no answer	{0, Not marked}...	6 - 9

SPSS Processor is ready

It looks a bit messy, but at least you can now find them more easily.

[37] The 2005 wave of ESS has the mnemonic names actually displayed on the (electronic) questionnaire, so that may help a bit, but it's very confusing to look at.

How do we analyse this question?

3.3: Multiple response

You could run separate frequency counts for each variable, and then add them all up, but it's far better to use the SPSS command **MULT RESPONSE**

Question **C17** allows for more than one response and most analysis will therefore need to use the SPSS procedure **mult response**. On the questionnaire the coding for question **C17** ranges from 1 to 10 (with 88 as missing) but the actual data are entered as binary (0 or 1) with a separate column entry for each possible response, and also for some categories of non-response not shown ("No answer", "Not applicable", "Refused").

The **mult response** command generates **temporary group variables** (which cannot be saved). It works either in **dichotomous mode** (using a **single** value across all variables in the group) which displays **variable labels**, or in **general mode** (using a specified **range** of **different** values across all variables in the group) for which **value labels** will be displayed.   In either case, SPSS limits all labels to 40 printed characters in frequency counts and limits value labels to 20 for row variables and 16 for column variables in contingency tables. Because of this restraint (unchanged since the the procedure was first introduced in the 1970s) valuable information can be lost, or may not appear on the output.

Some examples will illustrate this. The first, because of the way the original data file was generated, uses multiple response in **dichotomous mode** on the original file, the second and third do the same thing, but with the modified variable names and labels described above. The fourth involves some (temporary and complex) file manipulation and the addition of value labels.

To run **mult response** on the **original** data file, we write:

```
mult response groups =
   discrim 'Reasons for perceived discrimination'
   (dscrrce to dscrna (1))
   /freq discrim.
```

```
Group DISCRIM  Reasons for perceived discrimination
     (Value tabulated = 1)

                                                              Pct of  Pct of
Dichotomy label                           Name        Count  Responses  Cases

Discrimination of respondent's group: co DSCRRCE        82      3.8      4.0
Discrimination of respondent's group: na DSCRNTN        28      1.3      1.4
Discrimination of respondent's group: re DSCRRLG        44      2.0      2.1
Discrimination of respondent's group: la DSCRLNG         5       .2       .2
Discrimination of respondent's group: et DSCRETN        21      1.0      1.0
Discrimination of respondent's group: ag DSCRAGE        50      2.3      2.4
Discrimination of respondent's group: ge DSCRGND        37      1.7      1.8
Discrimination of respondent's group: se DSCRSEX        18       .8       .9
Discrimination of respondent's group: di DSCRDSB        18       .8       .9
Discrimination of respondent's group: ot DSCROTH        74      3.4      3.6
Discrimination of respondent's group: do DSCRDK          1       .0       .0
Discrimination of respondent's group: re DSCRREF         1       .0       .0
Discrimination of respondent's group: no DSCRNAP      1771     82.4     86.3
                                                    -------    -----    -----
                                 Total responses      2150    100.0    104.8

0 missing cases;  2,052 valid cases
```

For frequency counts, variable labels in **mult response** are limited to 40 characters.   As we can see, this has caused the actual perceived reasons for discrimination to be completely lost, and it's not much clearer even after modifying the value labels by adding question number and response code :

```
Group DISCRIM  Reasons for perceived discrimination
     (Value tabulated = 1)

                                                          Pct of  Pct of
Dichotomy label                          Name       Count  Responses  Cases

C17-1: Discrimination of respondent's gr DSCRRCE       82      3.8      4.0
C17-2: Discrimination of respondent's gr DSCRNTN       28      1.3      1.4
C17-3: Discrimination of respondent's gr DSCRRLG       44      2.0      2.1
C17-4: Discrimination of respondent's gr DSCRLNG        5       .2       .2
C17-5: Discrimination of respondent's gr DSCRETN       21      1.0      1.0
C17-6: Discrimination of respondent's gr DSCRAGE       50      2.3      2.4
C17-7: Discrimination of respondent's gr DSCRGND       37      1.7      1.8
C17-8: Discrimination of respondent's gr DSCRSEX       18       .8       .9
C17-9: Discrimination of respondent's gr DSCRDSB       18       .8       .9
C17-10: Discrimination of respondent's g DSCROTH       74      3.4      3.6
C17-DK: Discrimination of Respondent's g DSCRDK         1       .0       .0
C17-ref: Discrimination of respondent's  DSCRREF        1       .0       .0
C17-nap: Discrimination of respondent's  DSCRNAP     1771     82.4     86.3
                                                  -------    -----    -----
                                 Total responses     2150    100.0    104.8

0 missing cases;  2,052 valid cases
```

What is needed is a modification of the labels to chop out some redundant information and bring forward the substantive part of the coding frame to yield something like this:

```
Group DISCRIM  Reasons for perceived discrimination
     (Value tabulated = 1)

                                                          Pct of  Pct of
Dichotomy label                          Name       Count  Responses  Cases

C17-1: Discrimination: colour or race    C17_1         82      3.8      4.0
C17-2: Discrimination: nationality       C17_2         28      1.3      1.4
C17-3: Discrimination: religion          C17_3         44      2.0      2.1
C17-4: Discrimination: language          C17_4          5       .2       .2
C17-5: Discrimination: ethnic group      C17_5         21      1.0      1.0
C17-6: Discrimination: age               C17_6         50      2.3      2.4
C17-7: Discrimination: gender            C17_7         37      1.7      1.8
C17-8: Discrimination: sexuality         C17_8         18       .8       .9
C17-9: Discrimination: disability        C17_9         18       .8       .9
C17-10: Discrimination: other grounds    C17_10        74      3.4      3.6
C17-DK: Discrimination: don't know       C17_DK         1       .0       .0
C17-ref: Discrimination: refusal         C17_REF        1       .0       .0
C17-nap: Discrimination: not applicable  DSCRNAP     1771     82.4     86.3
                                                  -------    -----    -----
                                 Total responses     2150    100.0    104.8

0 missing cases;  2,052 valid cases
```

…which now looks odd because the variable names appear twice, once in the variable list and once at the beginning of the variable labels. We could always change the variable labels back to knock off the question number, but there's an altogether better way of analysing the responses to this question using **mult response** in **general mode**.

This requires some complex manipulation of the data file to change the values of the variables from binary to sequential, to suppress the missing values (otherwise cases with values 6 to 9 for responses “Age”, “Gender”, “Sexuality” and “Disability” will be left out) and to change the value labels. If we wish to keep the original values and missing values, the recodes need to be **temporary**. Let’s assume we’re starting from scratch.

Step 1: Check initial values
Step 2: Temporarily change the codes from binary to sequential
Step 3: Check recoded values
Step 4: Disable missing values
Step 5: Change value labels (first variable only)
Step 6: Run mult response in general mode

It is good research practice to check all data before running any statistical analysis. A useful procedure is **LIST** (not available in the drop-down menus) which displays the data values for (all or selected) variables and cases. It can be used to check data values before and after recoding to make sure you’ve produced what you actually want. In this example we shall inspect the initial data values for variables **c17-1** to **c17-10**, and again after recoding them from binary to sequential values.

Step 1: Check initial values

```
list var c17_1 to C17_10
     / cases 5.
```

```
C17_1 C17_2 C17_3 C17_4 C17_5 C17_6 C17_7 C17_8 C17_9 C17_10

    1     0     0     0     0     0     1     0     0      0
    1     1     1     1     1     0     0     0     0      0
    1     0     0     0     0     0     0     0     0      0
    1     0     0     0     0     0     0     0     0      0
    0     0     0     0     0     0     0     0     0      1

  Number of cases read:  5    Number of cases listed:  5
```

Step 2: Temporarily change the codes from binary to sequential

```
temp.

recode
     c17_1  to c17_10 (6 thru hi = sysmis)
     /c17_2  (1=2)
     /c17_3  (1=3)
     /c17_4  (1=4)
     /c17_5  (1=5)
     /c17_6  (1=6)
     /c17_7  (1=7)
     /c17_8  (1=8)
     /c17_9  (1=9)
     /c17_10 (1=10)
     /c17_dk (1=11)
     /c17_ref (1=12)
     /c17_nap (1=13)
     /c17_na (1=14).
```

Step 3: Check recoded values

```
list var c17_1 to C17_10
     / cases 5.
```

```
C17_1 C17_2 C17_3 C17_4 C17_5 C17_6 C17_7 C17_8 C17_9 C17_10

    1     0     0     0     0     0     7     0     0      0
    1     2     3     4     5     0     0     0     0      0
    1     0     0     0     0     0     0     0     0      0
    1     0     0     0     0     0     0     0     0      0
    0     0     0     0     0     0     0     0     0      *

Number of cases read:   5    Number of cases listed:   5
```

*[**NB:** the * = value 10: it would display in full with print format F2.0]*

Step 4: Disable missing values

```
missing values
     c17_1 to c17_14 ( ).
```

(not many people know that trick, but don’t save the file or you’ll lose the lot!)

Step 5: Specify new value labels

```
value labels c17_1
     (1) 'Colour or race'
     (2) 'Nationality'
     (3) 'Religion'
     (4) 'Language'
     (5) 'Ethnic group'
     (6) 'Age'
     (7) 'Gender'
     (8) 'Sexuality'
     (9) 'Disability'
     (10)  'Other'
     (11) "Don't know"
     (12) 'Refusal'
     (13) 'Not applicable'
     (14) 'No answer'.
```

*[**NB:** in **general** mode SPSS* **mult response** *reads labels from **first variable only**]*

Step 6: Specify group variable and get frequency count

```
Mult response groups =
        discrim 'Q17 Perceived reasons for discrimination'
        (c17_1 to c17_14 (1,14))
        /freq discrim.
```

```
Group DISCRIM  Q17 Perceived reasons for discrimination
                                                           Pct of  Pct of
Category label                            Code      Count Responses  Cases

Colour or race                               1        82      3.8     4.0
Nationality                                  2        28      1.3     1.4
Religion                                     3        44      2.0     2.1
Language                                     4         5       .2      .2
Ethnic group                                 5        21      1.0     1.0
Age                                          6        50      2.3     2.4
Gender                                       7        37      1.7     1.8
Sexuality                                    8        18       .8      .9
Disability                                   9        18       .8      .9
Other                                       10        74      3.4     3.6
Don't know                                  11         1       .0      .0
Refusal                                     12         1       .0      .0
Not applicable                              13      1771     82.4    86.3
                                                 --------    -----   -----
                               Total responses      2150    100.0   104.8

0 missing cases;  2,052 valid cases
```

To restrict analysis to those who answered, “Yes” to question **c16**, use **select if (c16 = 1).** which will include the “Don’t know” and “Refused” (11 and 12). To exclude the latter, change the specification to:

```
mult response groups =
    discrim 'Q17 Perceived reasons for discrimination'
    (c17_1 to c17_10 (1,10))
    /freq discrim.
```

```
Group DISCRIM  C17 Perceived reasons for discrimination
                                                           Pct of  Pct of
Category label                            Code      Count Responses  Cases

Colour or race                               1        82     21.8    29.4
Nationality                                  2        28      7.4    10.0
Religion                                     3        44     11.7    15.8
Language                                     4         5      1.3     1.8
Ethnic group                                 5        21      5.6     7.5
Age                                          6        50     13.3    17.9
Gender                                       7        37      9.8    13.3
Sexuality                                    8        18      4.8     6.5
Disability                                   9        18      4.8     6.5
Other                                       10        74     19.6    26.5
                                                 --------    -----   -----
                               Total responses       377    100.0   135.1

1,773 missing cases;  279 valid cases
```

The SPSS job to produce the above output is only one way of doing it. An alternative to **recode** is:

```
do repeat
       x = (c17_1 to c17_na)
       /y = 1 to 14.
compute x=y .
end repeat.
```

4: Syntax or drop-down menus?

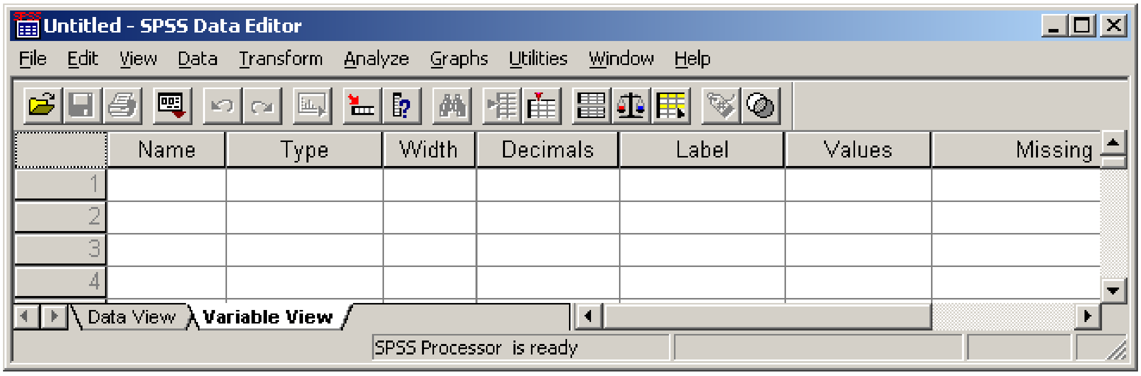
4.1: Data input

Can you do data entry to SPSS with drop-down menus? Not really. You have to use **File… New… Data** to open up a blank **Data Editor** in **Data View** and then type your data into the blank matrix. If you already have some data in a file and wish to add more, open the **Data Editor** select **Data View** click on **Data… Insert Cases** and type in your new data in **rows**. Or, if you prefer, you can also use **Data… Insert Variables** and type in your data in **columns**. Outside SPSS you will need to use a spreadsheet (SPSS can import from Excel), or possibly a word-processing package to type data in a fixed-width font (eg Courier New) and save the file as ***.dat** (assumed to be WordPerfect) or ***.txt** (in Word). Much raw data from surveys deposited with the UK Data Archive (Essex University) is supplied as Wordfect ***.dat** files in **Times New Roman** font. This proportional font is impossible to inspect visually and needs to be changed to **Courier New** to get the columns properly aligned. Typing your own data is prone to error, especially for large data sets. You have been warned!

Your data may already be in an existing SPSS or Excel file, in which case you can import them, but if they are raw (Hollerith type) data in an external file, you will have to use the **data list** command in syntax direct. You can use the mouse to drag and drop lines of data between **begin data ~ ~ ~ end data**, which is how I started using SPSS for Windows, when it consistently failed to find a raw data file in the directory I was working in. I did eventually discover that SPSS needs the whole address (eg **"C:\Documents and Settings\JFH\Desktop\Social Research\British Social Attitudes\bsa89 SCPR version\bsa89 essex version\bsa89.dat" !!!)** but it was quicker to copy the raw data to a blank 1.4mb floppy disk in drive **a:** and read the data from there (eg as **'a:QL1UK.dat'**.). This was fine for small data sets, but not for the huge (3.5mb +) files from **British Social Attitudes** and **European Social Survey**, which are too big. I used to drag and drop these into the syntax file (it took forever!) but I now have a 1gb memory stick (permanently plugged in as drive **f:**) and copy all raw data files to this. I can then read them easily into SPSS as external files (with nice short names!).

In this example the data are in **bsa86.dat** on drive **f:** and the external file name has to be declared to SPSS enclosed in single primes, ie **'f:bsa86.dat'**

Open a new Data Editor and adjust the columns:


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Width	Decimals	Label	Values	Missing
1							
2							
3							
4							

SPSS Processor is ready

Now click on **File … New … Syntax**, type in your SPSS commands for a **title** and the **data list** using **positional** variable names:

```
title 'Page 43b of BSA 1986'.
data list file 'f:bsa86.dat' records 23
    /15 v1508 8-9 v1510 10 v1511 11 v1512 12-13.
```

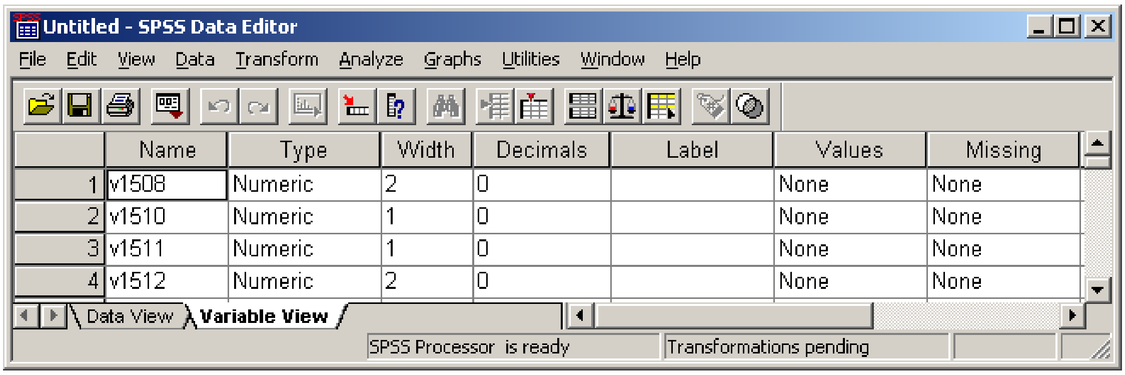
..and type **[CTRL]+R** or click on **RUN** to produce a data table (on which it's easy to check the accuracy of the specifications as the names and start column should match up).

```
Page 43b of BSA 1986
Data List will read 23 records from F:\bsa86.dat

Variable    Rec   Start     End        Format

V1508        15       8       9        F2.0
V1510        15      10      10        F1.0
V1511        15      11      11        F1.0
V1512        15      12      13        F2.0
```

The **Data File** also fills up...


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready
Transformations pending

	Name	Type	Width	Decimals	Label	Values	Missing
1	v1508	Numeric	2	0		None	None
2	v1510	Numeric	1	0		None	None
3	v1511	Numeric	1	0		None	None
4	v1512	Numeric	2	0		None	None

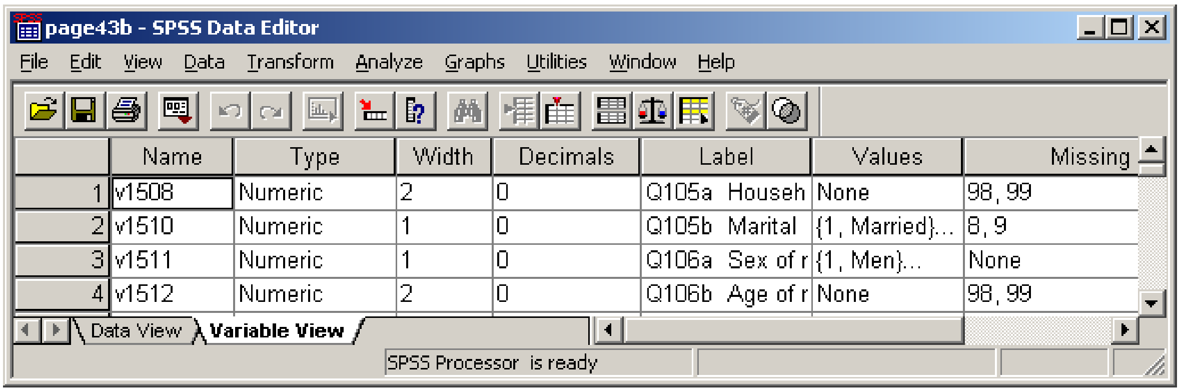
My advice would be always to construct your file definitions in direct syntax. It's also much easier to keep visual track of what you're doing, especially for beginners, if you you use tabs to inset continuation lines. Syntax files are much easier to check, amend or correct later. If yours is a large data set, it's probably better to do it in two or three or even more stages. For this example, type in:

```
missing values
    v1508 v1512 (98,99) v1510 (8,9).

var labels
    v1508 'Q105a  Household size'
    v1510 'Q105b  Marital status'
    v1511 'Q106a  Sex of respondent'
    v1512 'Q106b  Age of respondent last birthday'.

value labels
    v1510 1 'Married' 2 'Living  together' 3 'Sep or  div'
            4 'Widowed' 5 'Not     married' 8 'DK' 9 'N/A'
   /v1511 1 'Men' 2 'Women'.
```

..and **[CTRL]+R** or RUN to get:


page43b - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready

	Name	Type	Width	Decimals	Label	Values	Missing
1	v1508	Numeric	2	0	Q105a Househ	None	98, 99
2	v1510	Numeric	1	0	Q105b Marital	{1, Married}...	8, 9
3	v1511	Numeric	1	0	Q106a Sex of r	{1, Men}...	None
4	v1512	Numeric	2	0	Q106b Age of r	None	98, 99

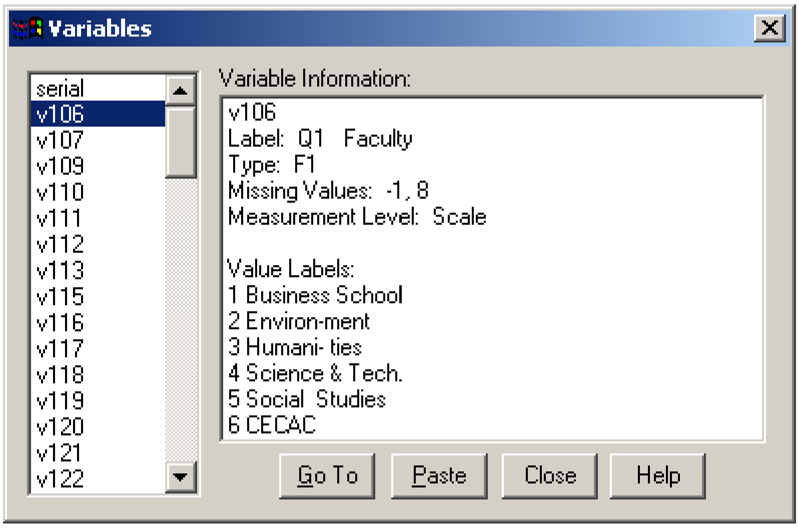
To save **page43b.sav** to drive **f:** for future use **File… Save as…**

or type: **save out 'f:page43b.sav'.** and then **[CTRL]+R** or **RUN**

4.2: Utilities

When working on a file, even a small one containing your own data, it is useful to have available a printed summary of the file contents. When (as in the following example from the **Polytechnic of North London Course Evaluation Survey 1986**) it's a large file and possibly not your own data, such summaries are essential.

Although you can get a quick check by sliding to each variable in the file from: **Utilities … Variables…**



Variable Information:

```
v106
Label:  Q1   Faculty
Type:  F1
Missing Values:  -1, 8
Measurement Level:  Scale

Value Labels:
1 Business School
2 Environ-ment
3 Humani- ties
4 Science & Tech.
5 Social  Studies
6 CECAC
```

the output from: **Utilities..**
File info..

```
              List of variables on the working file
Name                                                                Position

SERIAL     Serial number                                                   1
           Measurement Level:  Scale
           Column Width:  8  Alignment:  Right
           Print Format:  F4
           Write Format:  F4

V106       Q1    Faculty                                                   2
           Measurement Level:  Scale
           Column Width:  8  Alignment:  Right
           Print Format:  F1
           Write Format:  F1
           Missing Values:  *,  8

           Value     Label

               1     Business School
               2     Environ-ment
               3     Humani- ties
               4     Science & Tech.
               5     Social  Studies
               6     CECAC
```

…gives far more detailed information than you may need. You can get much shorter data summaries, **but not from drop-down menus**. For these you need to use the **display** command in syntax direct.
display. …displays the variables currently in the file (to be read in **columns downwards**).

```
Currently Defined Variables

SERIAL    V117      V127      V137      V147      V157      V166      V177
V106      V118      V128      V138      V148      V158      V167      V178
V107      V119      V129      V139      V149      V159      V168      V179
V109      V120      V130      V140      V150      V160      V169      V180
V110      V121      V131      V141      V151      V161      V171      AGESTART
V111      V122      V132      V142      V152      V162      V172      COURSE
V112      V123      V133      V143      V153      V163      V173      FACULTY
V113      V124      V134      V144      V154      V164      V174      YEAR
V115      V125      V135      V145      V155      V165      V175
V116      V126      V136      V146      V156      V170      V176
```

display labels. …is useful for checking presence and accuracy of variable labels.

```
      List of variables on the working file

      Name     Position  Label

      SERIAL           1  Serial number
      V106             2  Q1   Faculty
      V107             3  Q2   Course
      V109             4  Q3a  Full time or part time
      V110             5  Q3b  Daytime or evening
      V111             6  Q3c  Sandwich course
      V112             7  Q3d  Year of course
      V113             8  Q4   Age started course
      V115             9  Q5   Sex
      ~~~~           ~~~~~~~~~~~~~~~~~~~~~~~
      V118            12  Q7a  Lectures
      V119            13  Q7b  Seminars
      V120            14  Q7c  Academic tutorials
```

display variables. …is mostly useful for checking presence and accuracy of missing values.

```
           List of variables on the working file

Name       Pos  Level    Print Fmt     Write Fmt     Missing Values

SERIAL       1  Scale    F4            F4
V106         2  Scale    F1            F1            *, 8
V107         3  Scale    F2            F2            -1
V109         4  Scale    F1            F1            *, 8
V110         5  Scale    F1            F1            *, 8
V111         6  Scale    F1            F1            *, 8
V112         7  Scale    F1            F1            *, 8
V113         8  Scale    F2            F2            -1
V115         9  Scale    F1            F1            *, 8
~~~~       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
V118        12  Scale    F1            F1            *, 8
V119        13  Scale    F1            F1            *, 8
V120        14  Scale    F1            F1            *, 8
```

Printed copies of such summaries can be annotated and used later for amendments and corrections.

All of the above commands (**none of which are available in drop-down menus**) can be abbreviated.

```
disp.

disp lab.

disp var.
```

Another useful facility (also **not available in drop-down menus**) which can be used as an alternative to **Data View** in the **Data Editor** is the **LIST** command. Be careful: if used on its own it will list all values for all cases in the file!

```
LIST.
```

PNL Survey Analysis Workshop[38]

```
                    V V V V V V
          V V V V V 1 1 1 1 1 1
SERIAL 4 5 6 7 8 0 1 2 4 6 7 V18 V19 V20 SEX V24 AGE METRES FEET INCHES    HEIGHT

    1    3 5 2 1 4 1 1 2 5 1 2  5    .   .   1   3  32     .     5    10      1.78
    2    1 4 5 3 2 2 3 3 3 1 2  3    4   .   2   1  44     .     5     7      1.70
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   69    5 1 4 2 3 2 3 2 4 1 2  5    .   .   2   1  40   1.68    .     .      1.68
   70    1 2 5 4 3 2 3 3 2 1 2  3    .   .   2   4  21   1.58    .     .      1.58
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   97    1 5 4 2 3 1 2 3 5 1 2  3    4   5   2   4  24     .     5     1      1.55
   98    1 4 3 2 5 2 2 3 5 1 2  .    .   .   2   4  40     .     5    10      1.78
   99    4 3 2 1 5 1 3 2 5 2 .  .    .   .   1   4  28     .     6     6      1.98

Number of cases read:  169    Number of cases listed:  169
```

*[**NB:** Even after 1990, few students knew their height in metres. **HEIGHT** was calculated later]*

You can specify both the variables and the number of cases to be listed.

```
list     serial sex age height
         /cases 5.
```

```
SERIAL SEX AGE   HEIGHT

   1     1   32      1.78
   2     2   44      1.70
   3     1   32      1.73
   4     2   39      1.68
   5     2   34      1.60

Number of cases read:     5
Number of cases listed:   5
```

LIST is especially useful for checking values before and after recoding[39].

[38] (Cumulative data set from fun questionnaire completed by students at the start of my course)

4.3: Data Analysis

Detailed comparisons of syntax with drop-down menus are given in section 5 later.

	Paste from SPSS	*Syntax direct*
Frequencies	**FREQUENCIES VARIABLES=agegroup sex /ORDER= ANALYSIS .**	**freq var agegroup sex.**
Crosstabs	**CROSSTABS /TABLES=sex BY agegroup /FORMAT= AVALUE TABLES /CELLS= COUNT .**	**cro sex by agegroup.**
Correlation	**CORRELATIONS /VARIABLES=v2018 v2019 v2020 v2021 v2022 v2023 /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE .**	**corr v2018 to v2023 /pri nos.**

4.4 Data transformation

RECODE

For the 1986 British Social Attitudes survey, drop-down menus took several minutes to create a new variable **agegroup** by grouping the values of **v1512** (age in years) and produced this:

Paste from SPSS

```
RECODE
  v1512
  (18 thru 29=1)  (30 thru 44=2)  (45 thru 59=3)  (60 thru 97=4)
  (ELSE=SYSMIS)  INTO  agegroup .
VARIABLE LABELS agegroup 'Age group of respondent'.
EXECUTE .
```

….but how do you put the labels in? ….. by typing them into the **Data Editor** which takes even longer!! It’s far quicker to write the whole thing yourself in a syntax file or even in a Word file and then copy the text into a syntax file.

COMPUTE

Create a variable **antiprot** from the sum of **v2018** to **v2023** and subtract 6 from the total to give a true zero point. Won’t even do it! ..or perhaps it will if I enter the variables separately?

```
COMPUTE antiprot = v2018 + v2019 + v2020 + v2021 + v2022 + v2023 - 6 .
EXECUTE .
```

This takes up valuable time and it’s much quicker to write it yourself as:

```
comp antiprot = sum.6 (v2018 to v2023) - 6 .
```

[39] See the worked example from the European Social Survey on pp 33-34

4.5: First checks on raw data (and on data transformations)

It's sometimes a good idea to read in raw data in alpha format or even as column-binary, particularly from surveys processed on Hollerith cards. It used to be standard practice in fieldwork agencies to code more than one variable in the same column (eg sex, marital status, household status) and also, where multiple response was allowed, to punch more than one response code in the same column. A frequency count will then tell you what's actually in the data. Frequencies on serial and record number can be revealing. Sorting by record number and then running correlation on serial numbers can throw up errors caused by duplicate or missing data lines. Listing serial numbers finds duplicate cases: this may seem wasteful, but at least these days there's no paper involved.

I once used such a check on a data set provided by a research agency to a client who had contracted (and paid) for double sampling of young people. The check revealed around 200 duplicate serial numbers, and a subsequent check revealed that the agency had duplicated the cases rather than conducting additional interviews. The client was less than pleased at being deceived and the MD of the agency was furious at being thus exposed.

4.6: Using functions to generate groups

Household composition types (eg single, couple with or without responsibility for children) or groups taking account of different (for now) State Retirement Pension ages for men and women. For instance, I once did some secondary analysis[40] for a client who wanted to generate a new **family status** variable with four categories, 1 single – no child responsibility; 2 couple – no child responsibility; 3 single – responsible for child(ren); 4 couple - responsible for child(ren) from the respondent's **marital status** and whether he/she had any **responsibility for children**.

First we needed to check the data by frequenciy counts for each variable:

V670 RESPONSIBLE FOR CHILDREN

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	432	30.6	30.6	30.6
	2 No	953	67.5	67.5	98.2
	3	26	1.8	1.8	100.0
	Total	1411	100.0	100.0	

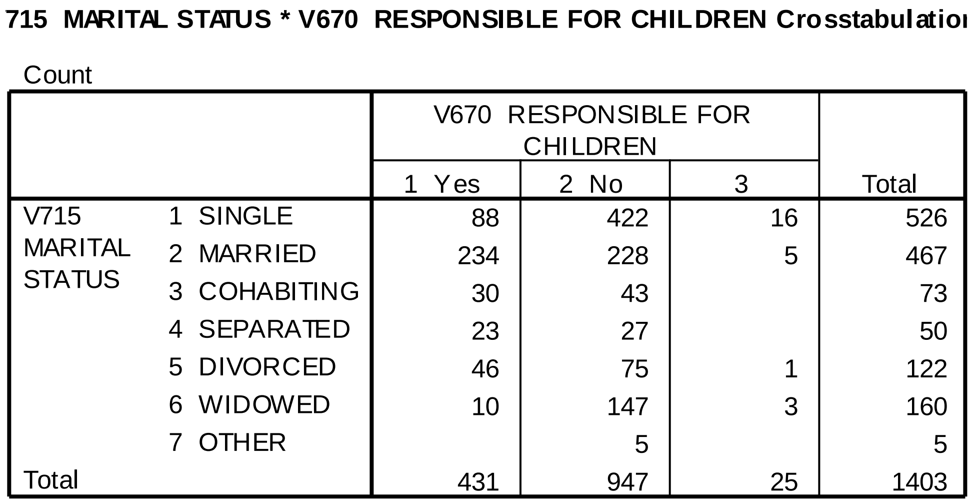
*[**NB**: **v670** code 3 has no label and is probably a DK or Inapp response]*

V715 MARITAL STATUS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 SINGLE	526	37.3	37.5	37.5
	2 MARRIED	467	33.1	33.3	70.8
	3 COHABITING	73	5.2	5.2	76.0
	4 SEPARATED	50	3.5	3.6	79.5
	5 DIVORCED	122	8.6	8.7	88.2
	6 WIDOWED	160	11.3	11.4	99.6
	7 OTHER	5	.4	.4	100.0
	Total	1403	99.4	100.0	
Missing	System	8	.6		
Total		1411	100.0		

[40] for Islington Welfare Rights Unit on the Bloomsbury and Islington **Health and Lifestyle Survey** See: Griffiths S**, Unfair Shares: an investigation of the effects the Social Security Act in Islington** Welfare Rights Unit, Islington 1989

... and also a crosstabulation:



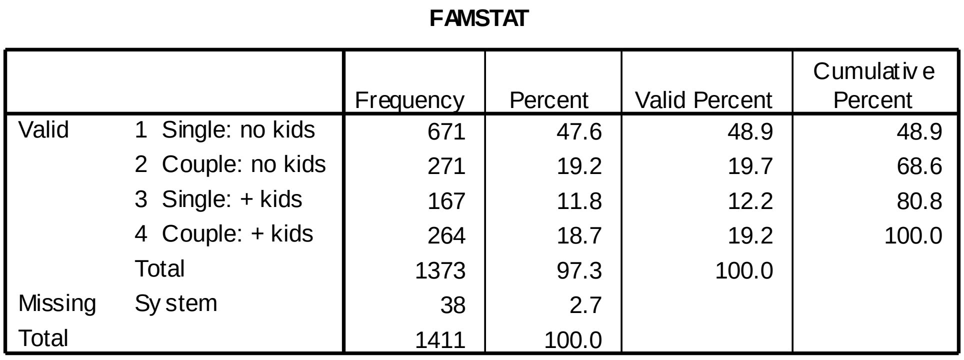
715 MARITAL STATUS * V670 RESPONSIBLE FOR CHILDREN Crosstabulatioı

Count

		V670 RESPONSIBLE FOR CHILDREN			
		1 Yes	2 No	3	Total
V715 MARITAL STATUS	1 SINGLE	88	422	16	526
	2 MARRIED	234	228	5	467
	3 COHABITING	30	43		73
	4 SEPARATED	23	27		50
	5 DIVORCED	46	75	1	122
	6 WIDOWED	10	147	3	160
	7 OTHER		5		5
Total		431	947	25	1403

There are several ways of generating the new variable (the longest using a succession of **IF** commands or **DO IF ~~~ ELSE IF**), but a quick way is:

```
compute     famstat=v670*10+v715.
recode      famstat (21 24 25 26=1)(22 23=2)
                   (11 14 15 16=3)(12 13=4)(else=sysmis).
value labels famstat
                    1 'Single: no kids' 2 'Couple: no kids'
                    3 'Single: + kids' 4 'Couple: + kids'.
freq var    famstat.
```



FAMSTAT

		Frequency	Percent	Valid Percent	Cumulativ e Percent
Valid	1 Single: no kids	671	47.6	48.9	48.9
	2 Couple: no kids	271	19.2	19.7	68.6
	3 Single: + kids	167	11.8	12.2	80.8
	4 Couple: + kids	264	18.7	19.2	100.0
	Total	1373	97.3	100.0	
Missing	Sy stem	38	2.7		
Total		1411	100.0		

*[**NB:** Row labels limited to 20 characters in those days]*

compute can also be used in other ways. For instance a quick way of looking at grouped distributions is to use the **truncation** function e.g.

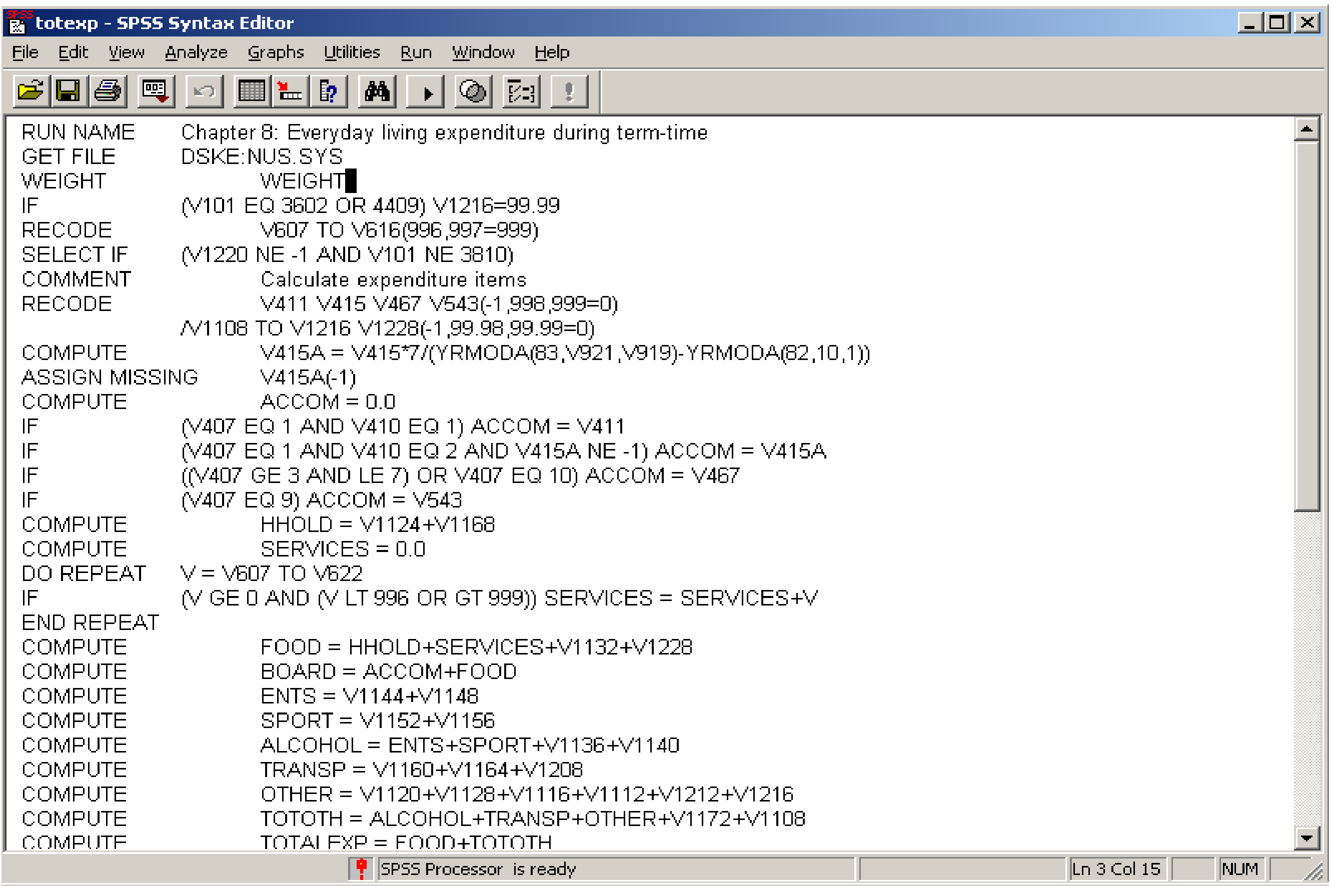
compute agegrp10 = trunc (age/10)

. . . which divides age by 10 and knocks off the decimal part to leave an integer. The same principle can be applied to year of birth (e.g. decade of birth for people born 1900-1999: subtract 1900 and divide by 10) or income in ££ (groups dividing by 100,200, 500 etc). This can get complicated, but it can also save time.

4.7: Complex calculations taking account of missing data

Missing data can be full of pitfalls for the unwary. For instance, in the 1982 Undergraduate Income and Expenditure Survey for the National Union of Students [41], detailed diaries were kept of expenditure under different headings. Values which were declared as missing for one purpose sometimes needed to be recoded to zero for others. Some of the SPSS setup files for such analysis ran to well over 100 lines of **IF**, **COUNT** and **COMPUTE** commands.

NUS Undergraduate Income and Expenditure Survey 1982


totexp - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
SPSS Processor is ready
Ln 3 Col 15
NUM

```
RUN NAME            Chapter 8: Everyday living expenditure during term-time
GET FILE            DSKE:NUS.SYS
WEIGHT                        WEIGHT
IF                  (V101 EQ 3602 OR 4409) V1216=99.99
RECODE                        V607 TO V616(996,997=999)
SELECT IF           (V1220 NE -1 AND V101 NE 3810)
COMMENT                       Calculate expenditure items
RECODE                        V411 V415 V467 V543(-1,998,999=0)
                    /V1108 TO V1216 V1228(-1,99.98,99.99=0)
COMPUTE                       V415A = V415*7/(YRMODA(83,V921,V919)-YRMODA(82,10,1))
ASSIGN MISSING                V415A(-1)
COMPUTE                       ACCOM = 0.0
IF                  (V407 EQ 1 AND V410 EQ 1) ACCOM = V411
IF                  (V407 EQ 1 AND V410 EQ 2 AND V415A NE -1) ACCOM = V415A
IF                  ((V407 GE 3 AND LE 7) OR V407 EQ 10) ACCOM = V467
IF                  (V407 EQ 9) ACCOM = V543
COMPUTE                       HHOLD = V1124+V1168
COMPUTE                       SERVICES = 0.0
DO REPEAT           V = V607 TO V622
IF                  (V GE 0 AND (V LT 996 OR GT 999)) SERVICES = SERVICES+V
END REPEAT
COMPUTE                       FOOD = HHOLD+SERVICES+V1132+V1228
COMPUTE                       BOARD = ACCOM+FOOD
COMPUTE                       ENTS = V1144+V1148
COMPUTE                       SPORT = V1152+V1156
COMPUTE                       ALCOHOL = ENTS+SPORT+V1136+V1140
COMPUTE                       TRANSP = V1160+V1164+V1208
COMPUTE                       OTHER = V1120+V1128+V1116+V1112+V1212+V1216
COMPUTE                       TOTOTH = ALCOHOL+TRANSP+OTHER+V1172+V1108
COMPUTE                       TOTAL EXP = FOOD+TOTOTH
```

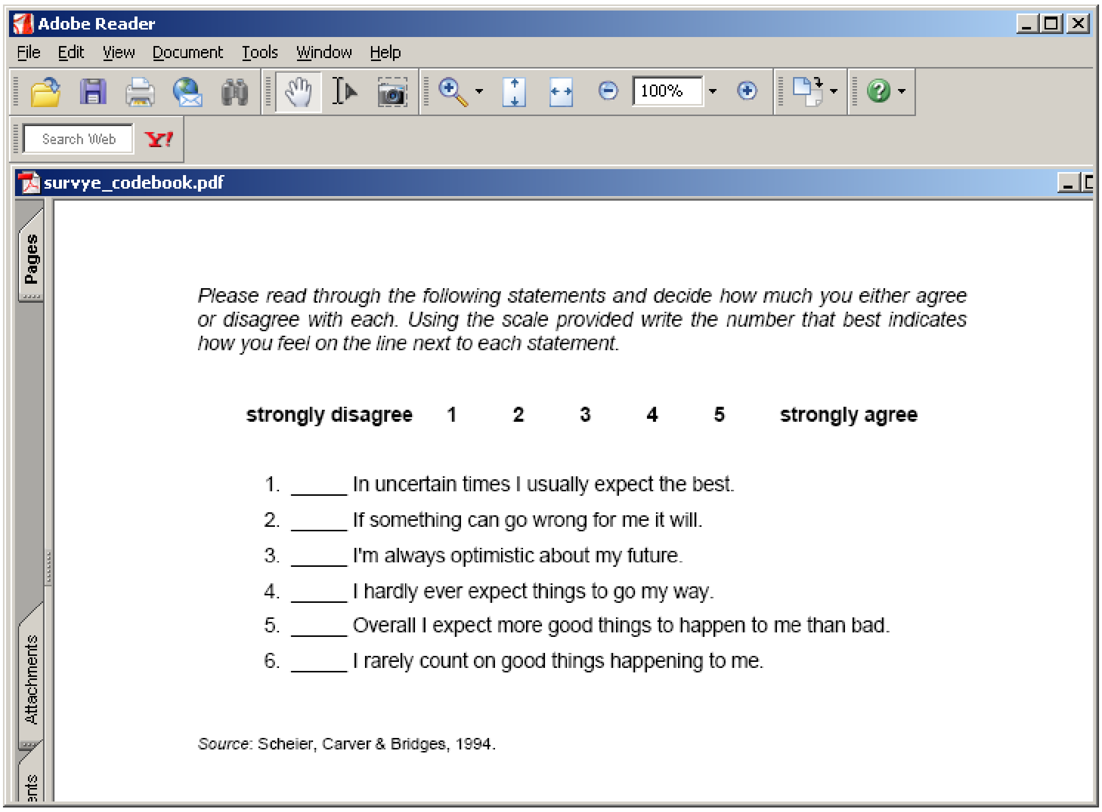
…and this is one of the less complex setups! If you want to have a shot at doing this with drop-down menus, feel free.

[41] When the DES (aka Thatcher government) refused to fund this regular survey, the NUS (who had a modest budget set aside) turned to the PNL Survey Research Unit for help. A quick telephone call to Nick Moon of NOP produced a fieldwork quote, and a couple of pints with NUS reps over lunch in a pub in Islington produced a literal back of a serviette costing of £12,000 for the whole exercise: hands were shaken. The survey was completed well within both time and budget, and at a fraction of what the DES would have expected to pay. See: Richard Goring, Jim Ring and John Hall, **Undergraduate Income and Expenditure,** PNL Survey Research Unit, 1983: J. Saxby, **Undergraduate Income and Expenditure,** National Union of Students, 1984: National Union of Students, **Undergraduate Income and Spending**, NUS booklet, 1984.

5: Exercises from Pallant 2005 (and one from 2001)

(for **RECODE**, **COMPUTE** and **CORRELATIONS** using questionnaire[42] and data set as supplied.)

Facsimile of Optimism-Pessimism scale



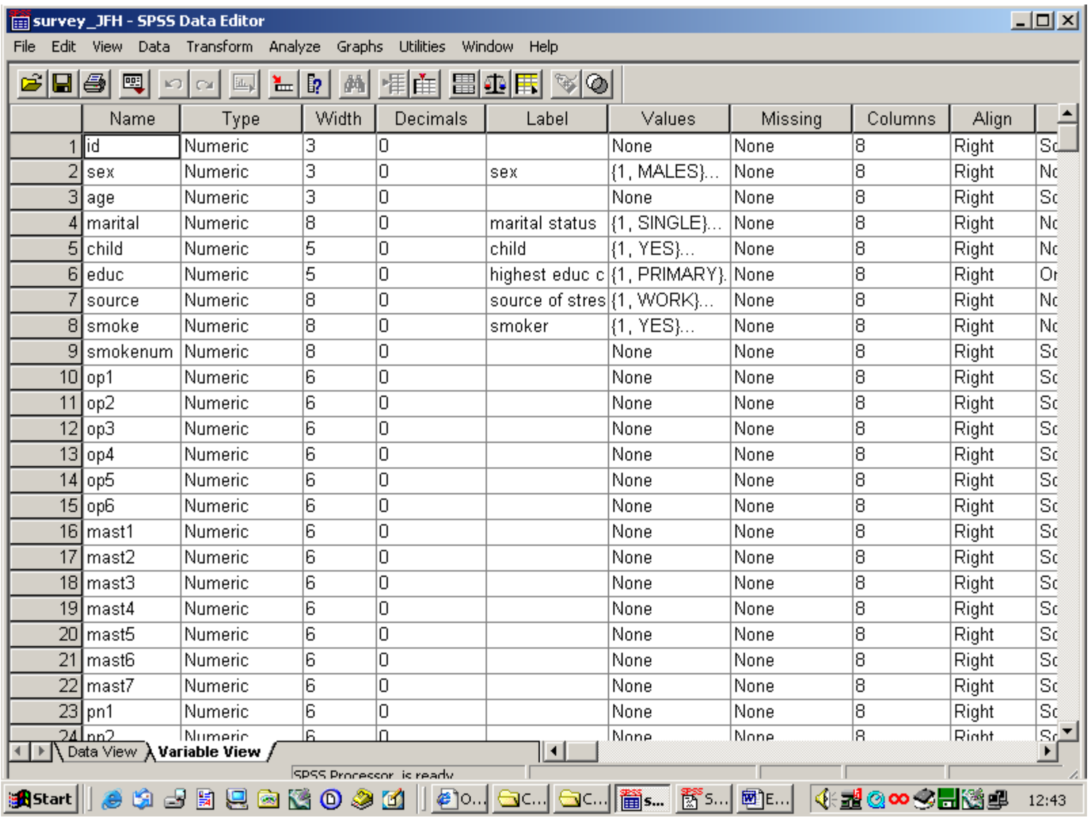
Please read through the following statements and decide how much you either agree or disagree with each. Using the scale provided write the number that best indicates how you feel on the line next to each statement.

strongly disagree 1 2 3 4 5 strongly agree

1. _____ In uncertain times I usually expect the best.
2. _____ If something can go wrong for me it will.
3. _____ I'm always optimistic about my future.
4. _____ I hardly ever expect things to go my way.
5. _____ Overall I expect more good things to happen to me than bad.
6. _____ I rarely count on good things happening to me.

Source: Scheier, Carver & Bridges, 1994.

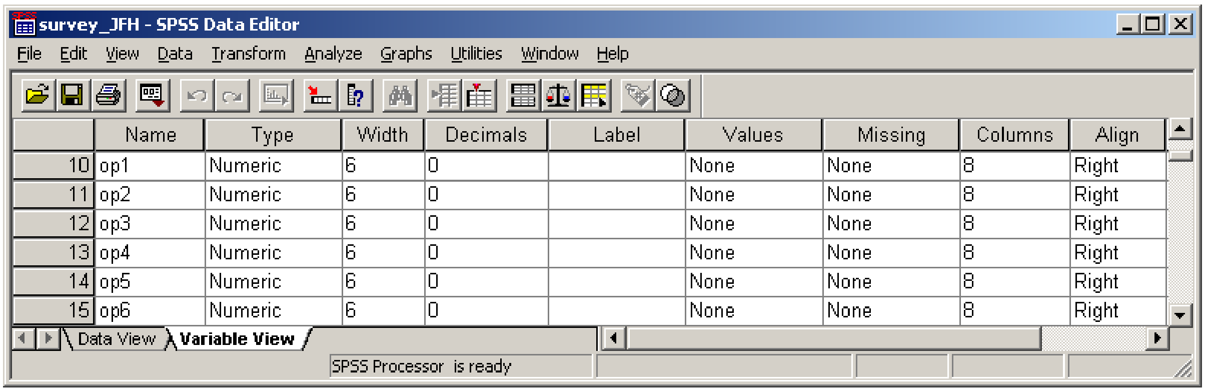
Open the file **survey.sav** as supplied by double-clicking the icon:


survey_JFH - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready
Start
12:43

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	
1	id	Numeric	3	0		None	None	8	Right	Sc
2	sex	Numeric	3	0	sex	{1, MALES}...	None	8	Right	Nc
3	age	Numeric	3	0		None	None	8	Right	Sc
4	marital	Numeric	8	0	marital status	{1, SINGLE}...	None	8	Right	Nc
5	child	Numeric	5	0	child	{1, YES}...	None	8	Right	Nc
6	educ	Numeric	5	0	highest educ c	{1, PRIMARY}.	None	8	Right	Or
7	source	Numeric	8	0	source of stres	{1, WORK}...	None	8	Right	Nc
8	smoke	Numeric	8	0	smoker	{1, YES}...	None	8	Right	Nc
9	smokenum	Numeric	8	0		None	None	8	Right	Sc
10	op1	Numeric	6	0		None	None	8	Right	Sc
11	op2	Numeric	6	0		None	None	8	Right	Sc
12	op3	Numeric	6	0		None	None	8	Right	Sc
13	op4	Numeric	6	0		None	None	8	Right	Sc
14	op5	Numeric	6	0		None	None	8	Right	Sc
15	op6	Numeric	6	0		None	None	8	Right	Sc
16	mast1	Numeric	6	0		None	None	8	Right	Sc
17	mast2	Numeric	6	0		None	None	8	Right	Sc
18	mast3	Numeric	6	0		None	None	8	Right	Sc
19	mast4	Numeric	6	0		None	None	8	Right	Sc
20	mast5	Numeric	6	0		None	None	8	Right	Sc
21	mast6	Numeric	6	0		None	None	8	Right	Sc
22	mast7	Numeric	6	0		None	None	8	Right	Sc
23	pn1	Numeric	6	0		None	None	8	Right	Sc
24	pn2	Numeric	6	0		None	None	8	Right	Sc

[42] Supplied as **survye_codebook** (sic!!) and is only partial, so I can't evaluate the items. Not sure the questionnaire format couldn't do with improvement either.

Adjust the window to contain only variables of interest:


survey_JFH - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
10	op1	Numeric	6	0		None	None	8	Right
11	op2	Numeric	6	0		None	None	8	Right
12	op3	Numeric	6	0		None	None	8	Right
13	op4	Numeric	6	0		None	None	8	Right
14	op5	Numeric	6	0		None	None	8	Right
15	op6	Numeric	6	0		None	None	8	Right

Note the **lack of labels**: this may be OK for a lone researcher working on her own material for a PhD, but pity anyone else trying to use the data set. I know psychologists dislike single item measures and like to go straight to scale scores derived from items surviving from someone else's factor analysis (usually with researcher administered pencil and and paper schedules on captive subjects in an office or classroom rather than interviewer administered or self-completed questionnaires to respondents from general populations in their own homes) but my advice is always to **check your data first** with an initial frequency count. . .

OP1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	21	4.8	4.8	4.8
	2	67	15.3	15.4	20.2
	3	169	38.5	38.9	59.1
	4	132	30.1	30.3	89.4
	5	46	10.5	10.6	100.0
	Total	435	99.1	100.0	
Missing	System	4	.9		
Total		439	100.0		

OP2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	15	3.4	3.4	3.4
	2	49	11.2	11.2	14.7
	3	111	25.3	25.5	40.1
	4	131	29.8	30.0	70.2
	5	130	29.6	29.8	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

OP3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	15	3.4	3.4	3.4
	2	28	6.4	6.4	9.9
	3	118	26.9	27.1	36.9
	4	187	42.6	42.9	79.8
	5	88	20.0	20.2	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

OP4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	14	3.2	3.2	3.2
	2	45	10.3	10.3	13.5
	3	101	23.0	23.2	36.7
	4	156	35.5	35.8	72.5
	5	120	27.3	27.5	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

OP5

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	10	2.3	2.3	2.3
	2	29	6.6	6.7	8.9
	3	78	17.8	17.9	26.8
	4	179	40.8	41.1	67.9
	5	140	31.9	32.1	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

OP6

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	18	4.1	4.1	4.1
	2	56	12.8	12.8	17.0
	3	91	20.7	20.9	37.8
	4	130	29.6	29.8	67.7
	5	141	32.1	32.3	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

*[**NB:** On 18 Oct 2006 I found another way of doing this via drop-down menus (which I wouldn't otherwise have found) but I couldn't find a way of displaying the variable names for each column, so they were inserted manually.]*

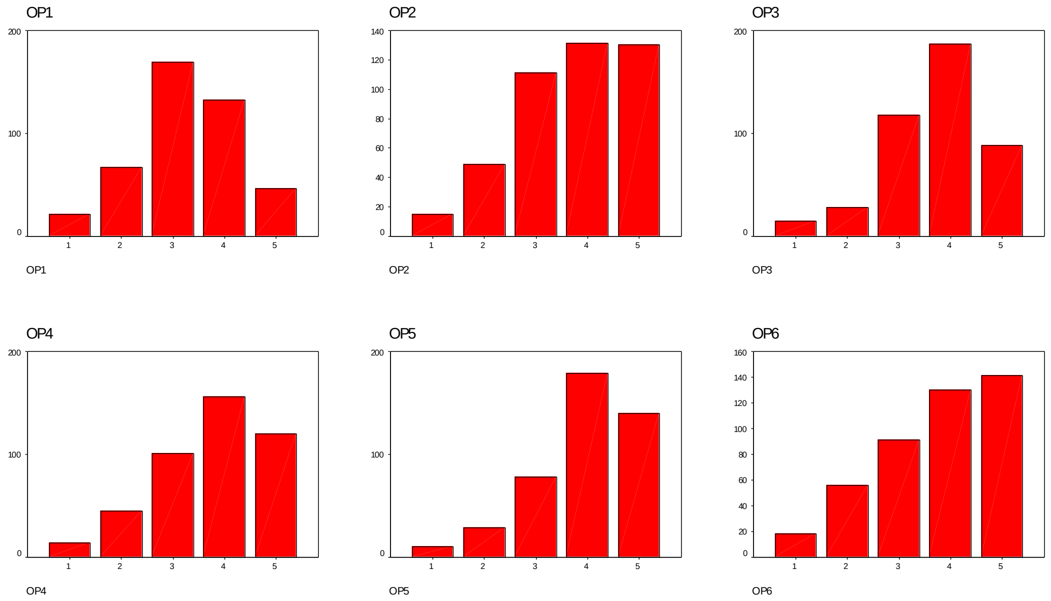
Analyse…

Custom Tables…

Tables of Frequencies

	op1	op2	op3	op4	op5	op6
	Count	Count	Count	Count	Count	Count
1	21	15	15	14	10	18
2	67	49	28	45	29	56
3	169	111	118	101	78	91
4	132	131	187	156	179	130
5	46	130	88	120	140	141

For a visual check you can use a graphic such as a barchart[43].



The result of this doesn't make sense! *The frequency distributions for the negative items have the same shape (negatively skewed) as the positive items, when they would be expected to be skewed in opposite directions.*

A frequent professional trick with batteries of attitude and similar items is to run off a correlation matrix to check strength and direction of association (and to see if the coding matches the questionnaire).

Correlations

		OP1	OP2	OP3	OP4	OP5	OP6
OP1	Pearson Correlation	1	.285	.382	.381	.337	.261
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
	N	435	435	435	435	435	435
OP2	Pearson Correlation	.285	1	.345	.584	.402	.435
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000
	N	435	436	436	436	436	436
OP3	Pearson Correlation	.382	.345	1	.419	.419	.250
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000
	N	435	436	436	436	436	436
OP4	Pearson Correlation	.381	.584	.419	1	.424	.510
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000
	N	435	436	436	436	436	436
OP5	Pearson Correlation	.337	.402	.419	.424	1	.502
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000
	N	435	436	436	436	436	436
OP6	Pearson Correlation	.261	.435	.250	.510	.502	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.
	N	435	436	436	436	436	436

What's going on? The problem is that Pallant does not give specimen results to use as a check. Clearly the variables have already been recoded in the file as distributed, so I've recoded them back again and this time the initial frequencies make intuitive sense.

[43] In syntax: **freq var op1 to op6 /for not /hbar.**

OP2

		Frequency	Percent	Valid Percent	Cumulativ e Percent
Valid	1	130	29.6	29.8	29.8
	2	131	29.8	30.0	59.9
	3	111	25.3	25.5	85.3
	4	49	11.2	11.2	96.6
	5	15	3.4	3.4	100.0
	Total	436	99.3	100.0	
Missing	Sy stem	3	.7		
Total		439	100.0		

OP4

		Frequency	Percent	Valid Percent	Cumulativ e Percent
Valid	1	120	27.3	27.5	27.5
	2	156	35.5	35.8	63.3
	3	101	23.0	23.2	86.5
	4	45	10.3	10.3	96.8
	5	14	3.2	3.2	100.0
	Total	436	99.3	100.0	
Missing	Sy stem	3	.7		
Total		439	100.0		

OP6

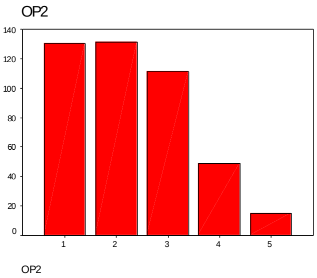
		Frequency	Percent	Valid Percent	Cumulativ e Percent
Valid	1	141	32.1	32.3	32.3
	2	130	29.6	29.8	62.2
	3	91	20.7	20.9	83.0
	4	56	12.8	12.8	95.9
	5	18	4.1	4.1	100.0
	Total	436	99.3	100.0	
Missing	Sy stem	3	.7		
Total		439	100.0		

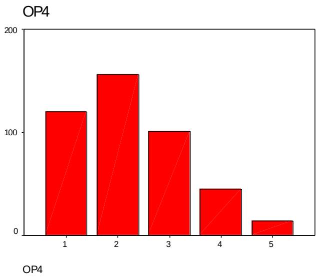
…and in condensed form:

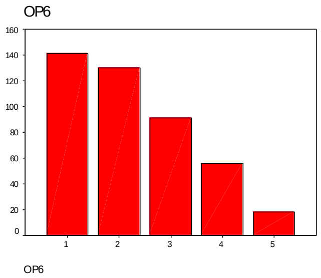
Analyse
Custom Tables
Frequencies

	op1	op2	op3	op4	op5	op6
	Count	Count	Count	Count	Count	Count
1	21	130	15	120	10	141
2	67	131	28	156	29	130
3	169	111	118	101	78	91
4	132	49	187	45	179	56
5	46	15	88	14	140	18

Graphically:







and, as an additional check, the correlation matrix has negatives in the right places, as one might expect.

Correlations

		OP1	OP2	OP3	OP4	OP5	OP6
OP1	Pearson Correlation	1	-.285	.382	-.381	.337	-.261
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
	N	435	435	435	435	435	435
OP2	Pearson Correlation	-.285	1	-.345	.584	-.402	.435
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000
	N	435	436	436	436	436	436
OP3	Pearson Correlation	.382	-.345	1	-.419	.419	-.250
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000
	N	435	436	436	436	436	436
OP4	Pearson Correlation	-.381	.584	-.419	1	-.424	.510
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000
	N	435	436	436	436	436	436
OP5	Pearson Correlation	.337	-.402	.419	-.424	1	-.502
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000
	N	435	436	436	436	436	436
OP6	Pearson Correlation	-.261	.435	-.250	.510	-.502	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.
	N	435	436	436	436	436	436

This file is deficient by my standards as it has very few variable or value labels. I'm not sure it wouldn't also benefit from renaming most of the variables in file order (using the row numbers) to make them easier to find, or else putting the names on the questionnaire as well so that it can be used as documentation.

Failing that, it helps to have something like this (reading downwards in columns):

```
Currently Defined Variables[44]

ID        OP6       PN7       LIFSAT1   PSS10     M4        PC8       TNEGAFF
SEX       MAST1     PN8       LIFSAT2   SEST1     M5        PC9       TLIFESAT
AGE       MAST2     PN9       LIFSAT3   SEST2     M6        PC10      TPSTRESS
MARITAL   MAST3     PN10      LIFSAT4   SEST3     M7        PC11      TSLFEST
CHILD     MAST4     PN11      LIFSAT5   SEST4     M8        PC12      TMARLOW
EDUC      MAST5     PN12      PSS1      SEST5     M9        PC13      TPCOISS
SOURCE    MAST6     PN13      PSS2      SEST6     M10       PC14      EDUC2
SMOKE     MAST7     PN14      PSS3      SEST7     PC1       PC15      AGEGP3
SMOKENUM  PN1       PN15      PSS4      SEST8     PC2       PC16      AGEGP5
OP1       PN2       PN16      PSS5      SEST9     PC3       PC17      AGEBAND3
OP2       PN3       PN17      PSS6      SEST10    PC4       PC18      EDUCREC
OP3       PN4       PN18      PSS7      M1        PC5       TOPTIM
OP4       PN5       PN19      PSS8      M2        PC6       TMAST
OP5       PN6       PN20      PSS9      M3        PC7       TPOSAFF
```

So, having restored the original data values to the variables of interest, and with summary documentation to hand, we are now ready to follow Pallant's steps.

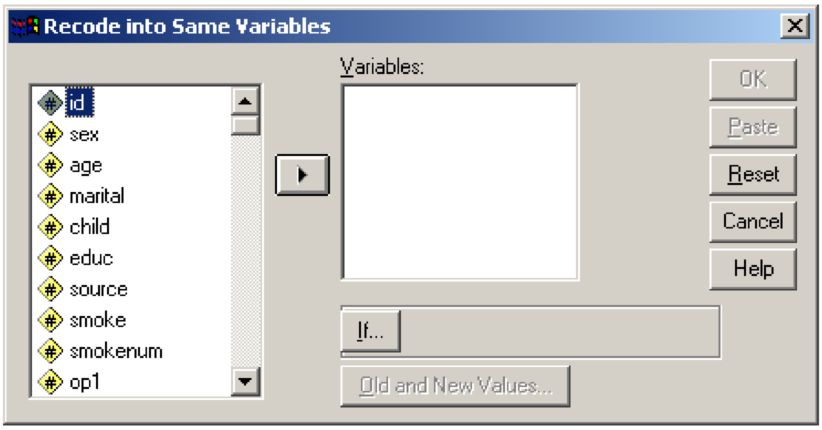
[44] Output from **DISPLAY**. Can't be done using SPSS drop-down menus

Exercise 1: Reversing scores of items 2, 4 and 6 (2005, pp79-80: 7 steps, one with 4 repeats)

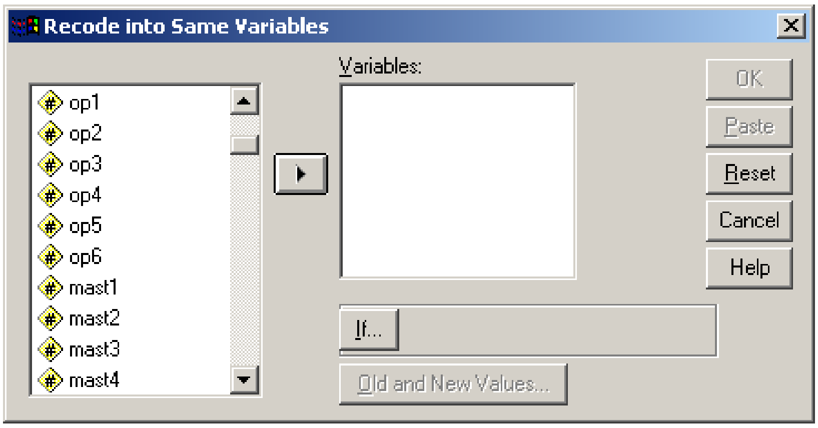
Edit, options, data is redundant as the file is already open in this mode, but I've saved a copy with a different name so as to preserve the original. Here we go!

Transform
Recode
Into Same Variables

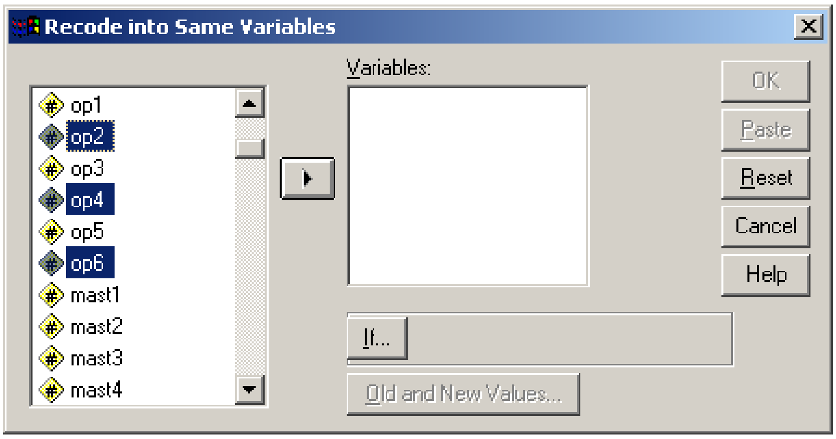
Let's assume you haven't read the previous section and don't know where the variables are in the file: luckily they're near the beginning and the first item in the battery, **op1**, appears at the bottom of the **Recode into Same Variables** box:



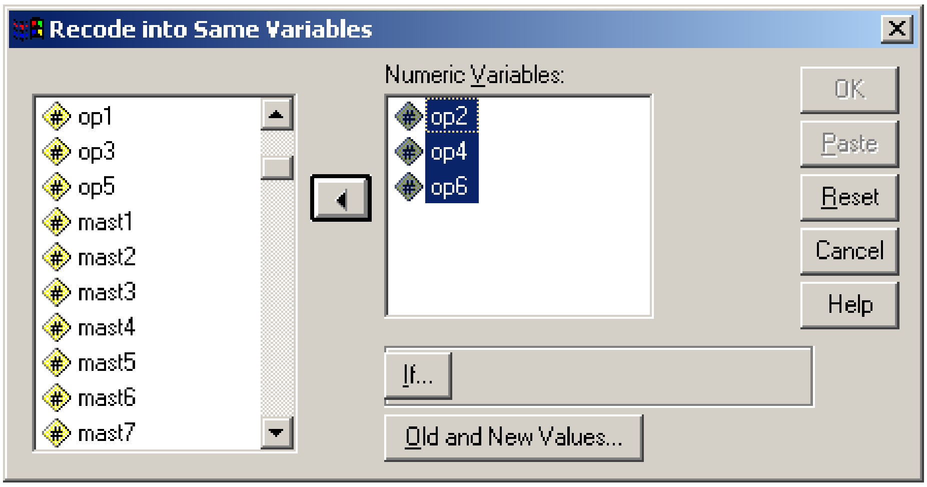
so by scrolling down we can display all six items



Pallant doesn't tell us how to select them either, but we need to highlight variables **op2 op4** and **op6**.

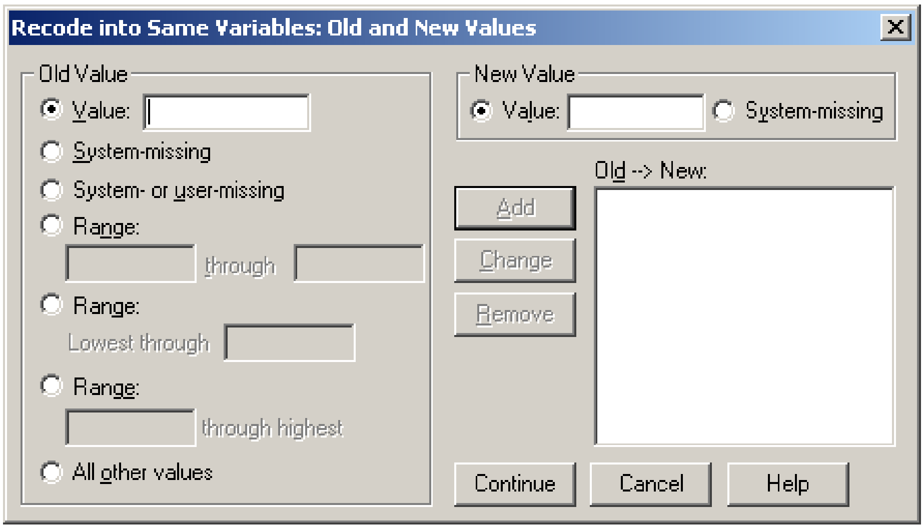


… and either move them one at a time by clicking first the variable and then the **right arrow**, or all at once by clicking on **op2** and using **[CTRL]+click** to highlight **op4** and **op6** (making sure we don't drag **id** with them as well) then clicking the **right arrow** to copy them into the **Numeric Variables** box.

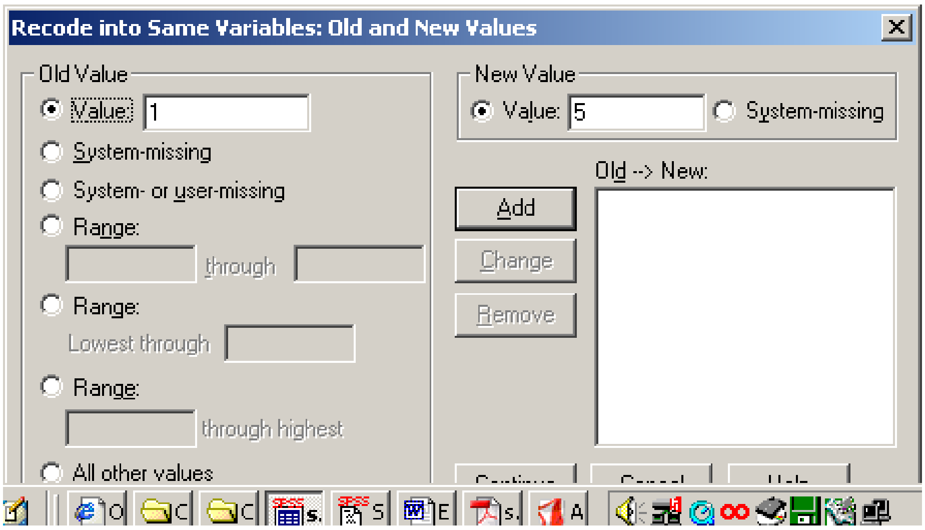


Numeric Variables: op2, op4, op6

Type **[Alt]+O** or click on **Old and New Values**



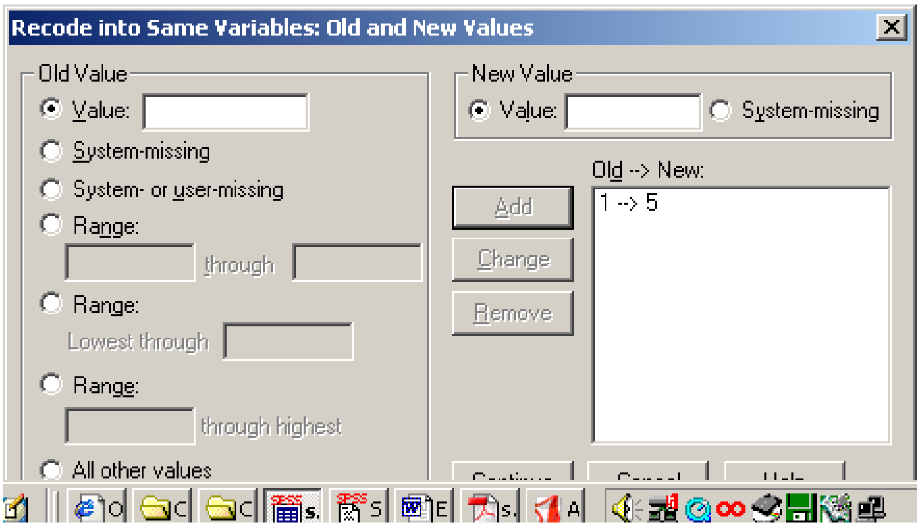
Enter **1** in **Old Value** and **5** in **New Value**



Old Value: ● Value: 1

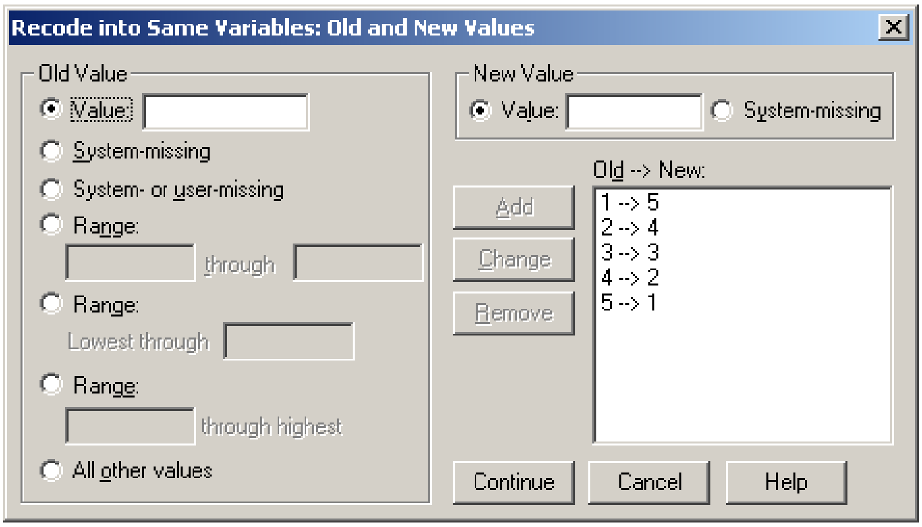
New Value: ● Value: 5

…and click **Add** for



Old --> New:
1 --> 5

Now do the whole thing **another four times (!!!)** to change **2** to **4**, **3** to **3**, **4** to **2**, and **5** to**1** until:



Old --> New:
1 --> 5
2 --> 4
3 --> 3
4 --> 2
5 --> 1

and click on **Continue**

Oops! I clicked PASTE first and had to start again as I lost the window and couldn't go back. Couldn't work out what happened, but discovered later that it pastes direct to the syntax file, so you need to open the latter then run the job with [CTRL]+R. You can also use the dialog recall box, which takes 2 clicks.

Clicking **OK** immediately effects the changes as permanent. It would have been quicker to write:

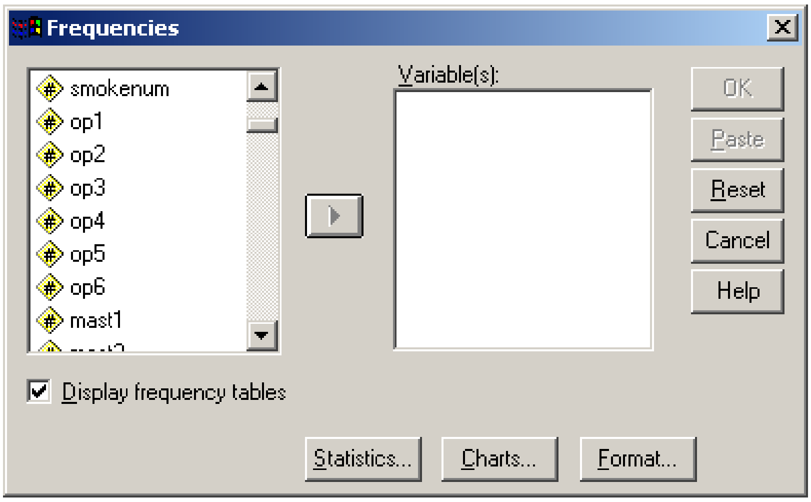
```
RECODE
     op2 op4 op6  (1=5)  (2=4)  (3=3)  (4=2)  (5=1) .
EXECUTE .
```

…as obtained from **PASTE**, but there is no need to recode **(3=3)** and you don't need to **EXECUTE** either (this overwrites your original data: it's safer to do this as a **TEMPORARY** recode anyway.

Now check the frequency counts for the recoded variables (In the book, Pallant rightly recommends this before and after data transformations, but doesn't do it for this example)

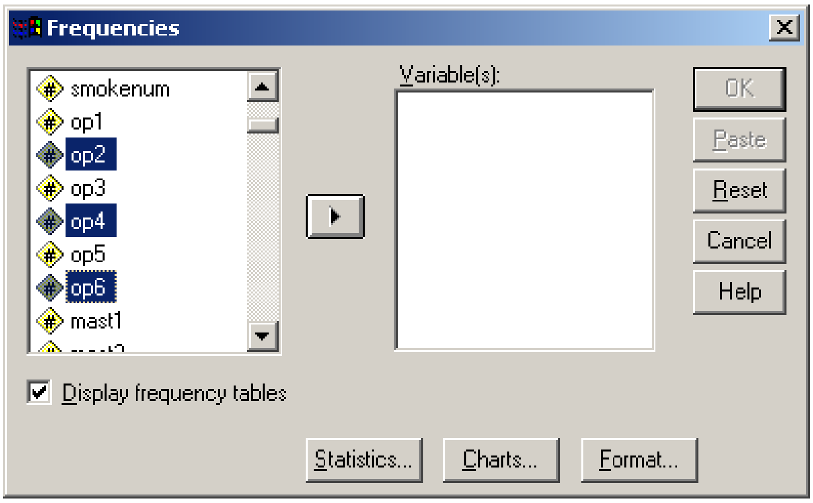
Analyse
Descriptives
Frequencies

… and scroll down find variables **op1** to **op6**:



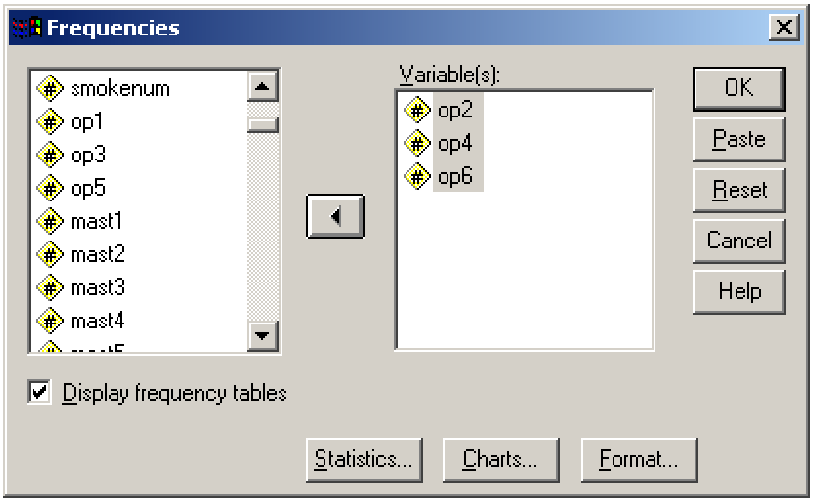
☑ Display frequency tables

Highlight variables **op2 op4** and **op6**:



☑ Display frequency tables

…and click **right arrow** transfer to **Variable(s)** box:



Variable(s):
op2
op4
op6

☑ Display frequency tables

..then click **OK** to run.

The frequencies for the recoded variables are now the right way round for the next step, but you really need to make this clear, possibly by changing the variable labels to something like ***OP6 reversed****. I would have either generated new variables and left the originals as per questionnaire, or more likely used* ***temp****)*

OP2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	15	3.4	3.4	3.4
	2	49	11.2	11.2	14.7
	3	111	25.3	25.5	40.1
	4	131	29.8	30.0	70.2
	5	130	29.6	29.8	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

OP4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	14	3.2	3.2	3.2
	2	45	10.3	10.3	13.5
	3	101	23.0	23.2	36.7
	4	156	35.5	35.8	72.5
	5	120	27.3	27.5	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

OP6

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	18	4.1	4.1	4.1
	2	56	12.8	12.8	17.0
	3	91	20.7	20.9	37.8
	4	130	29.6	29.8	67.7
	5	141	32.1	32.3	100.0
	Total	436	99.3	100.0	
Missing	System	3	.7		
Total		439	100.0		

It's a lot quicker to write:

```
freq var op1 to op6.
```

or

```
freq var op2 op4 op6.
```

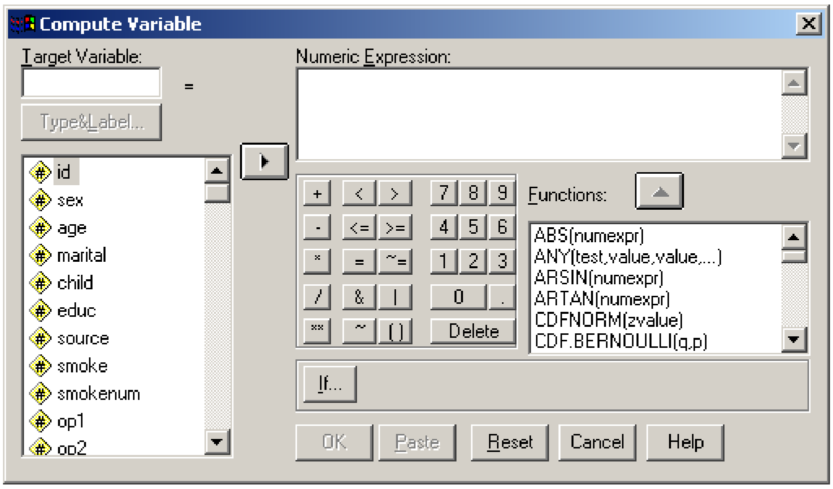
Another way of doing the recode is:

```
do repeat
     X=op2, op4, op6
compute x=6-x.
end repeat.
```

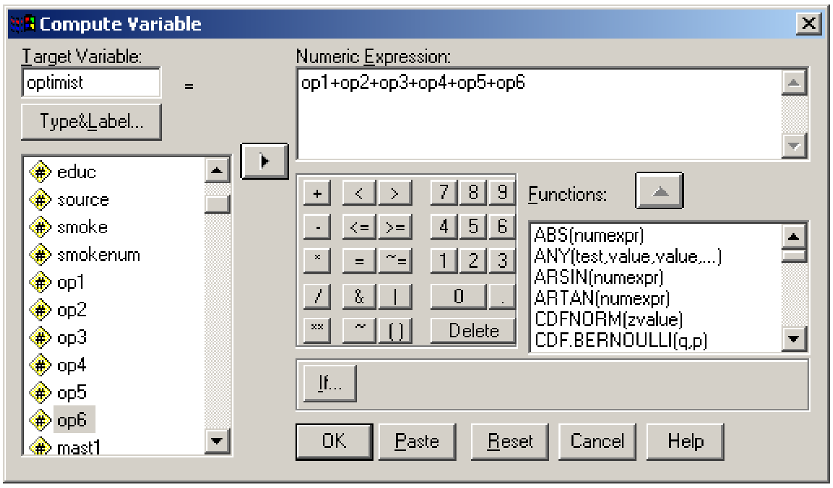
Exercise 2: Adding up items to yield scale score (2005, pp80-81: 9 steps, one with 4 repeats)

Transform

Compute



Drop-down menus then require **no fewer than 9 steps !!!** (one with **4 repeats!!**) to produce::



Target Variable: optimist

Numeric Expression: op1+op2+op3+op4+op5+op6

Now check the frequency count for the newly computed variable*. Pallant doesn't do it for this example either, but by now I've given up entirely on drop-down menus. It takes for ever and you get a* ***painful bunion on your wrist!.***

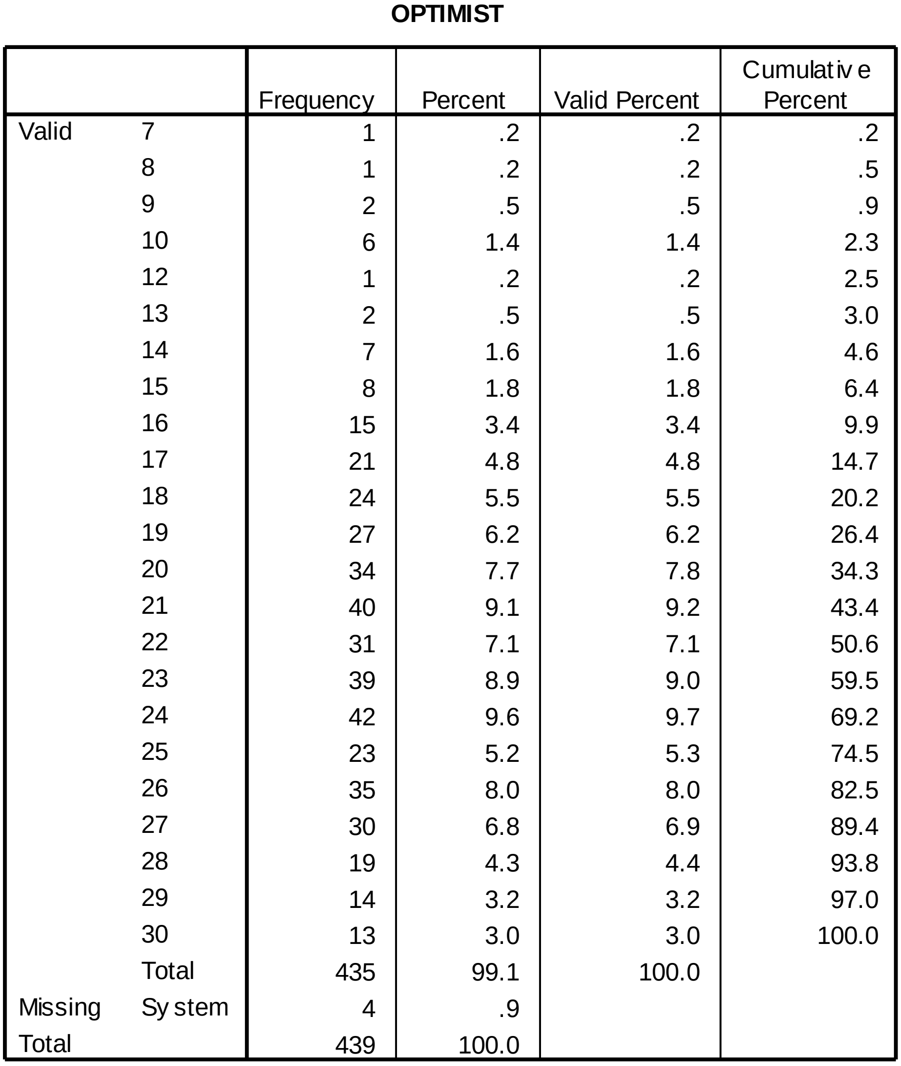
Again, it would be quicker to write:

```
COMPUTE
  optimist = op1+op2+op3+op4+op5+op6.
FREQUENCIES
  VARIABLES=optimist
  /ORDER=  ANALYSIS .
```

as obtained from **PASTE**, but all the above can be done (**in seconds**) directly in syntax:

```
recode  op2 op4 op6   (1=5)(2=4)(4=2)(5=1).
comp    optimist = op1+op2+op3+op4+op5+op6.
freq    var optimist.
```

…which is not much shorter than the original **PASTE**, but would be if the score was calculated over a longer list: it is however a lot quicker to use syntax than drop-down menus.



OPTIMIST

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	7	1	.2	.2	.2
	8	1	.2	.2	.5
	9	2	.5	.5	.9
	10	6	1.4	1.4	2.3
	12	1	.2	.2	2.5
	13	2	.5	.5	3.0
	14	7	1.6	1.6	4.6
	15	8	1.8	1.8	6.4
	16	15	3.4	3.4	9.9
	17	21	4.8	4.8	14.7
	18	24	5.5	5.5	20.2
	19	27	6.2	6.2	26.4
	20	34	7.7	7.8	34.3
	21	40	9.1	9.2	43.4
	22	31	7.1	7.1	50.6
	23	39	8.9	9.0	59.5
	24	42	9.6	9.7	69.2
	25	23	5.2	5.3	74.5
	26	35	8.0	8.0	82.5
	27	30	6.8	6.9	89.4
	28	19	4.3	4.4	93.8
	29	14	3.2	3.2	97.0
	30	13	3.0	3.0	100.0
	Total	435	99.1	100.0	
Missing	System	4	.9		
Total		439	100.0		

[ie a range from 7 to 30 out of a theoretical range of 6 to 30]

There is another way of calculating the total score using the **SUM** function:

```
comp optimist = sum (op1 to op6) .
```

…but to ensure a score is calculated only if all six items have been answered we need to write:

```
comp optimist = sum.6 (op1 to op6) .
```

The resulting score has a range of 6 to 30, but we really ought to subtract 6 from the total to give the score a true zero point, and make it a ratio scale (so that we can say 20 is twice as optimistic as 10?).

```
comp optimist = sum.6 (op1 to op6) - 6 .
```

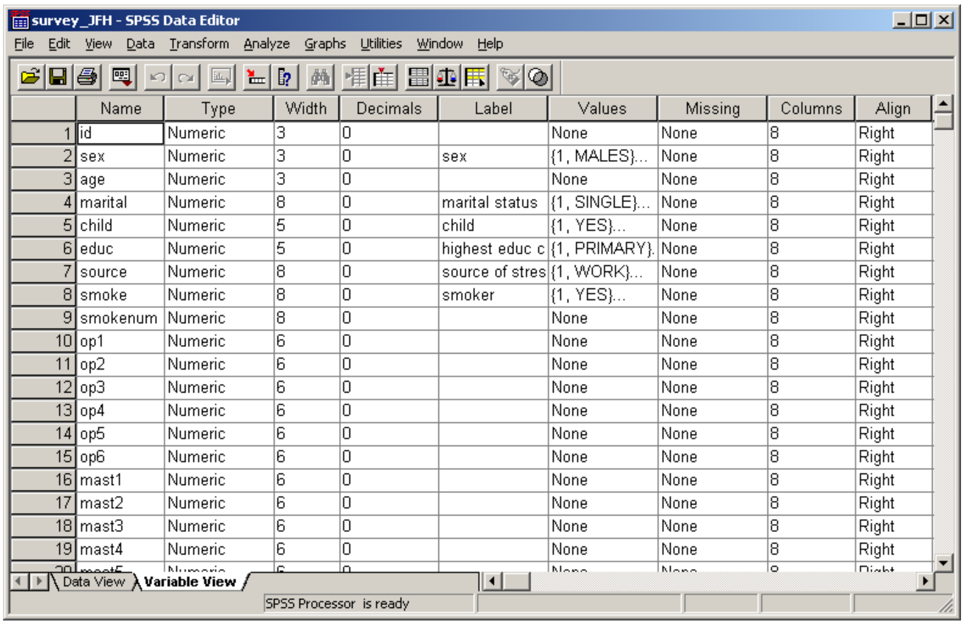
Pallant doesn't do this: presumably neither do her sources.

Exercise 3: **Grouping the values of a continuous variable** (Pallant 2005, pp85-87).

Task: **Create 3 approximately equal-sized groups for AGE**

Step 1: **Find the cutting points**[45] (Pallant 2001, pp 81-84)

Open Data Editor


survey_JFH - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

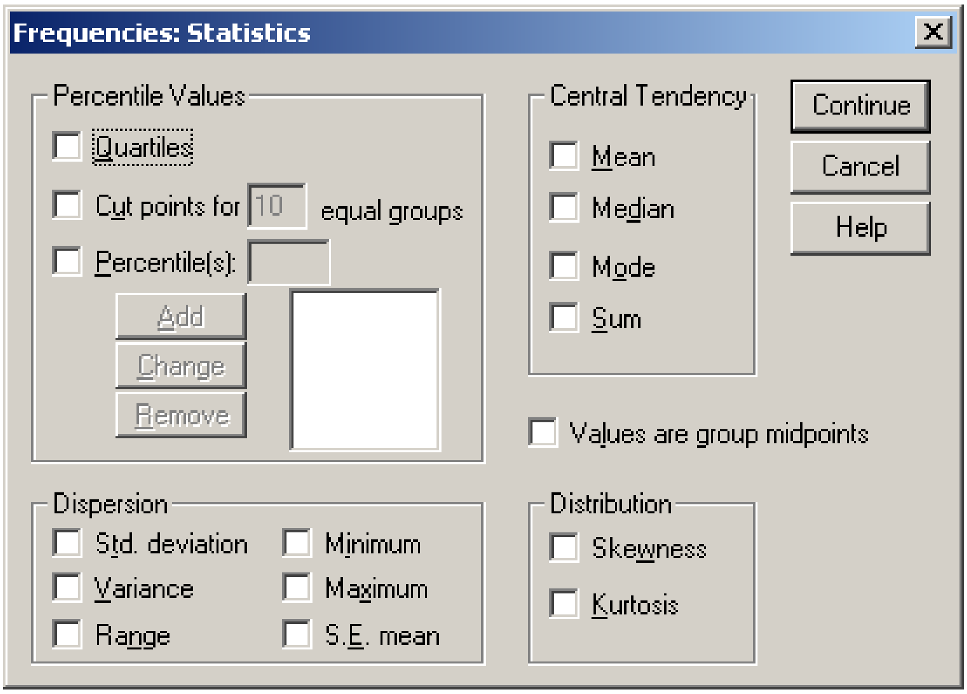
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	id	Numeric	3	0		None	None	8	Right
2	sex	Numeric	3	0	sex	{1, MALES}...	None	8	Right
3	age	Numeric	3	0		None	None	8	Right
4	marital	Numeric	8	0	marital status	{1, SINGLE}...	None	8	Right
5	child	Numeric	5	0	child	{1, YES}...	None	8	Right
6	educ	Numeric	5	0	highest educ c	{1, PRIMARY}.	None	8	Right
7	source	Numeric	8	0	source of stres	{1, WORK}...	None	8	Right
8	smoke	Numeric	8	0	smoker	{1, YES}...	None	8	Right
9	smokenum	Numeric	8	0		None	None	8	Right
10	op1	Numeric	6	0		None	None	8	Right
11	op2	Numeric	6	0		None	None	8	Right
12	op3	Numeric	6	0		None	None	8	Right
13	op4	Numeric	6	0		None	None	8	Right
14	op5	Numeric	6	0		None	None	8	Right
15	op6	Numeric	6	0		None	None	8	Right
16	mast1	Numeric	6	0		None	None	8	Right
17	mast2	Numeric	6	0		None	None	8	Right
18	mast3	Numeric	6	0		None	None	8	Right
19	mast4	Numeric	6	0		None	None	8	Right
20	mast5	Numeric	6			None	None	8	Right

SPSS Processor is ready

Standard practice would be to check the initial frequency count first, but Pallant goes direct to optional statistics from frequencies to find her cutting points and summary statistics, and produces a frequency count almost by default. Age itself may be a continuous variable, but the variable recorded here is ***age last birthday****. Pallant's calculated mean is actually inaccurate: for a more accurate mean you should really add 0.5, but the cutting points remain the same, even after this adjustment.*

Frequencies…

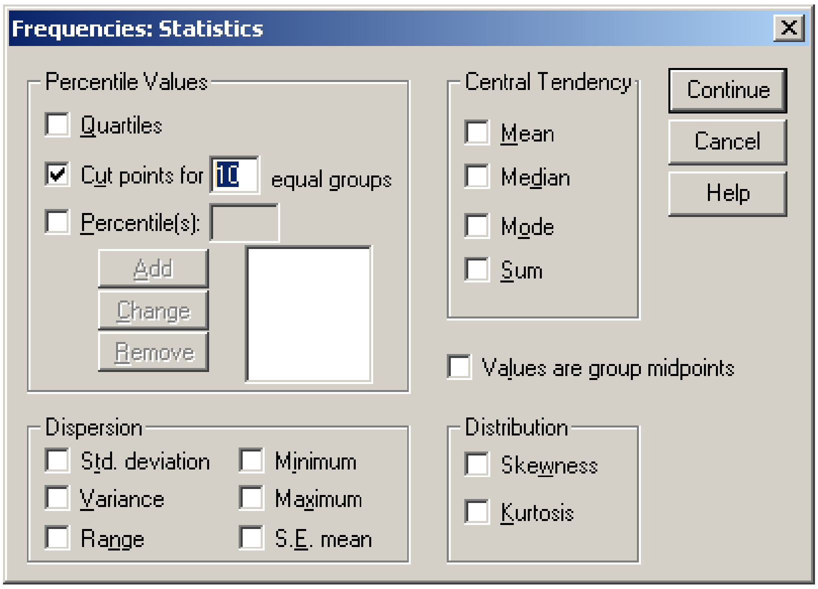
Statistics…



□ Quartiles
□ Cut points for 10 equal groups
□ Percentile(s):
□ Mean
□ Median
□ Mode
□ Sum
□ Values are group midpoints
□ Std. deviation
□ Minimum
□ Variance
□ Maximum
□ Range
□ S.E. mean
□ Skewness
□ Kurtosis

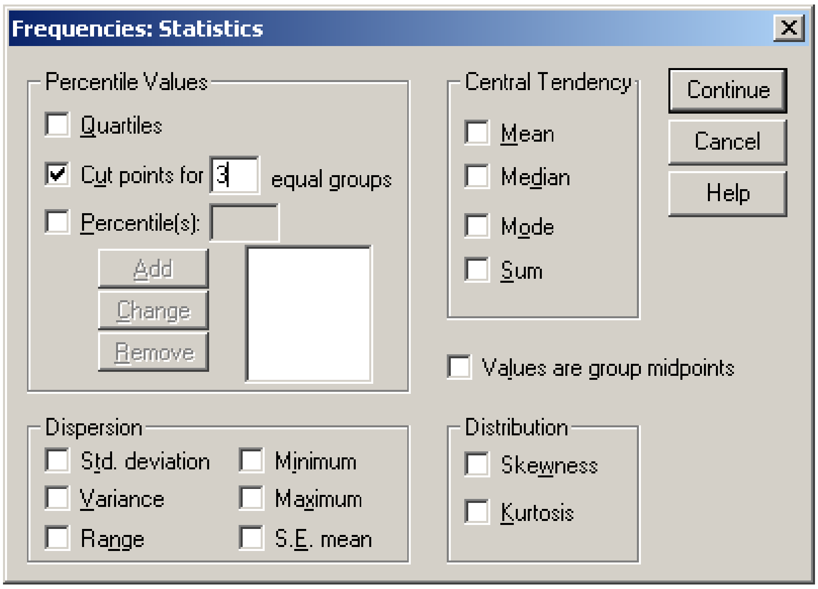
[45] Pallant 2005 uses a new facility **Visual Bander** to find the cutting points for this, so I've used the 2001 book, p82 instead.

In the **Percentile Values** panel, click on **Cut points for ….**



☑ Cut points for 10 equal groups

and enter **3** in the box **for [   ] equal groups**:



☑ Cut points for 3 equal groups

NB Pallant has missed a step out here as she doesn't show how to get the required statistics: are they automatic? No, but they're produced in the summary table:

Continue…

OK

Descriptive Statistics

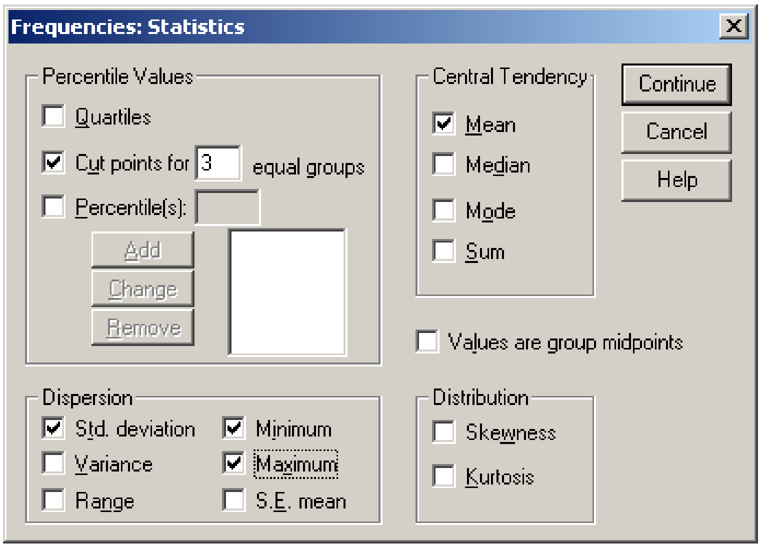
	N	Minimum	Maximum	Mean	Std. Deviation
AGE	439	18	82	37.44	13.202
Valid N (listwise)	439				

Statistics

AGE

N	Valid	439
	Missing	0
Percentiles	33.33333333	29.00
	66.66666667	44.00

To get Pallant's version, you have to tick the boxes for Mean in the Central Tendency panel and the boxes for Std. deviation, Min and Max in the Dispersion panel. This could be very confusing for beginners.


Frequencies: Statistics

Percentile Values
□ Quartiles
☑ Cut points for 3 equal groups
□ Percentile(s):

Central Tendency
☑ Mean
□ Median
□ Mode
□ Sum

□ Values are group midpoints

Dispersion
☑ Std. deviation
☑ Minimum
□ Variance
☑ Maximum
□ Range
□ S.E. mean

Distribution
□ Skewness
□ Kurtosis

Continue which produces:

Statistics

AGE

N	Valid	439
	Missing	0
Mean		37.44
Std. Dev iation		13.202
Minimum		18
Maximum		82
Percentiles	33.33333333	29.00
	66.66666667	44.00

and also, not shown in the book, the frequency table, the Cumulative Percent column of which could just as easily have been used to find these or any other cutting points visually. The table can be suppressed by setting an upper limit to the number of categories under Formats, but Pallant does not demonstrate this. The condensed format subcommand for frequency counts of variables with many values, /format condense, is not implemented in SPSS11 for Windows.

]

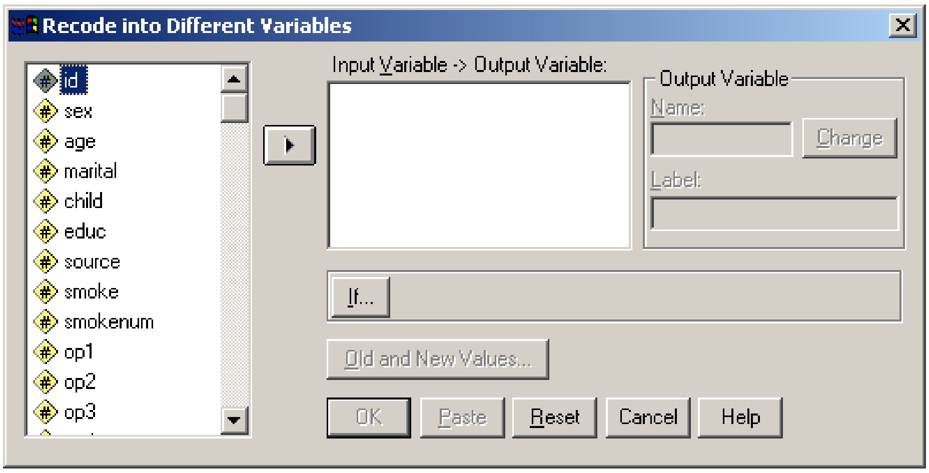
AGE

		Frequency	Percent	Valid Percent	Cumulativ e Percent
Valid	18	6	1.4	1.4	1.4
	19	5	1.1	1.1	2.5
	20	7	1.6	1.6	4.1
	21	22	5.0	5.0	9.1
	22	20	4.6	4.6	13.7
	23	17	3.9	3.9	17.5
	24	17	3.9	3.9	21.4
	25	12	2.7	2.7	24.1
	26	13	3.0	3.0	27.1
	27	15	3.4	3.4	30.5
	28	10	2.3	2.3	32.8
	29	5	1.1	1.1	33.9
	30	8	1.8	1.8	35.8
	31	15	3.4	3.4	39.2
	32	9	2.1	2.1	41.2
	33	12	2.7	2.7	44.0
	34	7	1.6	1.6	45.6
	35	14	3.2	3.2	48.7
	36	18	4.1	4.1	52.8
	37	10	2.3	2.3	55.1
	38	8	1.8	1.8	56.9
	39	7	1.6	1.6	58.5
	40	7	1.6	1.6	60.1
	41	13	3.0	3.0	63.1
	42	11	2.5	2.5	65.6
	43	4	.9	.9	66.5
	44	10	2.3	2.3	68.8
	45	9	2.1	2.1	70.8
	46	13	3.0	3.0	73.8
	47	8	1.8	1.8	75.6
	48	17	3.9	3.9	79.5
	49	10	2.3	2.3	81.8
	50	13	3.0	3.0	84.7
	51	6	1.4	1.4	86.1
	52	4	.9	.9	87.0
	53	4	.9	.9	87.9
	54	5	1.1	1.1	89.1
	55	6	1.4	1.4	90.4
	56	6	1.4	1.4	91.8
	57	3	.7	.7	92.5
	58	3	.7	.7	93.2
	59	1	.2	.2	93.4
	60	5	1.1	1.1	94.5
	61	2	.5	.5	95.0
	62	2	.5	.5	95.4
	63	3	.7	.7	96.1
	64	1	.2	.2	96.4
	65	2	.5	.5	96.8
	66	2	.5	.5	97.3
	67	2	.5	.5	97.7
	68	2	.5	.5	98.2
	69	1	.2	.2	98.4
	70	2	.5	.5	98.9
	74	2	.5	.5	99.3
	75	1	.2	.2	99.5
	78	1	.2	.2	99.8
	82	1	.2	.2	100.0
	Total	439	100.0	100.0	

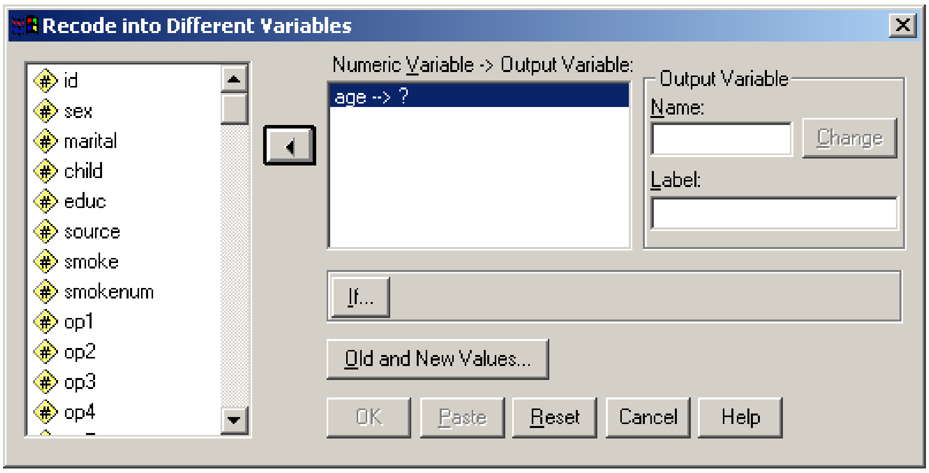
Step 2: Grouping the values

As in the previous exercise the new variable already exists in the file distributed by the Open University. SPSS will not allow it to be overwritten, so yet again I had to overcome this by deleting the variable from the file and recreating it from scratch.

Transform
Recode
Into Different Variables

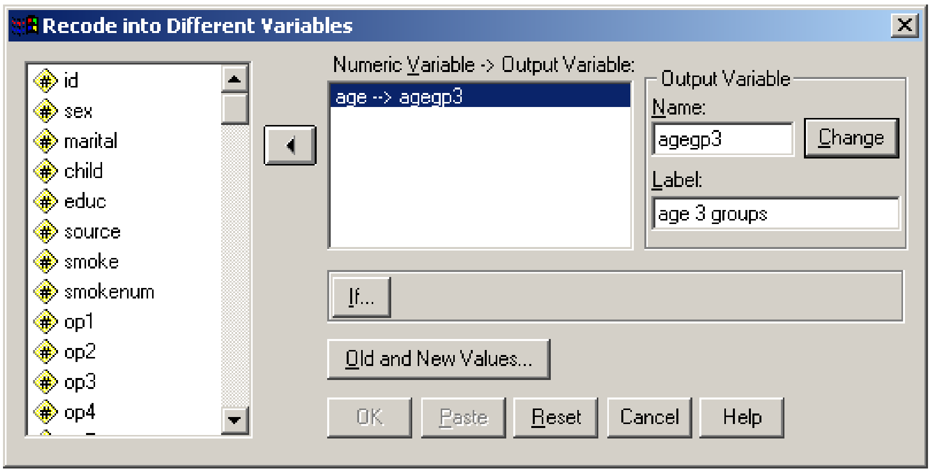


Highlight **age** and click on **right arrow** to transfer it to the **Input Variable** box which now becomes the **Numeric Variable → Output Variable** box:



Numeric Variable -> Output Variable: age --> ?

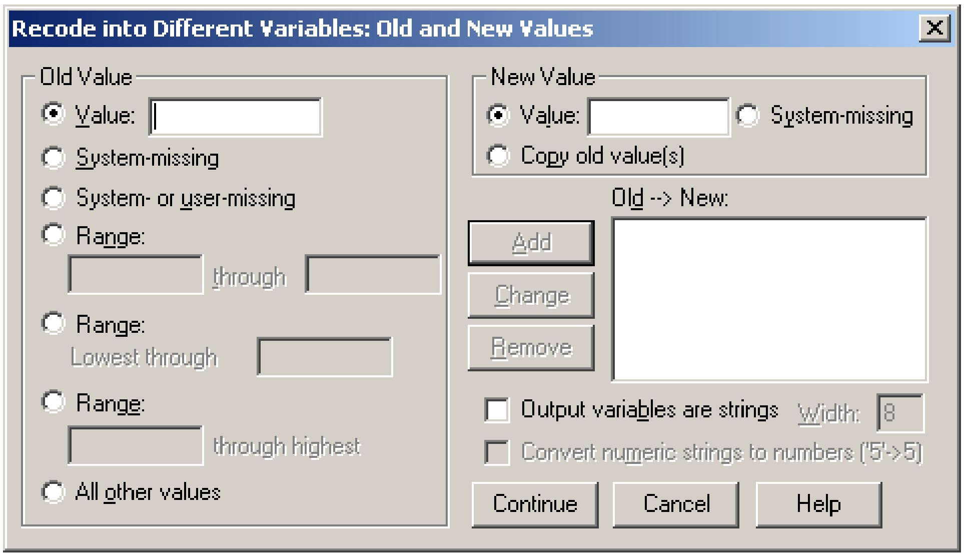
Type **agegp3** in **Output Variable Name** box, and **age 3 groups** in the label box then **change**:



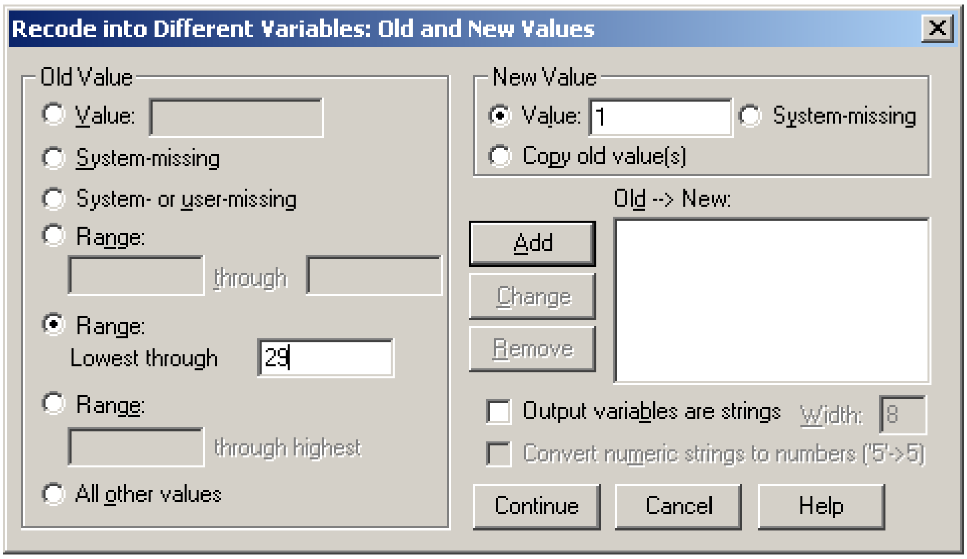
Numeric Variable -> Output Variable: age --> agegp3

Output Variable
Name: agegp3
Label: age 3 groups

Click **Old and New Values**



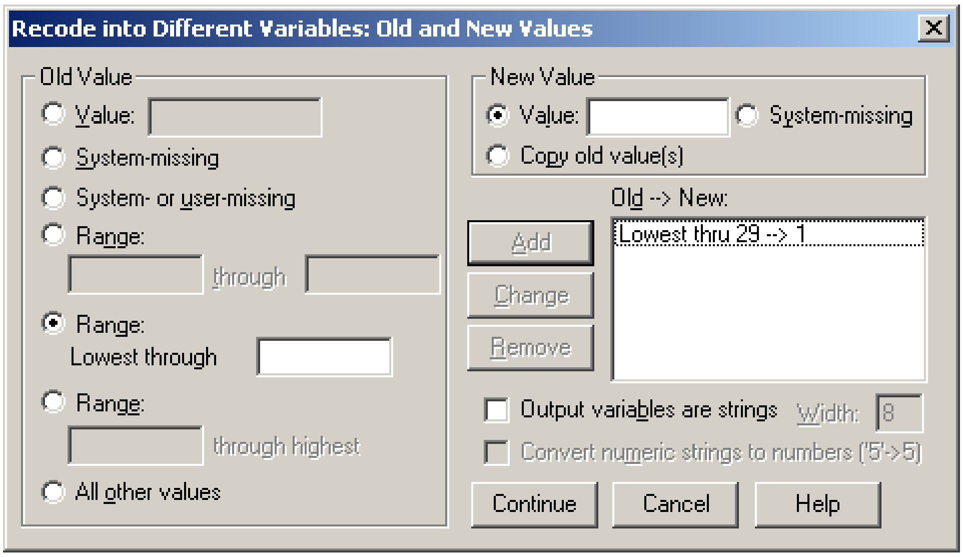
then tick **Range Lowest thru** and enter **29** in the box and **1** in the **Value** box of the **New Value** panel



Lowest through: 29

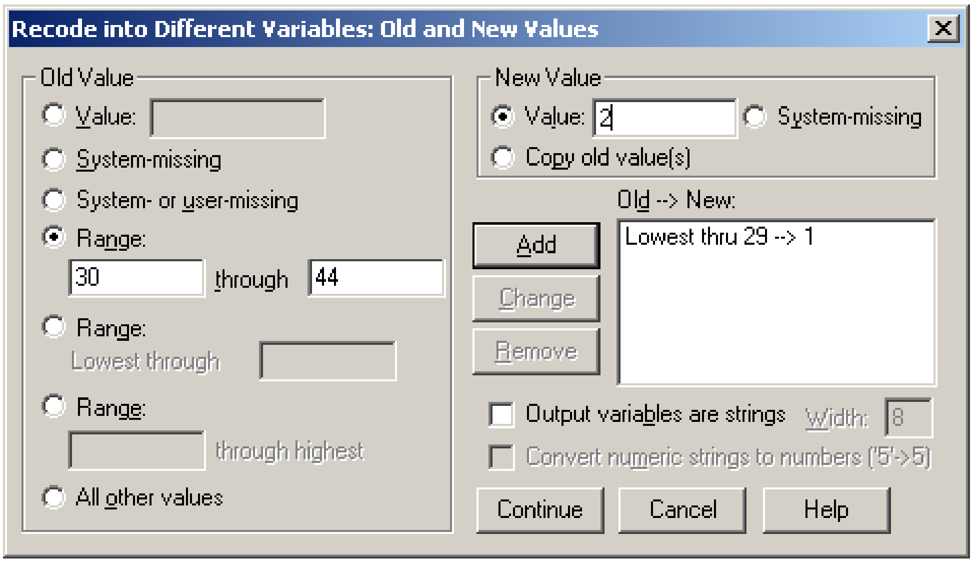
New Value: 1

and click **Add**



Old --> New: Lowest thru 29 --> 1

Click on **Range** and enter values **30 [through] 44**: enter **2** in the **Value** box of the **New Value** panel



Recode into Different Variables: Old and New Values

Old Value

● Range: 30 through 44

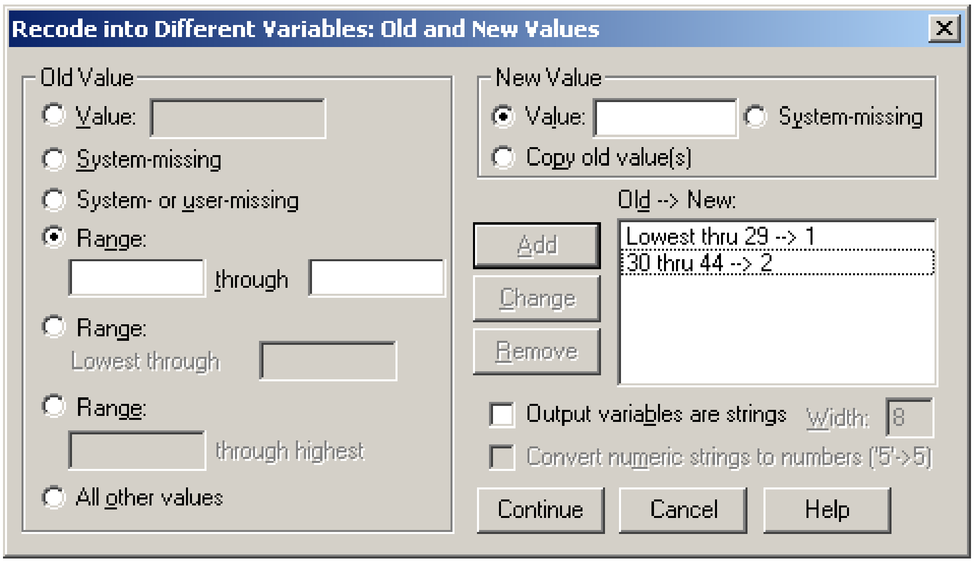
New Value

● Value: 2

Old --> New:

Lowest thru 29 --> 1

Click on **Add**



Recode into Different Variables: Old and New Values

Old Value

● Range: through

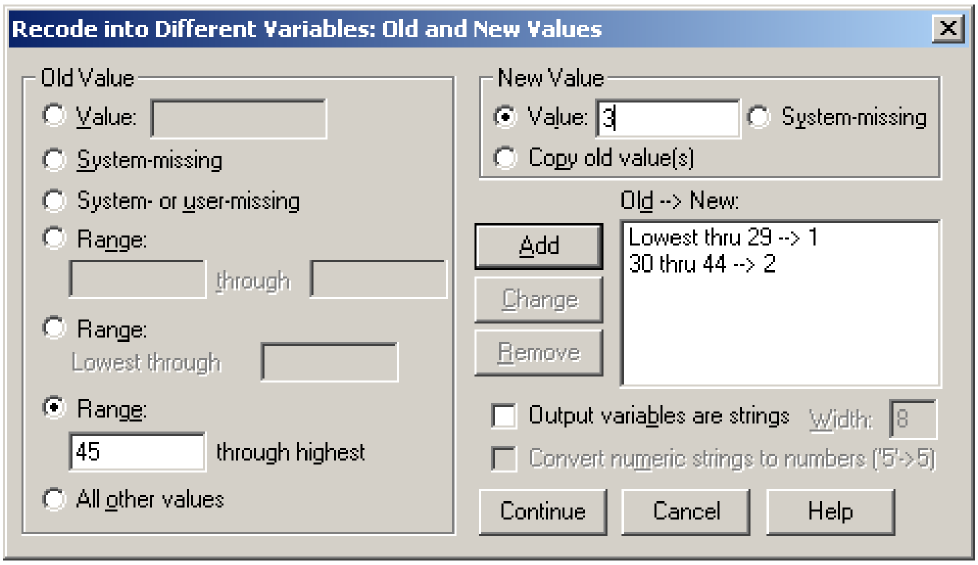
New Value

● Value:

Old --> New:

Lowest thru 29 --> 1
30 thru 44 --> 2

Finally tick **Range through highest**, enter **45**: then enter **3** in the **Value** box of the **New Value** panel



Recode into Different Variables: Old and New Values

Old Value

● Range: 45 through highest

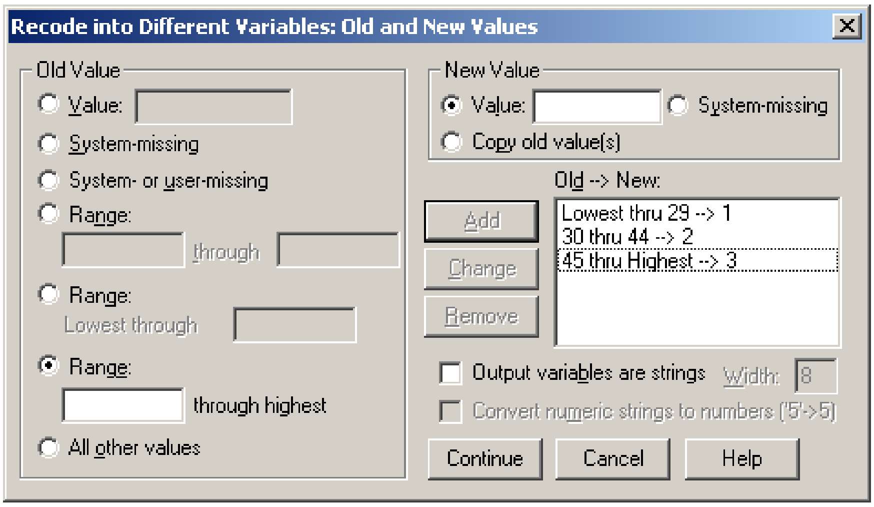
New Value

● Value: 3

Old --> New:

Lowest thru 29 --> 1
30 thru 44 --> 2

and click on **Add**



Recode into Different Variables: Old and New Values

Old Value
- ○ Value:
- ○ System-missing
- ○ System- or user-missing
- ○ Range: through
- ○ Range: Lowest through
- ● Range: through highest
- ○ All other values

New Value
- ● Value: ○ System-missing
- ○ Copy old value(s)

Old --> New:

Lowest thru 29 --> 1
30 thru 44 --> 2
45 thru Highest --> 3

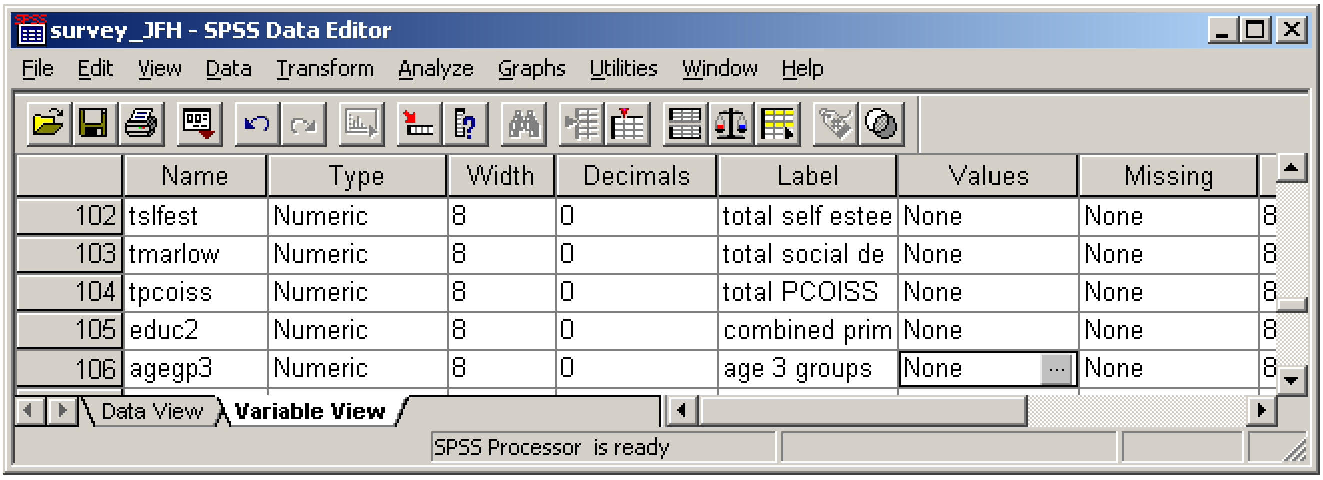
□ Output variables are strings Width: 8
□ Convert numeric strings to numbers ('5'->5)

and **Continue**

***Although this is only an exercise, the use of* Lowest thru *and* thru Highest *can be dangerous unless you are certain there are no non-missing values outside the valid range. A safer way would have been to use:* 18 thru 29 --> 1 30 thru 44 --> 2, 45 thru 82 --> *3 and* ELSE -- > SYSMIS.**

Step 3: Adding in Value Labels

[CTRL]+End or scroll down to the new variable **agegp3** (which will be appended at the end of the file) and click on the **Values** cell:



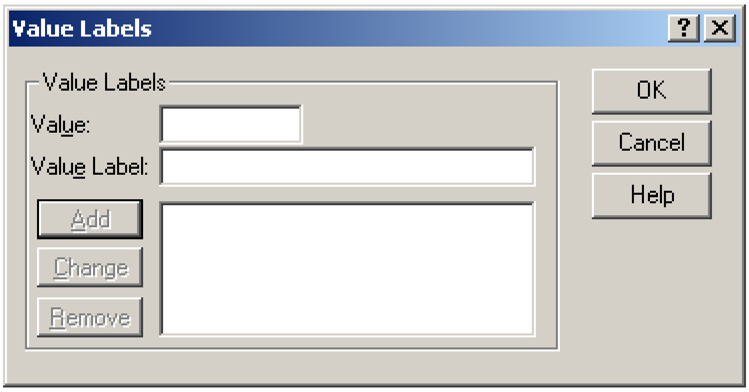
survey_JFH - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	
102	tslfest	Numeric	8	0	total self estee	None	None	8
103	tmarlow	Numeric	8	0	total social de	None	None	8
104	tpcoiss	Numeric	8	0	total PCOISS	None	None	8
105	educ2	Numeric	8	0	combined prim	None	None	8
106	agegp3	Numeric	8	0	age 3 groups	None ...	None	8

SPSS Processor is ready

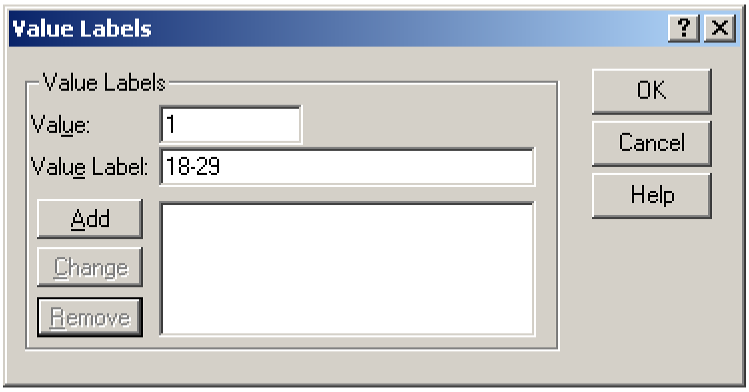
..then on the gray patch to bring up a new window, the **Value Labels** dialog panel:



Value Labels

Value Labels
Value:
Value Label:

Enter **1** in the **Value** box and **18-29** in the **Value Label** box

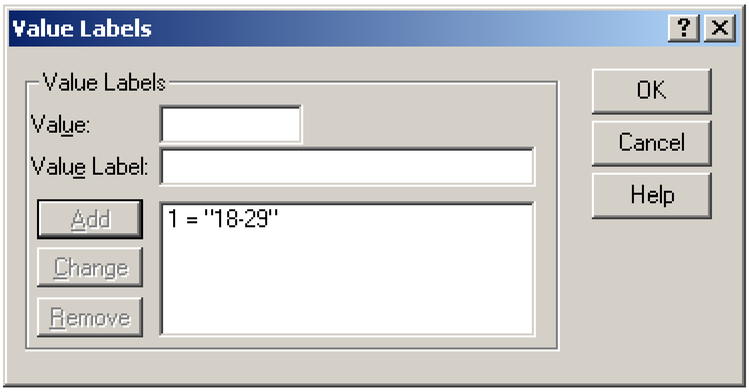


Value Labels

Value: 1

Value Label: 18-29

then **Add**



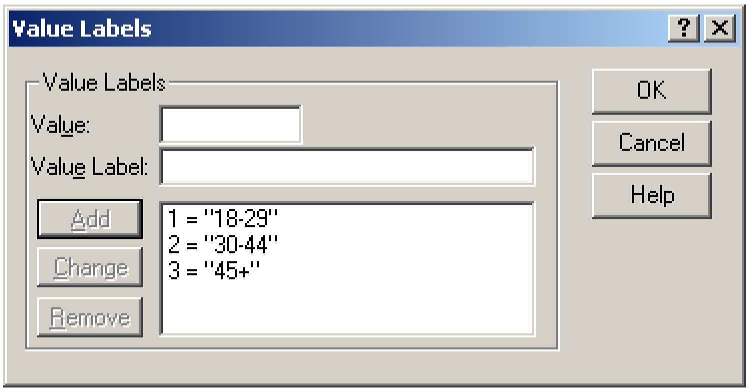
Value Labels

Value:

Value Label:

1 = "18-29"

Now do the same for values **2 = "30-44"** and **3 = "45+"** then **Add**



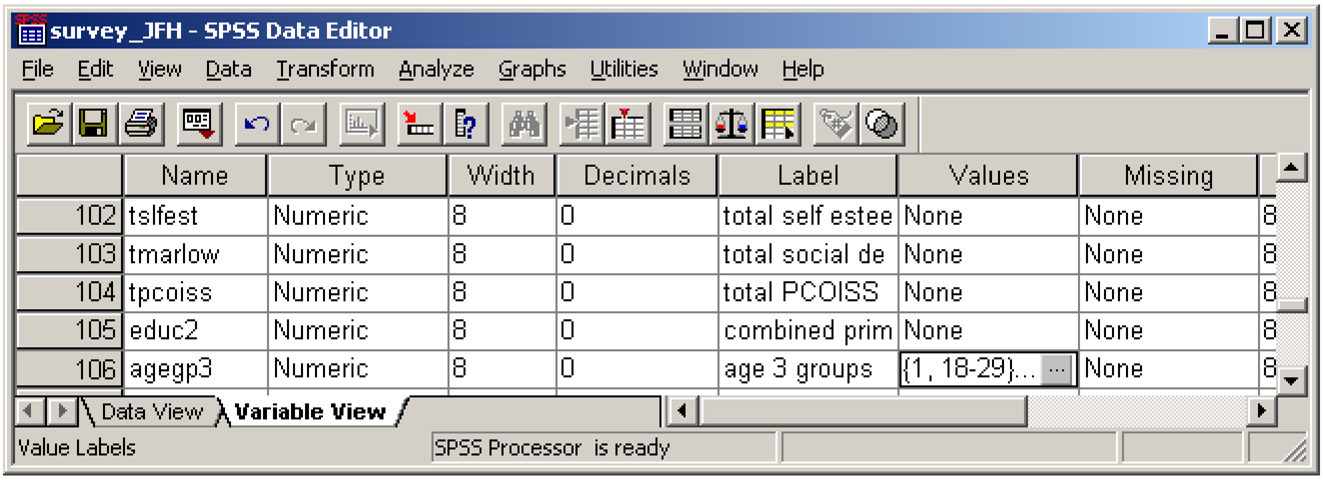
Value Labels

Value:

Value Label:

1 = "18-29"
2 = "30-44"
3 = "45+"

and click **OK**. The cell in the **Values** column in the **Data Editor** will now have changed:


survey_JFH - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View

	Name	Type	Width	Decimals	Label	Values	Missing	
102	tslfest	Numeric	8	0	total self estee	None	None	8
103	tmarlow	Numeric	8	0	total social de	None	None	8
104	tpcoiss	Numeric	8	0	total PCOISS	None	None	8
105	educ2	Numeric	8	0	combined prim	None	None	8
106	agegp3	Numeric	8	0	age 3 groups	{1, 18-29}...	None	8

Value Labels

SPSS Processor is ready

[NB: I have not found a paste facility for this: is there one??]

Again standard procedure would be to check the frequencies for the new variable: Pallant does not do it for this example. In general, it is better to check on initial frequencies for variables to be recoded, as there is always a risk of outliers and missing values being erroneously included in the recodes. Anyway:

Analyse…
Frequencies …
etc., etc…

AGEGP3 age 3 groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 18 - 29	149	33.9	33.9	33.9
	2 30 - 44	153	34.9	34.9	68.8
	3 45+	137	31.2	31.2	100.0
	Total	439	100.0	100.0	

[NB: If you do this with drop-down menus, previous options and statistics are retained (because of the automatic EXECUTE and you will finish up with more output than you wanted! ie:]

Statistics

AGEGP3 age 3 groups

N	Valid	439
	Missing	0
Mean		1.97
Std. Deviation		.808
Minimum		1
Maximum		3
Percentiles	33.33333333	1.00
	66.66666667	2.00

which is clearly nonsense! As before syntax is quicker (from **PASTE**):

```
RECODE
   age (Lowest thru 29=1)  (30 thru 44=2)  (45 thru Highest=3)  INTO  agegp3 .
VARIABLE LABELS agegp3 'age 3 groups'.
EXECUTE .
FREQUENCIES
  VARIABLES=age
 /NTILES= 3
 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN
 /ORDER=  ANALYSIS .
```

but to suppress the frequency count table (from **PASTE**):

```
FREQUENCIES
  VARIABLES=age
 /FORMAT=LIMIT(10)  /NTILES= 3
 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN
 /ORDER=  ANALYSIS .
```

[**NB: 10** is the default limit for **format**, but it can be set at any value]

But forget **PASTE**. I would have done (as a preliminary check on value range, outliers, unexpected values etc.)

```
freq var age.
```

…then to get Pallant’s cutting points: (which you can do just as easily from the frequency count)

```
freq var age /for not/ per 33 67.
```

```
              Statistics

AGE
N             Valid          439
              Missing          0
Percentiles   33           29.00
              67           44.00
```

For the recode (since you already have the valid range from the frequency check):

```
recode age
          (18 thru  29 =1)
          (30 thru 44 = 2)
          (45 thru 82 = 3)
          (else=sysmis)
    into agegrp3.
```

…for the labels:

```
var lab  agegrp3 ‘Grouped age (3 groups)’.
val lab  agegrp3 1 ’18 - 29’ 2 ’30 - 44’ 3 ‘45+’.
```

…and for the frequency check on the new group:

```
freq var agegrp3.
```

Thus, all the above laborious steps in drop-down menus could have been done (again in seconds) with:

```
recode  age  (18 thru  29 =1)(30 thru 44 = 2)
             (45 thru 82 = 3)(else=sysmis)
        into agegrp3.
var lab  agegrp3 ' Grouped age (3 groups)'.
val lab  agegrp3 1 '18 - 29' 2 '30 - 44' 3 '45+'.
freq var  agegrp3.
```

AGEGRP3 Grouped age (3 groups)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 18 - 29	149	33.9	33.9	33.9
	2 30 - 44	153	34.9	34.9	68.8
	3 45+	137	31.2	31.2	100.0
	Total	439	100.0	100.0	

…but to be statistically correct:

```
comp age = age+0.5.
freq var age /for not /sta mea std /per 33 67.
```

Statistics

AGE

N	Valid	439
	Missing	0
Mean		37.94
Std. Deviation		13.202
Percentiles	33	29.50
	67	44.50

…which is more accurate, but doesn't change the groupings needed for **recode** into **agegrp3**

Exercise 4: Correlation (2005, pp78ff)

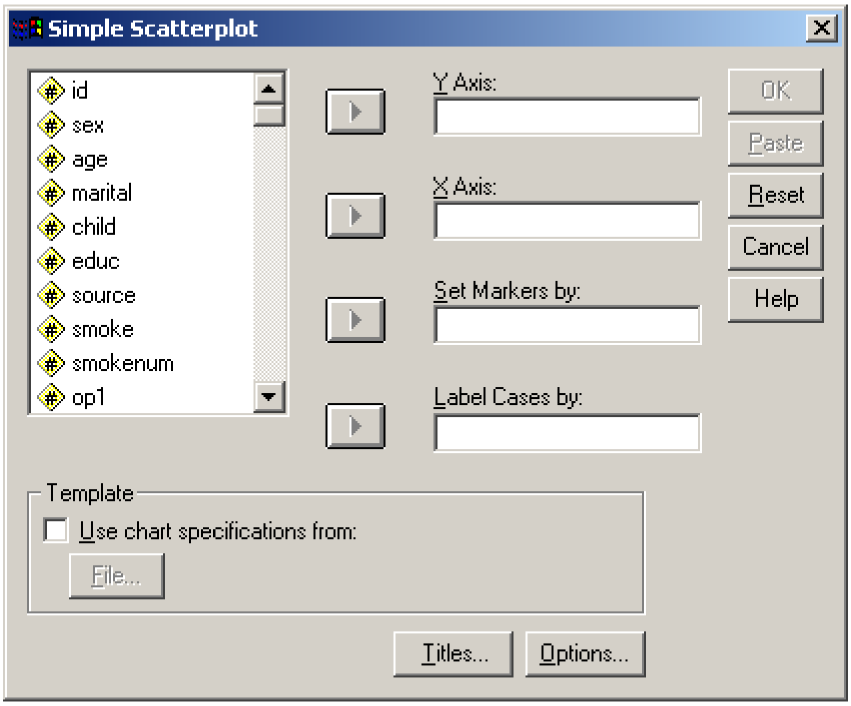
I found a major problem repeating this exercise from the book. Although the variable names are given in an initial decision-making table, they are not given in the actual exercises. The sample outputs use variable ***labels****, not variable* ***names****, and this is confusing, especially for the second exercise for which,* ***"Click on the first variable"*** *is not particularly helpful. However, having worked out which variables (****tpcoiss*** *and* ***tpstress****) are supposed to be used:*

Step 1: Scatterplot

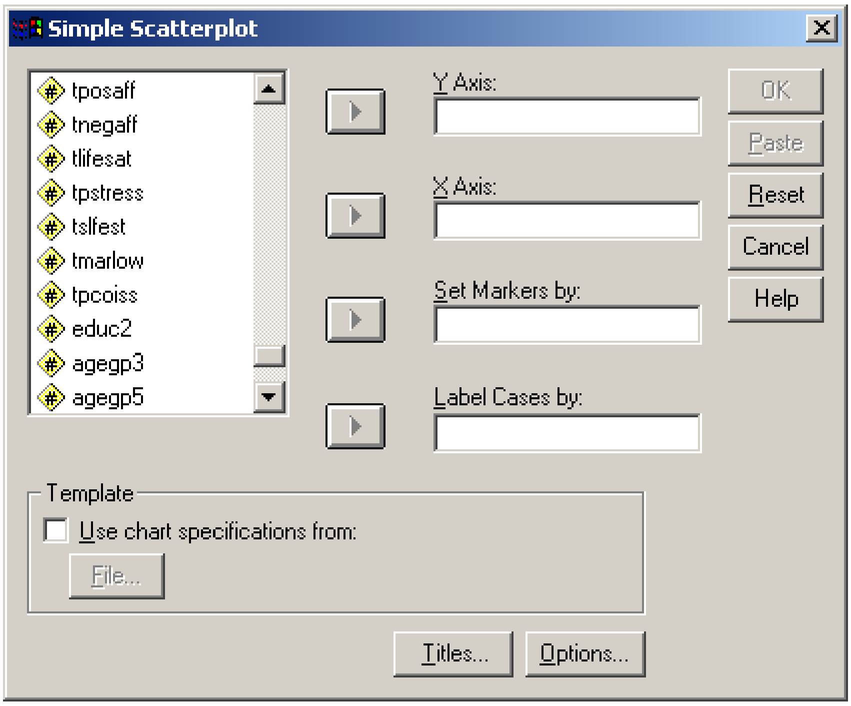
Graphs

Scatter [**Simple** will be highlit]

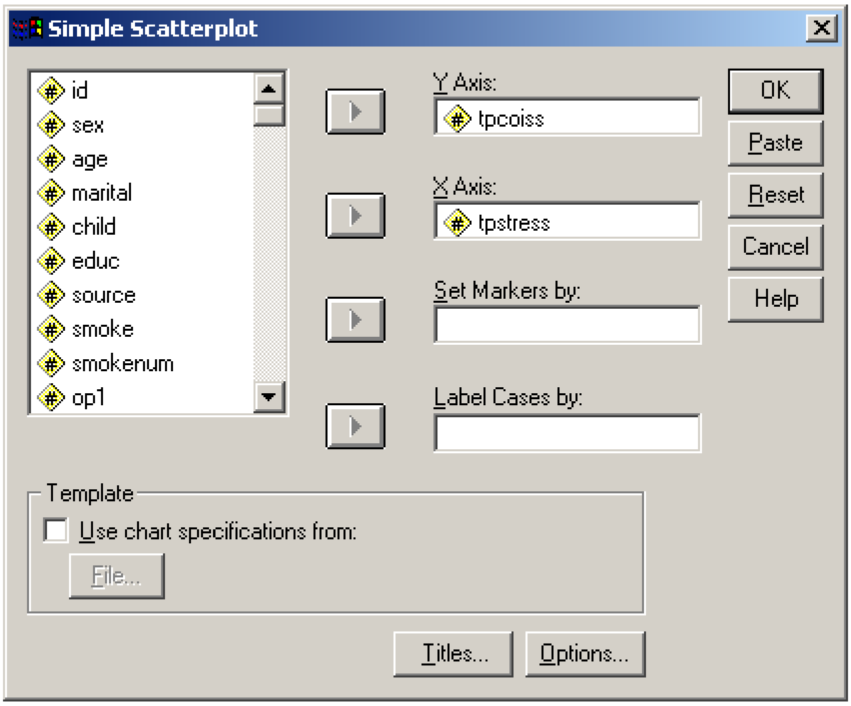
Define



Scroll down to find the variables



Click and drag **tpstress** to the **Y Axis** box and **tpcoiss** to the **X Axis** box



Y Axis: tpcoiss

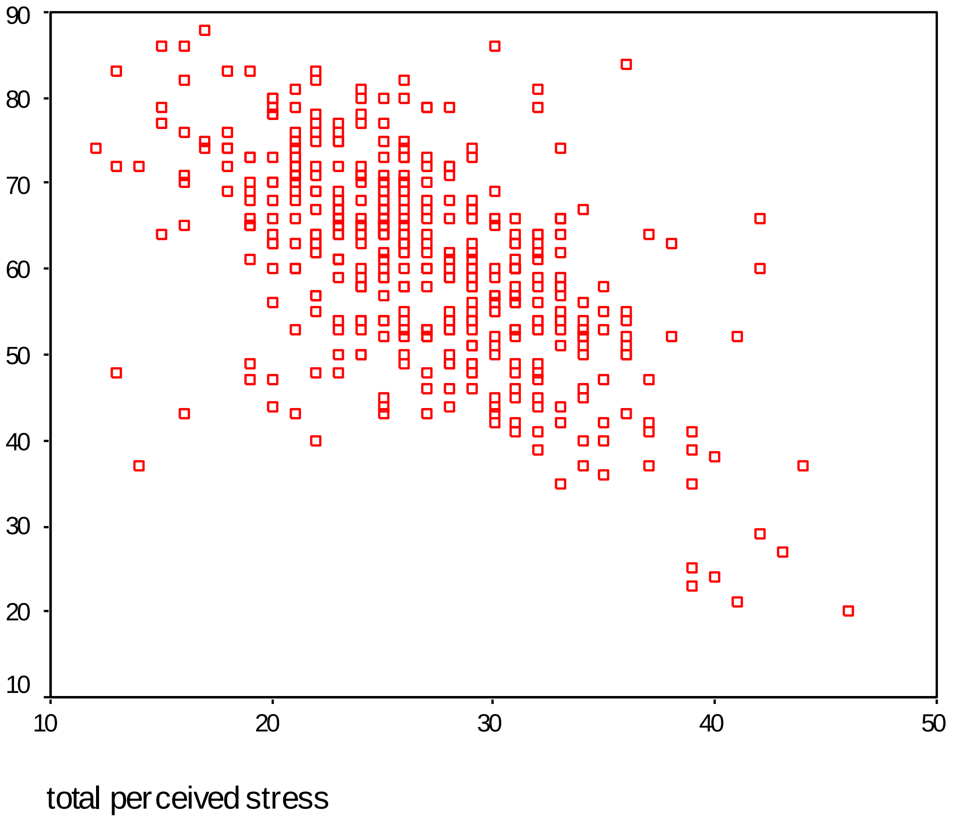
X Axis: tpstress

Set Markers by:

Label Cases by:

☐ Use chart specifications from:

then click **OK**



PASTE produces:

```
GRAPH
  /SCATTERPLOT(BIVAR)=tpstress WITH tpcoiss
  /MISSING=LISTWISE .
```

…but it's quicker to write: **GRAPH /SCAT tpstress WITH tpcoiss.**
since **LISTWISE** is the default.

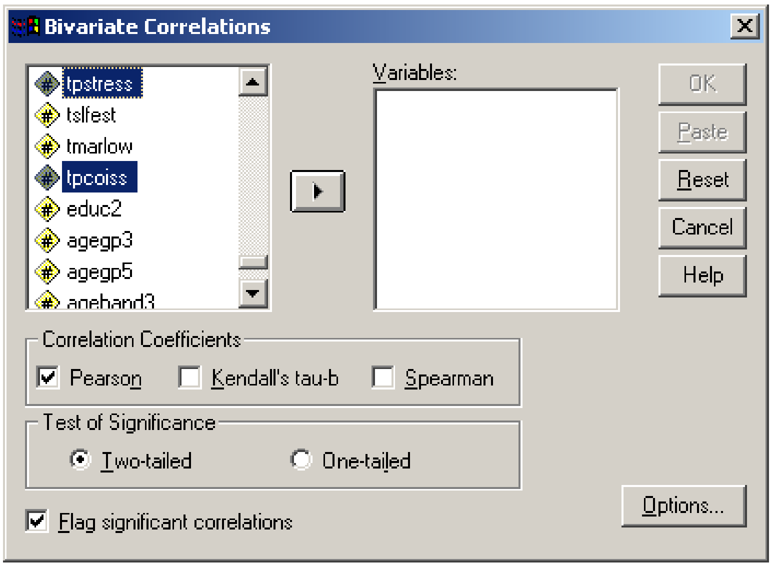
Step 2: Correlation

Analyse

Correlate

Bivariate

Scroll down and highlight variables **tpstress** and **tpcoiss**:



Correlation Coefficients

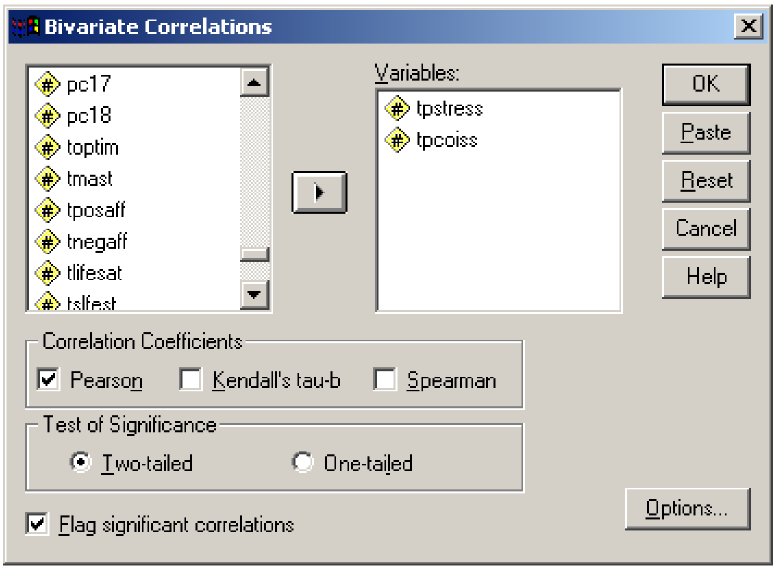
☑ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance

● Two-tailed ○ One-tailed

☑ Flag significant correlations

Click on **right arrow** to place **tpstress** and **tpcoiss** in the **Variables** box:



Correlation Coefficients

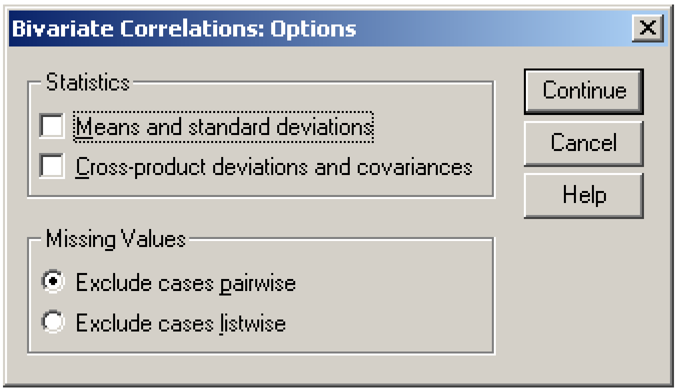
☑ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance

● Two-tailed ○ One-tailed

☑ Flag significant correlations

Check that **Pearson** is ticked in the **Correlation Coefficients** box and **Two-tailed** selected in **Test of Significance** then click on **Options**



Statistics

☐ Means and standard deviations

☐ Cross-product deviations and covariances

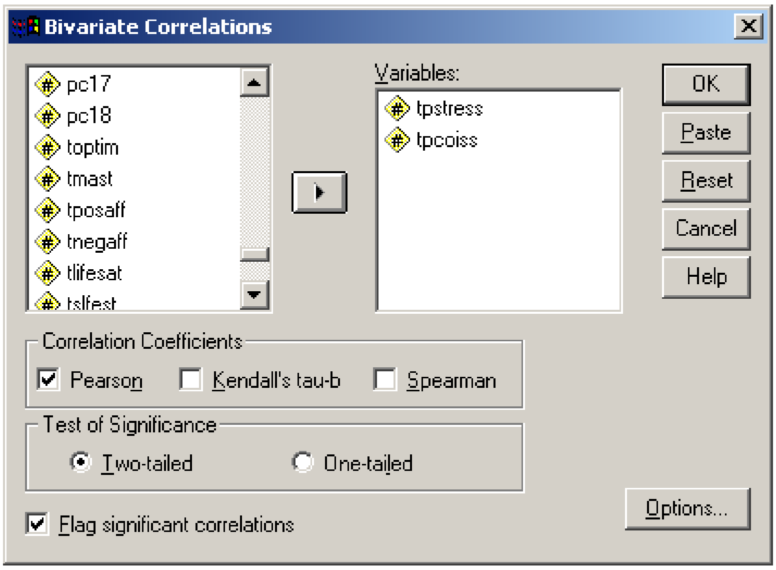
Missing Values

● Exclude cases pairwise

○ Exclude cases listwise

Check in **Missing Values** box that **Exlude cases pairwise** is selected then click

Continue



pc17
pc18
toptim
tmast
tposaff
tnegaff
tlifesat
tslfest

Variables:
tpstress
tpcoiss

Correlation Coefficients
☑ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance
● Two-tailed ○ One-tailed

☑ Flag significant correlations

…and **OK**

Correlations

		TPCOISS total PCOISS	TPSTRESS total perceived stress
TPCOISS total PCOISS	Pearson Correlation	1	-.581**
	Sig. (2-tailed)	.	.000
	N	430	426
TPSTRESS total perceived stress	Pearson Correlation	-.581**	1
	Sig. (2-tailed)	.000	.
	N	426	433

**. Correlation is significant at the 0.01 level (2-tailed).

PASTE produces:

```
CORRELATIONS
  /VARIABLES= tpcoiss tpstress
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .
```

. . . but it’s quicker to write:

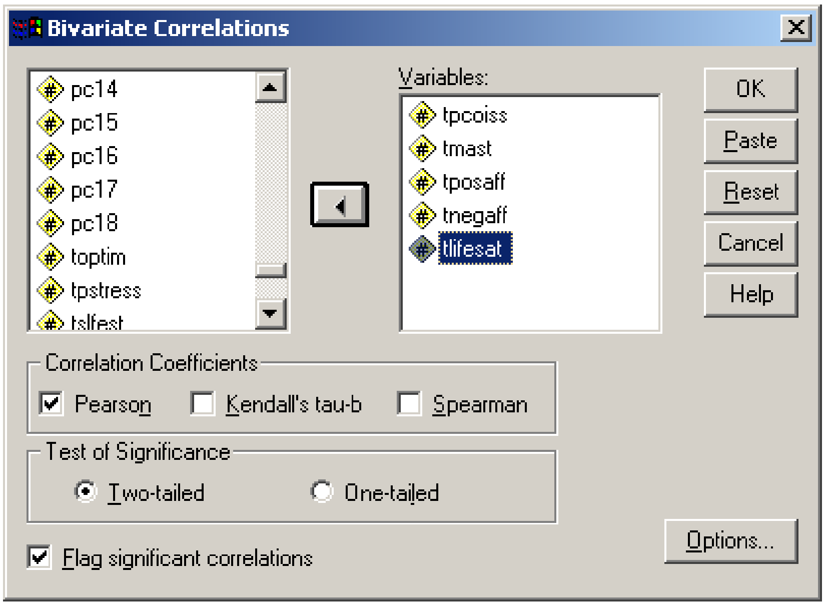
```
corr tpcoiss tpstress
  /pri nos.
```

. . . since **Pearson**, **two-tail** test and **pairwise** deletion are the default.

The same process can be used to generate a matrix of correlations, but Pallant's example of a triangular matrix on page 122 has variable names which do not match those in the file (or their labels). However:

Analyse
Correlate
Bivariate

Scroll down, highlight and transfer variables to **Variables** box:



Variables:
tpcoiss
tmast
tposaff
tnegaff
tlifesat

Correlation Coefficients
☑ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance
● Two-tailed ○ One-tailed

☑ Flag significant correlations

Click on **OK**

Correlations

		TPCOISS total PCOISS	TMAST total mastery	TPOSAFF total positive affect	TNEGAFF total negative affect	TLIFESAT total life satisfaction
TPCOISS total PCOISS	Pearson Correlation	1	.521**	.456**	-.484**	.373*
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	430	429	429	428	429
TMAST total mastery	Pearson Correlation	.521**	1	.432**	-.464**	.444*
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	429	436	436	435	436
TPOSAFF total positive affect	Pearson Correlation	.456**	.432**	1	-.294**	.415*
	Sig. (2-tailed)	.000	.000	.	.000	.000
	N	429	436	436	435	436
TNEGAFF total negative affect	Pearson Correlation	-.484**	-.464**	-.294**	1	-.316*
	Sig. (2-tailed)	.000	.000	.000	.	.000
	N	428	435	435	435	435
TLIFESAT total life satisfaction	Pearson Correlation	.373**	.444**	.415**	-.316**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.
	N	429	436	436	435	436

**. Correlation is significant at the 0.01 level (2-tailed).

Finally, in only the second example in the entire book, Pallant is forced to use syntax (and indirectly at that) to produce correlations of one set of variables with another set. In the above example, **PASTE** produces:

```
CORRELATIONS
  /VARIABLES=tpcoiss tmast tposaff tnegaff tlifesat
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .
```

…and she has to resort to **syntax mode** to insert **with**

```
CORRELATIONS
  /VARIABLES=tpcoiss tmast tposaff with tnegaff tlifesat
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .
```

…to produce:

Correlations

		TNEGAFF total negative affect	TLIFESAT total life satisfaction
TPCOISS total PCOISS	Pearson Correlation	-.484**	.373**
	Sig. (2-tailed)	.000	.000
	N	428	429
TMAST total mastery	Pearson Correlation	-.464**	.444**
	Sig. (2-tailed)	.000	.000
	N	435	436
TPOSAFF total positive affect	Pearson Correlation	-.294**	.415**
	Sig. (2-tailed)	.000	.000
	N	435	436

**. Correlation is significant at the 0.01 level (2-tailed).

which could just as easily have been produced by:

```
corr
  tpcoiss tmast tposaff with tnegaff tlifesat
  /pri nos.
```

Far better to **stick to syntax** or at least use **PASTE**. That way, if you make a mistake you can go (back) to the syntax file, check or amend it and run the job again with **[CTRL]+R** or **RUN**. You can't do this with drop-down menus, except by recalling the **dialog box**, clicking on whichever set of procedures you were using and then amending the dialog box(es), but it would be nice to have a proper **<back** button (**Undo** doesn't seem to work inside **Data Editor**).

After all that, I had to buy one of these!

No, it's not from the Family Planning Clinic: it's a mouse-pad with a soft wrist-support! Perhaps I should have used the title Pilgrim's Progress or Bunyan's Bunion. I would have used Not Waving, but Drowning! but Sylvia Plath got there first.

Point made?

Trebles all round!

Any questions?

Time for tea...

Copyright BBC

...said Zebedee

Copyright 2002 Pathé Pictures

…and that’s one thing you can’t do in syntax!