

A STEP-BY-STEP GUIDE TO
ACCESSING THE
1991 SAR VIA USAR

Ian Turton and Stan Openshaw

WORKING PAPER 94/6

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

A step-by-step guide to accessing the 1991 SAR via USAR

Ian Turton and Stan Openshaw

School of Geography

University of Leeds

Leeds, LS2 9JT

email: ian@geog.leeds.ac.uk, stan@geog.leeds.ac.uk

Contents

1	Introduction	1
2	Getting Started	2
3	USAR Recipes	5
	Recipe 1: A simple tabulation recipe for USAR	5
	Recipe 2: Adding a third variable to the table	20
	Recipe 3: Applying a filter to the dataset	21
	Recipe 4: Formatting and printing tables	29
	Recipe 5: Saving and loading tables	32
	Recipe 6: Other ways of grouping variables	35
	Recipe 7: Table optimisation	41
	Recipe 8: Searching the SAR	44
	Recipe 9: Exploring the SAR	48
4	Advanced Details	54
	4.1 Log Files	54
	4.2 Batch jobs	55

1 Introduction

This guide provides a simple introduction to installing and using the USAR package which is specially designed for use with the 1991 SAR. The 1991 SAR provides a 1% sample of anonymised household records and a 2% sample of anonymised personal records from the 1991 census. These data have been purchased on behalf of the academic community and are available

⁰This manual may be freely copied provided this notice is left intact and the result is not sold for profit.

free of charge for research from the CMU at the University of Manchester, provided that you complete the relevant licenses.

However analysis of the SAR via standard statistical packages is not easy because of the specialist nature of the data. This guide describes how users of unix systems can avoid many of the problems of using the SAR by making use of specially developed software (USAR - Unix SAR system); see (Turton and Openshaw 1994). This provides a quick and easy means of generating tables, and performing various exploratory analysis with the SAR data using a simple interactive system that runs under UNIX. It might be expected that useful results can be obtained in a very short time with USAR. USAR is designed to provide a simple one finger user interface that is robust.

This guide is a complete manual covering all aspects of acquiring the SAR data, installing USAR and using USAR to perform various analyses.

2 Getting Started

Step 1: Are you a registered SAR user at a registered SAR site?

If yes then go to step 2; otherwise you will need to contact your CMU representative. This will usually be the same person that is responsible for links between your institution and the MCC. If you are unable to find out who this is contact:

Ruth Durrell
Administrator
Census Microdata Unit
Faculty of Economic and Social Studies
University of Manchester
Manchester, M13 9PL
email: cmu@man.ac.uk
tel: 061-275-4721
fax: 061-275-4722

who will be able to assist you.

Once you have located your CMU representative they will provide you with the necessary user registration form.

Commercial users of the data must contact the CMU for licensing details.

Step 2: Have you got a copy of USAR on your workstation or on a unix system that you can access?

If yes, then go to section 3.

If no, then there are three options:

1. you can load USAR on your workstation;
2. you can ask your computing services to install USAR either on your workstation or on a machine you can access; or,
3. you can obtain sufficient disk space on a general unix machine and install USAR there.

To proceed you will need at least 30Mb and preferably 65Mb of disk space. The best way of transferring the software is to use ftp, if you have access to JANET (or other section of the INTERNET). Failing that then you will need to contact the School of Geography at the University of Leeds to discuss other methods. However it is quicker and easier if you get the software directly yourself.

This description describes a generic unix ftp session, which may be slightly different at your site. Talk to your local experts if you have any difficulties. Text like this is from the computer, text like this is what you should type. Comments describing the process are in [].

```
prompt: ftp gam.leeds.ac.uk
Connected to gam.leeds.ac.uk
220 gam FTP server (Version 5.60) ready.
Name (gam:ian): anonymous
331 Guest login ok, send ident as password.
Password: your user name
230 Guest login ok, access restrictions apply.
ftp> cd pub/usar [change directory - note / not \]
250 CWD command successful.
ftp> ls [list the files in the directory]
200 PORT command successful.
150 Opening ASCII mode data connection for file list.
usar-0.96b.tar.Z
usar-0.96.tar.Z
usar-0.96a.tar.Z
usar-0.97.tar.Z
usar-0.98.tar.Z
226 Transfer complete.
62 bytes received in 0.0023 seconds (26 Kbytes/s)
ftp> binary [All these files are binary so we must set this]
200 Type set to I.
```

```

ftp> get usar-0.98.tar.Z
[ in general choose the file with the highest number]
200 PORT command successful.
150 Opening BINARY mode data connection for usar-0.98.tar.Z (95281 bytes).
226 Transfer complete.
local: usar-0.98.tar.Z remote: usar-0.98.tar.Z
95281 bytes received in 1.1 seconds (82 Kbytes/s)
ftp> quit
221 Goodbye.
prompt:

```

Then you will need to unpack the file you fetched in the last section.

```
prompt: zcat usar-0.98.tar.Z | tar -xvf -
```

This will create the source files for USAR. Then you can edit the Makefile to reflect your system's setup. First if you don't have the gcc compiler add a comment (#) at the start of the first two lines and delete the # from the next two lines. Next decide where the USAR datafiles are to live and add this directory to the line

```
f1 = \"/home1/ian/work/\" # .....
```

between the ". This must be a full directory path from the root and must end in a /. If you want to make use of the compressed data files in USAR then delete the comment character from the start of the line

```
# f2 = -DCOMPRESSED # .....
```

but bear in mind that this will double the time taken to create tables, whilst halving the disk space used.

Next check the libraries listed in the line that starts LIBS. Do not worry too much if you are not sure about these.

Then type make, and with any luck you should see a series of successful compiles.

If the compiler complains about being unable to find bcopy, edit the makefile and uncomment the line

```
# f4 = -DNEEDBCOPY # .....
```

and type make again.

If the linker complains about being unable to find functions such as tputs then you probably guessed wrong earlier about the libraries on your machine. Just comment out the line that starts LIBS and uncomment the next line.

If the linker is missing functions such as pow then your libraries are probably not in a standard place. Add the directory where they can be found to the line starting

```
# LPATH = -L
```

after the -L and remove the #.

If you get really stuck get in contact with Ian Turton at the School of Geography.

Step 3: Loading the datafiles

Once you are a registered SAR user at a registered SAR site, you can obtain the SAR datafiles in a variety of ways.

The easiest is to contact either CMU or Ian Turton to arrange how to get them. Or you can build the datafiles from the raw SAR files supplied by CMU. To do this place the compressed SAR files `indiv.sar.Z` and `hhld.sar.Z` in the USAR directory or link the files to this directory e.g.

```
prompt: ln -s /tmp/sarfiles/indiv.sar.Z /home/ian/work/usar
```

Then type `make data`. This will take some time since the USAR files are randomised during this process to save time later.

If you selected the compressed file option in the Makefile then you now need to type `compress *.in` (or modify the data lines in the Makefile if you are really short of space).

3 USAR Recipes

Recipe 1: A simple tabulation recipe for USAR

If you are an experienced USAR user go to recipe 2 or 3.

USAR is a screen based program which is operated by either the use of the arrow (cursor) keys of your terminal to select an option on the screen or by pressing the first letter of a command. The following examples are shown one screen to a page, with instructions about what to do under the picture of the screen underneath.

First start USAR. Exactly how to do this will vary from system to system, but if you or your system manager has placed USAR on your path then simply typing `usar` (remember UNIX is case sensitive typing USAR will not work) at your normal prompt should work. If you have any problems then talk to whoever installed USAR and see if they can tell you what to do next. You should see a welcome screen. Press any key to reach the main screen.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Data None
Variables
Make table

Filters

Search
Explore

Logfile: None
Print
Restore table

Quit

use arrow keys to choose command - return to select
or type first letter

This is the main USAR control screen. It allows you access to all of USAR's functions. The first thing to do is to select which dataset is to be used. Use the arrow keys to move the highlighted block around the screen until **Data** is highlighted. How your terminal highlights an area depends on what sort of terminal it is, but it may appear as reverse video or as a different colour. When **Data** is highlighted press return to select this option. Alternatively you can just press the first letter which in this case is **D**.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI. Problems to: ian@geog.leeds.ac.uk

Individual	sample size 100
Houses	
House/people	2% Individual data

use arrow keys to choose file - return to select
Changing the sample size will reduce the accuracy of your table
but shorten the time required to produce it

This screen is used to select which datafile is to be used for the table. The up and down arrow keys will cycle round the list of datafiles. At this point you can also change the size of the sample to be used. By default USAR will use all of the datafile. However on some systems this will be too slow and you may wish to trade off the accuracy of the table for an increase in speed. This is particularly useful if you want to generate lots of tables quickly and then select just a few for publication or more detailed examination. To change the sample size either move the cursor right to highlight **sample size** or press s. Then type in a number between 1 and 100. This becomes the new sample size. You can of course change the sample size again later if you wish. Move back to the left and highlight **Individual** and press return to select it. This returns you to the main screen. Now select **Variables** and press return, or press V to continue.

REGIONP	Quit	Rows
AGE	Make Table	
CESTSTAT	Filters	
CESTTYPE	Locate	
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		Cols
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

Region of SAR area

use arrow keys to choose variable - press rows or columns

reselect a variable to delete it

Select Maketable to build table or finish to return to menu

This screen is used to select the variables for a table. On the left you can see a list of OPCS abbreviated variable names with an explanation of the highlighted variable at the bottom of the column. On the right are two headings rows and cols which will indicate which variables you have chosen. In the center are four options; **Quit** which returns back to the main screen, **Make Table** which leads to the screen which controls making the table, **Filters** which is used to apply filters to the dataset as it is read and **Locate** which prompts for a variable name to be found.

A table is made up of rows and columns USAR requires a table to have at least one row and one column. However you may have more variables if you wish, the first variable you select will be the inner most (i.e. the closest to the figures in the table).

Use the arrow keys to move down the list of variables until **ECONPRIM** is highlighted, then press **r** to select this variable as a row variable. If you select the wrong variable then simply reselect it to delete it, i.e. if you press **r** when **ECONSEC** is highlighted by mistake press **r** again while it is highlighted and **USAR** will deselect this variable.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

NA	finish
Employee FT	Next group
Employee PT	Special
Self-emp with	Clear
Self-emp without	Total No
Govt scheme	Locate
Unemployed	
Student	
Perm Sick	
Retired	
Other Inactive	

ECONPRIM

Select class(es) for group 1

Special - to set up fixed size classes or a number of classes

To select all classes one per group just finish

This screen is used to group variable classes within a table. On the left is a list of the class names for this variable. On the right are some functions; **finish** accepts the grouping entered (or on class per group if none are entered); **Next Group** or pressing n finishes entering this group and starts the next (only if at least one class has been entered in the current group); **Special** allows some special operations to be performed on the classes such as a number of equal sized groups (see Recipe 6 for more details (35)); **Clear** clears the groups entered and starts again; **Total** turns totals on for this variable; and **Locate** which searches the list of classes for a specific name.

Use the arrow keys to highlight **Employee FT** and press return. A 1 (for group 1) will appear next to it and **Employee PT** will now be highlighted. Press return again. Now press

n to move to group 2 and select the two self employed classes, then press n to move to group 3 and select **Govt Scheme**.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

NA	finish
1 Employee FT	Next group
1 Employee PT	Special
2 Self-emp with	Clear
2 Self-emp without	Total No
3 Govt scheme	Locate
4 Unemployed	
5 Student	
6 Perm Sick	
6 Retired	
7 Other Inactive	

ECONPRIM

Select class(es) for group 7

Special - to set up fixed size classes or a number of classes

To select all classes one per group just finish

Continue to group the variable by placing **Unemployed** in group 4, **student** in group 5, **Perm Sick** and **Retired** in group 6 and **other inactive** in group 7. You should now see the screen above. Move to **finish** and press return to select this grouping.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

1 Employee FT-Employee PT	Finish
2 Self-emp with-Self-emp without	
3 Govt scheme	
4 Unemployed	
5 Student	
6 Perm Sick-Retired	
7 Other Inactive	

Select any of the labels to change by pressing return
Finish when done

This screen allows you to change the labels which are used to describe the groups you have just selected. By default USAR will use the first and last class in the group separated by a hyphen. Using the arrow keys to move up and down highlight **Employee FT-Employee PT** and press return. You can now enter a better label for this group such as **Employed**. Press return when you have finished. You can use the delete keys on your terminal to erase any mistakes you might make. Now change **Self-emp with -Self-emp without** to **self-employed**. Then select **Finish** and press return to select these changes.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP	Quit	Rows ECONPRIM
AGE	Make Table	
CESTSTAT	Filters	
CESTTYPE	Locate	
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		Cols
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

Ethnic group

use arrow keys to choose variable - press rows or columns

reselect a variable to delete it

Select Maketable to build table or finish to return to menu

You should now have returned to the variable selection screen. Move down and highlight **ETHGROUP** and press c to select this as a column variable. Move through the same process as for **ECONPRIM** but just select **finish** on each screen to select the defaults in each case.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP	Quit	Rows ECONPRIM
AGE	Make Table	
CESTSTAT	Filters	
CESTTYPE	Locate	
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		Cols ETHGROUP
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

Ethnic group
use arrow keys to choose variable - press rows or columns
reselect a variable to delete it
Select Maketable to build table or finish to return to menu

Having returned to the variable selection screen again, move to **Make Table** or press m.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Make table
Preview table
Optimise the table
Save the table
Flip table
Quit

Select an option

This menu allows you to set various options relating to the table, but for now we will just pass over them. Simply press M or select **Make table** to continue building the table.

Processing00%

USAR is now reading the datafile. This may take a little while depending on how busy your system is. If you find that it takes too long remember to set a smaller sampling fraction next time you build a table. USAR will also ask you if you wish to continue if it has taken more than 30 seconds to read the first 10% of the datafile. If you choose not to continue then USAR will display the table you created as if you had selected a sample size of 10%. The data is stored such that this is always a representative random sample of the SAR.

ETHGROUP	White	Black Caribb	Black Africa	Black Other	Indian
ECONPRIM					
Employed	400811	4036	1100	823	5325
self employe	55626	232	104	62	1234
Govt scheme	6455	134	68	47	147
Unemployed	46071	1041	465	279	1057
Student	40633	441	682	209	1356
Perm Sick-Re	202155	1072	133	97	1241
Other Inacti	101248	678	412	207	1675

use arrow keys to scroll table

Crown Copyright

Output to file, Q to return to menu, F to change format R to regroup table

Here is your first table. In fact it is the left hand side of your first table since the whole table is probably too wide to fit on your screen. Press the right arrow key to move across. You will hear a beep (or your screen may flash) when you reach the end.

ETHGROUP	Pakistani	Bangladeshi	Chinese	Other-Asian	Other
ECONPRIM					
Employed	1372	439	927	1492	1532
self employe	443	90	358	163	205
Govt scheme	76	40	25	51	64
Unemployed	819	293	144	274	386
Student	719	235	687	578	570
Perm Sick-Re	494	152	164	173	283
Other Inacti	1571	558	335	585	462

use arrow keys to scroll table At End Column Crown Copyright
 Output to file, Q to return to menu, F to change format R to regroup table

When you have finished moving backwards and forwards along the table you can press Q to return to the main screen and a further Q to quit USAR.

Congratulations, you have now discovered how to make a table using USAR! The rest of this manual builds on this, showing you how to make multi-dimensional tables, how to apply filters and how to explore the SAR.

Recipe 2: Adding a third variable to the table

The next step is to create a more complex table than in recipe 1. USAR allows tables to be constructed using more than two variables. This type of table is often described as multi-dimensional or n-dimensional where n is the number of variables used.

First lets make a 3-dimensional table. Adding more variables is just the same step again and again. Follow the instructions in recipe 1 until page 15 and then instead of pressing M for make table, select another variable and group it as shown on pages 8–15. USAR uses the convention that the first named variable is the closest to the table.

USAR allows you to have up to 10 variables as rows and columns, simply keep selecting them in the variable screen as above. But beware too many variables and your table will be mostly spaces, and will be impossible to interpret.

Recipe 3: Applying a filter to the dataset

A filter is a rule or set of rules that are applied to every record read from the datafile. If the record matches the rules it is sent to the table, if not then the next record is read.

Follow the instructions in recipe 1 until page 15 and then instead of pressing **M** for make table, select **Filters** or press **f**.

Filters can also be accessed from the main screen. The following screens apply to either route.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

	Finished	Filters
REGIONP		
AGE	Clear	
CESTSTAT	Locate	
CESTTYPE		
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

use arrow keys to choose variable - press return to select
Select finished to return to menu

This screen displays all the available variables for this data set. There are three commands; **Finished** which is selected when you've finished adding filters and returns to which ever screen you started on; **Clear** which allows you to clear all of the filters entered so far; **Locate** allows you to search for a variable in the list.

Use the arrow keys to move up and down the list of variables. When **REGIONP** is highlighted press return to select it.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI

Problems to: ian@geog.leeds.ac.uk

equals
not equal
greater than
less than

select type of filter

USAR uses this screen to request a selection rule. There are four choices; **equals** is the most common option; **not equal** the opposite of equals; and **greater than** and **less than**, are really only meaningful for ordered variables (e.g. age).

The arrow keys move the highlighted region up and down. Select **equals** and press return.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

	REGIONP EQ
North	
Yorks and Humb	
East Midlands	
East Anglia	
Inner London	
Outer London	
Rest of S.East	
South West	
West Midlands	
North West	
Wales	
Scotland	

choose the class for the filter

This screen displays the classes for the variable selected. If necessary the list can be scrolled up and down the screen. However this variable has less than a screen's worth of classes.

If we want to create a table that just includes data for Scotland, we move down to **Scotland** and press return.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP	Finished	Filters
AGE	Clear	REGIONP EQ Scotland
CESTSTAT	Locate	
CESTTYPE		
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

use arrow keys to choose variable - press return to select
Select finished to return to menu

You can now see your selection listed on the right hand side of the screen. If you have made a mistake you can either select **clear** to delete all the rules you have entered, or select the same variable, and class; USAR will delete the rule defined for them.

We can now set a second rule by selecting **SEX**. You may need to scroll down the screen to find this, and then setting the rule to **not equal** and then selecting **Female**. In fact this is equivalent to setting **SEX equals Male** but can be more useful in other cases for instance if you wanted to study England and Wales, filtering Not Scotland is easier then selecting all the regions in England and Wales.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP	Finished	Filters
AGE	Clear	REGIONP EQ Scotland
CESTSTAT	Locate	SEX NE Female
CESTTYPE		
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

use arrow keys to choose variable ~ press return to select
Select finished to return to menu

If you select two rules for the same variable, but with different classes then USAR creates an "or" join between the classes. For instance REGIONP = Scotland and REGIONP = Wales will select records from Scotland or Wales. If you enter two rules from different variables, as above REGIONP = Scotland and SEX = Male, then the records of males living in Scotland will be selected.

USAR allows an arbitrary number of filters to be entered, however the later ones may be displayed off the bottom of the screen. Also too many rules will lead to very low numbers of records matching.

Select finished or press f to return to the variable screen.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP	Quit	Rows ECONPRIM
AGE	Make Table	
CESTSTAT	Filters	
CESTTYPE	Locate	
COBIRTH	REGIONP EQ Scotland	
DISTMOVE	SEX NE Female	
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		Cols ETHGROUP
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

use arrow keys to choose variable - press rows or columns
reselect a variable to delete it
Select Maketable to build table or finish to return to menu

You will notice that the filters you have selected are shown in the center of the screen. Select make table to make the table.

ETHGROUP	White	Black Caribb	Black Africa	Black Other	Indian
ECONPRIM					
Employed	20262	4	11	4	24
Self Employe	3031	0	0	1	20
Govt scheme	488	0	1	0	0
Unemployed	3436	2	2	1	7
Student	1951	1	18	2	16
Perm Sick-Re	8489	4	1	2	6
Other Inacti	342	0	0	0	0

use arrow keys to scroll table

Crown Copyright

Output to file, Q to return to menu, F to change format R to regroup table

This is the table of Ethnic group against Economic Primary activity of Scottish men (in fact, Scots who are not female). Note that the whole of the second column is highlighted since USAR has calculated that these values are not significantly different from 0 at a 5% confidence level. Several other values in the table are similarly highlighted..

Recipe 4: Formatting and printing tables

MARSTAT	Single		Married		Remarried
SEX	Male	Female	Male	Female	Male
ETHGROUP					
White	227102	202613	215367	218639	30023
Black Caribb	2859	3092	1321	1291	189
Black Africa	1312	1259	670	619	57
Black Other	1388	1456	206	220	35
Indian	4181	3789	4003	3908	183
Pakistani	2883	2496	1856	1805	113
Bangladeshi	1074	910	614	570	33
Chinese	962	859	665	710	26
Other-Asian	1084	987	907	1059	43
Other	1983	1882	834	673	77

use arrow keys to scroll table

Crown Copyright

Output to file, Q to return to menu, F to change format R to regroup table

USAR provides a variety of different formats that can be used to display and print tables. These can be controlled by pressing F when a table is displayed or from within the **Print** option on the main screen

In this recipe we will examine some of these formats. Follow the steps in recipe 1 until you have a table displayed on screen, then press F.

Output type Ascii
Format Counts
Errors Flag

select O - to change output type, F - to change format of table,
E - to change format of errors
q or return to finish

This screen is used to select one of three output formats of USAR. The first group **Output type** takes one of two values **Ascii** and **Latex**. This determines if the output file written by USAR is in plain ascii, the sort of characters you can send to a line printer, or **LaTeX** which is a typesetting program used by the unix community. If you do not know what **LaTeX** is then do not worry about this option. If you are a **LaTeX** user then USAR writes a complete tabular environment that you can import directly into your document; however the begin and end table commands are up to you. You can also import ascii text into your spreadsheet, the separator are TAB characters.

The next group is **Format**. This determines how USAR displays the numbers in the table on the screen and when you output the table. The choices are **Counts** which gives the raw counts, **% columns** and **% rows** which display column and row percentages respectively.

The final group is **Errors**. you can choose from one of **Flag** which highlights any cell in the table which is not significantly different from 0 at a 5% confidence limit, when sampling errors are taken into account; **Show** displays the standard error for each cell in brackets after the count; and **Ignore** which suppresses all consideration of sampling errors. To write the output of a table to disk press **O** when looking at the table on the screen. USAR will prompt you for a file name. This may be a simple name, in which case the current directory will be used, a name beginning with a `~` which is expanded to be your home directory or a full path name starting `/` which will use that explicit path (e.g. `/tmp/test.tab`). If the file exists then USAR will ask you if you wish to replace the old file, append this output to the existing file or cancel the action. USAR will always use the current formats, set above, when saving the table to disk.

Recipe 5: Saving and loading tables

USAR allows you to save and then reload tables in a format that allows you to add filters or more variables to your table once it is reloaded. To test this create a table as shown in recipe 1 (pages 6–16).

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Make table
Preview table
Optimise the table
Save the table
Flip table
Quit

Select an option

To save your table select **save the table** or press **s**. You will now be prompted for a file name. This may be a simple name, in which case the current directory will be used, a name beginning with a **~** which is expanded to be your home directory or a full path name starting **/** which will use that explicit path (e.g. **/tmp/test.tab**). If the file does not exist then USAR will ask you to confirm that you wish to create it. If it already exists then the new table will be added to the end of the file. You will then be prompted for a label for the table. This is any short description of the table you are saving. *Do not use things like "table 1" if you want to be able to find them later.*

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Data None
Variables
Make table

Filters

Search
Explore

Logfile: None
Print
Restore table

Quit

use arrow keys to choose command - return to select
or type first letter

To reload a saved table, select **Restore table** or press **r**. USAR will prompt you for the name of a file. The same valid names as used above for saving are permitted. You are then presented with a list of tables stored within the file, use the arrow keys to move up and down this list and press return to select the table to load. You can now continue either by making the table in the usual way, or continue to add variables or filters to the table before making it.

Recipe 6: Other ways of grouping variables

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

0	finish
1	Next group
2	Special
3	Clear
4	Total No
5	Locate
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

AGE

Select class(es) for group 1

Special - to set up fixed size classes or a number of classes

To select all classes one per group just finish

Follow the instructions in recipe 1, until you reach the variable selection screen, select **AGE**.
On the screen above select **Special** or press **S**.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Fixed classes
Number of classes

choose either Fixed class intervals or a number of classes

USAR now offers you the choice of **Fixed Classes** and **Number of Classes**. Fixed classes allows you to specify the size of the groups you wish to apply to the variable e.g. 5 year age groups. Number of classes allows you to say how many groups you want e.g. 4 quartiles. Select **Fixed Classes** now, discussion on **Number of classes** is on page 39.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

50	First class 18
51	Last class 65
52	Step size 10
53	Return
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	

use arrow keys to move up/down class list
press F to set the first class, L to set the last class
and S to set the Size of the groups. R to finish

This screen allows you to specify the range of the groups, and the size of the groups. Use the arrow keys to move up and down the list of classes. To set the start of the range press F and the currently highlighted class will be used as the start. Move further down the list and press L to set the end of the range. To set the size of the groups either move across to Step or press S and then enter the size. Select Return when you have finished.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

1 18-27
2 28-37
3 38-47
4 48-57
5 58-65

Finish

Select any of the labels to change by pressing return
Finish when done

You can now continue as usual in the grouping of variables and change any of the labels for the groups as required.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

51 First class 16
52 Last class 65
53 Number of groups 10
54 Return
55
56
57
58
59
60
61
62
63
64
65
66

use arrow keys to move up/down class list
press F to set the first class, L to set the last class
and N to set the number of groups R to finish

If you select **Number of classes** you are taken to the screen above which allows you to select the beginning and end of the groups and how many groups are to be created. Use the arrow keys to move up and down the list of classes. To set the start of the range press **F** and the currently highlighted class will be used as the start. Move further down the list and press **L** to set the end of the range. To set the number of the groups either move across to **Number of groups** or press **N** and then type in the number of groups required. Press **R** or select **Return** when you have finished.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

- 1 16-20
- 2 21-25
- 3 26-30
- 4 31-35
- 5 36-40
- 6 41-45
- 7 46-50
- 8 51-55
- 9 56-60
- 10 61-65

Finish

Select any of the labels to change by pressing return
Finish when done

You can now continue as usual in the grouping of variables and change any of the labels for the groups as required. However you may notice that there are not the same number of groups as you specified earlier, this is a result of the fact that there may be no number of classes that exactly divides your range into the number of groups asked for. It is also possible for the last group to be smaller than the others for the same reason.

Recipe 7: Table optimisation

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Make table
Preview table
Optimise the table
Save the table
Flip table
Quit

Select an option

After selecting **Make Table** either on the variable selection screen or the main screen, you are presented with the table above. It is possible to use USAR to optimise the grouping of your table.

However certain conditions are required for this to be effective; the variable or variables to be optimised must be grouped so that there are multiple classes in at least one of the groups. This is because USAR keeps the number of groups constant and moves classes from one group to the next.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Entropy
Chi Squared
Small Numbers

Select the type of optimisation required and press return

A number of objective functions are provided in USAR, and it is hoped that users will submit more over time. At present there is a choice between minimisation of entropy (Shannon 1948), maximisation of χ^2 and the minimisation of small values.

Use the arrow keys to move to the function required and press return.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

AGE
LTILL

Finish
All

Select variable to be optimised and press return

Now select the variable or variables that you want to regroup, remember that they must have more than one class in at least one group. If you want to select all the the variables in the table select **All** or press **A**. Select **Finish** to return to the previous screen where selecting **Make Table** will produce your table. You will be asked to label the new groups before the table is displayed. This is exactly the same process as shown on page 13.

Recipe 8: Searching the SAR

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Data None
Variables
Make table

Filters

Search
Explore

Quit

Logfile: None
Print
Restore table

use arrow keys to choose command - return to select
or type first letter

The SAR is a very large dataset and contains many interrelated variables. To address this, USAR provides a method for searching the SAR to discover these relationships. To use this function you must first select a dataset (**D**, see pages 6-7). In this recipe we will use the individual data, then select **Search** or press **S**.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP (North) = 0 (0.00%)	Search
REGIONP (Yorks and Humb) = 0 (0.00%)	Quit
REGIONP (East Midlands) = 0 (0.00%)	Output
REGIONP (East Anglia) = 0 (0.00%)	Locate
REGIONP (Inner London) = 0 (0.00%)	
REGIONP (Outer London) = 0 (0.00%)	
REGIONP (Rest of S.East) = 0 (0.00%)	
REGIONP (South West) = 0 (0.00%)	
REGIONP (West Midlands) = 0 (0.00%)	
REGIONP (North West) = 0 (0.00%)	
REGIONP (Wales) = 0 (0.00%)	
REGIONP (Scotland) = 0 (0.00%)	
AGE (0) = 0 (0.00%)	
AGE (1) = 0 (0.00%)	
AGE (2) = 0 (0.00%)	
AGE (3) = 0 (0.00%)	

use arrow keys to select variable to be limited, reselect to remove

This screen shows you all the available combinations of variables and classes in the selected dataset. The arrow keys allow you to move up and down the list of combinations. You can also use the **Locate** option to search for a variable in the list. Pressing return selects a variable class pair to be searched for. If you make a mistake simply pressing return again will unselect that class. Note that the selections you have made are shown at the foot of the screen.

Select **REGIONP (Yorks and Humb)** and then move the cursor right to **Search** and press return or just press **s**.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Processing ...00%

use arrow keys to select variable to be limited, reselect to remove
REGIONP=Yorks and Humb

USAR will now process the data file.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP (Yorks and Humb) = 98304 (100.00%)	Search
GAELLANG (NA) = 98304 (100.00%)	Quit
WELSHLAN (NA) = 98304 (100.00%)	Output
LOWFLOOR (NA) = 98304 (100.00%)	Locate
CESTTYPE (NA) = 96025 (97.68%)	
CESTSTAT (NA) = 96025 (97.68%)	
BATH (Exclusive) = 95606 (97.26%)	
URVISIT (NA) = 95482 (97.13%)	
INSIDEWC (Exclusive) = 95426 (97.07%)	
ETHGROUP (White) = 93891 (95.51%)	
ECONSEC (NA) = 93464 (95.08%)	
COBIRTH (England) = 91717 (93.30%)	
RESIDSTA (Present) = 91591 (93.17%)	
QUALEVEL (NA) = 89885 (91.44%)	
QUALNUM (0) = 89539 (91.08%)	
QUALSUB (0) = 89539 (91.08%)	
use arrow keys to select variable to be limited, reselect to remove	
REGIONP=Yorks and Humb	

98304 matches out of 1116181 (8.81%)

The new list of variables displayed shows how many records matched the selection you made earlier. In this case 8.81% of the total file (look at the foot of the screen). We can now use the arrow keys to move up and down this sorted list to see if any interesting items have been discovered. If you interested in a particular variable select **Locate** to search for it. New variables can be added to the list of selections and if necessary the original selection can be deleted and the search restarted.

Recipe 9: Exploring the SAR

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

Data None
Variables
Make table

Filters

Search
Explore

Quit

Logfile: None
Print
Restore table

use arrow keys to choose command - return to select
or type first letter

The exploration function of USAR is used to carry out "fuzzy" searches of the SAR. This enables you to study how one variable is affected by others when you can not be sure which, of several variables, are responsible. Thus you can specify a group of variables and classes within those variables and a required number of matches (i.e. 4 out of 7). USAR will then return a list of combinations of those variable groups that match at least 4 of the specified variables and a count of the number of records.

To use this function you must first select a dataset (**D**, see pages 6-7). In this recipe we will use the individual data, then select **Explore** or press **E**. This recipe will investigate the housing tenure of lone parent families with respect to other related variables.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP		Filters
AGE		Explore
CESTSTAT		Required matches 1
CESTTYPE		Quit
COBIRTH	Matches	Locate
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

Select variable using arrow keys and press return

On this screen you can see a list of the variables available in the selected dataset on the left of the screen. On the right are four commands; **Filters** which is used to apply filters to the selected data as described in recipe 3 (pages 21-28); **Explore** which starts the exploration process; **Required matches** which is used to specify the number of matches; **quit** which returns to the main screen; and **Locate** which is used to search for a named variable in the list.

Select **TENURE** and press return.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

NA	Finish
* 00 outright	All
* 00 Buying	
* Rent priv Fur	
* rent priv unfur	
* Rent job/bus	
* HA rent	
* rent LA/NT E+W	
* rent LA S	
* rent NT S	
* rent Scot H	

use arrow keys to move up/down class list
Press return to select a class
move to finish to return

This screen is used to define which classes of a variable are required in the exploration. Pressing return selects the highlighted class and a * appears next to it. Pressing return again unselects the class. If you select All all the classes are highlighted. Thus if you want all but one of the classes displayed pressing All and then pressing return on the unrequired class will be the quickest way.

Now select Finish to return to the previous screen.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

WELSHLAN	Filters
WORKPLCE	Explore
BATH	Required matches 4
CENHEAT	Quit
INSIDEWC	Matches. Locate
CARS	TENURE 00 outright
LOWFLOOR	TENURE 00 Buying
HHSPTYPE	TENURE Rent priv Fur
DENSITY	TENURE rent priv unfur
TENURE	TENURE Rent job/bus
RESIDNTS	TENURE HA rent
DEPCHILD	TENURE rent LA/NT E+W
LTILLHH	TENURE rent LA S
PENSINHH	TENURE rent NT S
EARNERS	TENURE rent Scot H
ECONPOSFHP	SEX Male

Select variable using arrow keys and press return

Repeat the procedure to include **SEX**, **DEPCHILD** and **EARNERS** in the list of matches required.

Now select **Required Matches** or press r and enter 4 as the number of required matches.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

REGIONP	Finished	Filters
AGE	Clear	RELAT = Head
CESTSTAT	Locate	FAMTYPE = lpt Dep chil
CESTTYPE		FAMTYPE = lpt n/dep chil
COBIRTH		
DISTMOVE		
DISTWORK		
ECONPRIM		
ECONSEC		
ETHGROUP		
FAMTYPE		
GAELLANG		
HOURS		
INDUST		
LTILL		
MARSTAT		

use arrow keys to choose variable - press return to select
Select finished to return to menu

Select filters on the exploration screen and then follow the instructions on pages 21-28 to select **FAMTYPE** equals **Lpt Dep Child** and **Lpt n/dep child** and **RELAT** equals **Head**. This restricts the search to the heads of family in lone parent families.

Select **finish** to move back to the exploration screen.

```

57 57 57 57 57 57 57 57 57 57 57 38 38 60 60 60 63 63 63 Finished
 1  2  3  4  5  6  7  8  9 10 0  1  0  1  2  0  1  2
0  0  0  0  0  0  1  0  0  0  0  1  0  1  0  1  0  0  6266
0  1  0  0  0  0  0  0  0  0  0  1  0  1  0  0  1  0  3081
0  0  0  0  0  0  1  0  0  0  0  1  0  1  0  0  1  0  2368
1  0  0  0  0  0  0  0  0  0  0  1  1  0  0  0  1  0  1799
0  1  0  0  0  0  0  0  0  0  0  1  0  1  0  1  0  0  1615
0  1  0  0  0  0  0  0  0  0  0  1  1  0  0  0  0  1  1603
0  0  0  0  0  0  1  0  0  0  0  1  1  0  0  0  1  0  1463
0  1  0  0  0  0  0  0  0  0  0  1  1  0  0  0  1  0  1113
1  0  0  0  0  0  0  0  0  0  0  1  1  0  0  0  0  1  1111
0  0  0  0  0  0  1  0  0  0  0  1  1  0  0  0  0  1  959
0  0  0  0  0  0  0  1  0  0  0  1  0  1  0  1  0  0  921
0  0  0  0  0  1  0  0  0  0  0  1  0  1  0  1  0  0  920
0  0  0  0  0  0  1  0  0  0  0  1  1  0  0  1  0  0  881
0  1  0  0  0  0  0  0  0  0  1  0  1  0  0  0  0  1  749
Total number of sucessful matches = 37697 out of 1116181 (3.38%)
Select pattern of interest and press return for a description
Select finish to return

```

Selecting **Explore** or pressing **e** starts the exploration process. USAR displays a screen to show how much of the task has been completed. This is followed by the answer screen shown above, which is a sorted list of groups of matching patterns. However it can be hard to "decode". The top line shows the variable numbers, and the second line shows the classes. The following lines are the patterns: a 1 means it is in the pattern and a 0 is not in the pattern. However, pressing return displays a more understandable description of the pattern. When the top row of answers is highlighted, press return.

USAR 1.0 - (c) Ian Turton 1993, School of Geography, University of Leeds
ESRC/ JISC/ DENI Problems to: ian@geog.leeds.ac.uk

TENURE rent LA/NT E+W
SEX Female
DEPCHILD 1+
EARNERS 0

6266 matches from 37697
Press return to finish

This shows that the pattern with the largest number of matches is for female lone parents living in local authority housing with one or more dependent children and no earners in the household.

Press return to redisplay the results screen. When you have finished examining the results select **Finish** to return to the exploration main screen. **Quit** will return you to the main USAR screen.

4 Advanced Details

4.1 Log Files

USAR provides a comprehensive logging facility and the ability to run these log files as batch files. Users may also edit the log file to create complete table construction jobs. The log file is

set from the main screen, by selecting **Logfile** and entering the name of a file when prompted. If the file exists you will be asked if you wish to replace the file, append this log to the end of the file or cancel the operation. Once a logfile is started all operations related to the creation of tables are logged.

To replay the logged session, type `usar logfile`; where logfile is the name of your logfile. USAR then reads the logfile line by line following your previous instructions. You can continue from where you finished the logfile.

4.2 Batch jobs

Batch jobs are created using a superset of the logfile commands. The simplest batch jobs are in fact just logfiles. However for more flexibility a logfile can be created including any of the following commands:

data type where type is one of Individual, Houses, House/people to set the dataset to be used.

sample n where n is a number from 1 to 100 to set the sample size.

Row n VAR

Col n VAR where n is the number of the row or column, 1 is the innermost, VAR is the variable name required. There then follows a list of variable classes, in the form low value high value label. In general it is easiest to use USAR to create these lists and to cut and paste them into a batch file. The list is terminated with a line containing End Row or End Col.

delete Row n

delete Col n where n is the row or column number as described above. These commands are used to delete a row or column. Any variables that are higher than the deleted variable are moved down.

Total Row n

Total Col n where n is the row or column number as described above. These commands are used to turn totaling on for the row or column indicated.

Filters: This line is followed by a list of filters to be applied to the table. These lines have a format of VARIABLE type class, where type is one of EQ, NE, LT, GT meaning equals, not equals, less than and greater than respectively. Again it is easiest to use USAR to create lists of filters. The list of filters is terminated by a line End Filters. An empty list of filters will clear any already set.

Make table The first time this is encountered in a batch file USAR will move to the table preparation screen (16). The second time will cause it to make the table and display it on the screen. However if an output file is specified (see below) then USAR will not display the table to the screen, and the Make table command is ignored.

output filename USAR will create the table described in the batch file and output it to the named file. If the file exists then the output will be appended to the end of the existing file.

format Output n where n is 0 or 1; 0 means Ascii and 1 means \LaTeX ; described in more detail on page 30.

format Error n where n is 0, 1 or 2; 0 is ignore errors, 1 is flag small values and 2 is show numerical values of errors, see page 30 for more details.

format Format n where n is 0, 1 or 2; 0 is raw counts, 1 is show column percentages and 2 is show row percentages.

References

- Shannon, C. (1948). A mathematical theory of communication, *Bell System Technical Journal* 27: 379–423.
- Turton, I. and Openshaw, S. (1994). Putting the 1991 Census Sample of Anonymised Records on your Unix workstation, *Environment and Planning A*.

Index

- arrow key, 5–8, 10, 12, 17, 21, 22, 33, 44, 46
- Ascii, 29
- batch files, 54
- census, 1
- CMU, 1, 2, 4, 5
- column, 8, 13, 19, 27, 29, 54, 55
- data
 - household, 7
 - individual, 7
- datafile, 7
- Explore, 47, 48, 52
- explore, 18, 47
- Filter, 20, 48, 54
- filter, 8, 18, 20, 21, 24–26, 31, 33, 48, 51, 54
 - rules, 22
- Latex, 29
- logging, 53
- OPCS, 8
- optimisation, 40
- output, 30
- patterns, 52
- percentages, 29
- raw counts, 29
- row, 8, 19, 29, 52, 54, 55
- sampling, 7
- sampling errors, 30
- Search, 43, 44
- search, 1, 43, 44, 46, 47, 51
- table, 4, 7, 8, 10, 15–20, 23, 26–33, 53–55
 - multi-dimensional, 19
- optimisation, 40
- output, 30
- reloading, 33
- saving, 31
- structure, 8
- tables
 - format, 28
- USAR, 1–9, 12, 16, 18, 19, 22, 24, 25, 27–33, 43, 45, 47, 52–55
- variable, 8–11, 13, 14, 19, 21–25, 31, 33, 43, 44, 46–49, 52, 54
 - deleting, 9
 - grouping, 10, 34
 - labels, 12
 - ordering, 8

Produced By
School of Geography
University of Leeds
Leeds LS2 9JT
From Whom Copies May Be Ordered

April 1994