Lists and coded lists[1]

Producing lists of variables and observations seems trivial and not really worthwhile talking about. However lists are very important as a first step in analysis before you start to summarize, or later, when summaries seem strange, you might want to come back to the individual values. Furthermore you will see that well presented and analyzed lists are valuable tools in multivariate analysis and comparision.

There are basically two forms of lists: numerical lists (showing data values) and coded lists (symbols replacing the numbers). The EDA software offers two commands: LIST and SHOW. LIST offers many options to produce a wide variety of different lists; SHOW is useful if you want to display data values selectively (based on a condition, e.g. outliers only).

1 Numerical lists

By default observations are shown as they are stored (sequentially) in the data matrix.

```
>LIST 1-3

 variable  listing
country    g#  GNPAgr  GNPInd GNPServ

  1 AFGH   1     65     20     15
  2 AFRI   2      5     43     52
  3 ALBA   3     33     52     15
  4 ALGE   2     14     49     37
  5 D      3      3     45     52
  6 AND    3      6     37     57
  7 ANGO   2     14     36     50
  8 ANBA   4      5     22     73
  9 ANNE   4      1     26     73
 10 ARAB   1      7     45     48
 11 ARG    5     15     35     50
```

Besides the three data columns, the table shows three additional columns: the first contains the sequential number of each case, the second the case identifier (CASID).[2] and the third column (labelled #g) contains group numbers, i.e. in this case refer to the continent, e.g. the country “AFGH” is in group number 1, i.e. Afghanistan is in Asia. Note that this column is only present if groups are defined.[3]

```
>LIST 1-10 START=URUG END=ZIMB
```

will list only countries ‘URUG’ through ‘ZIMB’. [4]The START=cas# and END=cas# are used to limit the display to a specified range.[5] cas# can be either a case id (case ids are case sensitive!) or the sequential number. A command like LIST 1-10 END=20 would list the first 20 observations.[6]

```
>LIST 1-10 SORTED
```

shows variables 1 to 10 sorted on the first variable of the list. If you want to use a different variable as sort key use the KEY=v option, where v is a variable reference.[7]

---

1. E. Horber, 13.12.98 : LIST.mss

2. All work areas have case identifiers. By default EDA uses the numerical ids (sequence) when a WA is initially created. Note that in this case the id numbers only agree with the first column if the order of the observations has not been changed and no observations have been removed. Note also that these casids are alphanumeric: in commands referring to these case ids, you will have to quote them.

3. Case groups are defined as a GVAR (=grouping variable). We will see later how GVARS are defined and manipulated.

4. We should have written START='URUG', i.e. using single quotes to show the casids. The quotes can be omitted if the casid does not start with a numerical character nor contain spaces or other special characters.

5. START defaults to the first case and END to the last.

6. Note that if the case ids are numbers, a difficulty arises, i.e. there are situations where you need to distinguish between the 20th case and a case named ‘20’. END=20 refers always to the 20th case, END='20' refers to a case named ‘20’. Make sure to understand this when cases have numerical names; you should not be confused by the fact that in many situations the 20th case is named ‘20’; this will not always be the case.

7. You may use the ASCENDING or DESCENDING option to control sort order and to override the default sort order as set by the SET ORDER command (set to ASCENDING initially).

A number of options control the display format. DECIMALS=ndec lets you specify the number of decimals[8] You may also ask to show only INTEGER values (INTEGER means truncation) or ROUNDED values. In both cases zeroes are shown as blanks, unless you add a ZERO option. Note that setting the number of decimals to 0 is the same as using the INTEGER option.

Blanking lets you blank out all observations corresponding to some criterion, i.e. values not corresponding to the criterion are shown as usual (as numbers), all others appear as blanks.

```
>LIST 4-5 BLANK IF>10
```

All values greater than 10 (on all variables) appear as blanks, i.e. the criterion is applied to every variable.

```
>LIST 20-22 BLANK IF<10
```

case		GNPAgr	GNPInd	GNPSrv
1	ALBA	33	52	15
2	D		45	52
3	AND		37	57
4	A		37	60
5	B		30	68
6	BULG	13	68	19
7	CHYP		27	66
8	DK		29	66
9	E		39	56
10	FI		36	58
11	F		27	67
12	GREC	16	29	55
13	HNGR	20	45	35
14	IRLA	11	37	52
15	ISLA	23	26	51

```
>LIST 20-22 BLANK IF>50
```

case		GNPAgr	GNPInd	GNPServ
1	ALBA	33		15
2	D	3	45	
3	AND	6	37	
4	A	3	37	
5	B	2	30	
6	BULG	13		19
7	CHYP	7	27	
8	DK	5	29	
9	E	5	39	
10	FI	6	36	
11	F	6	27	
12	GREC	16	29	
13	HNGR	20	45	35
14	IRLA	11	37	
15	ISLA	23	26	

Besides the coded form, explained below, the LIST command offers a number of other options, the most important ones are:

- CASE=cas#: List the values for a single case.
- STREAM: List variables in stream form (unformatted list of all observations for the variables of the variable list.
- ORDER: You cannot only sort the list by observations to some variable, but you you can also change the order of the variables, specifying a specific observation as sorting criterion.[9]

2 The SHOW command

The SHOW command lists observations selectively according to some criterion.

A first category of options is related to the location of an observation within the distribution of a variable, displaying for instance the FAR out values for a variable:

```
>SHOW 22 FAR

  30 countries
Showing :GNPServ(22) %GNP for Services
Median= 59.0 LHinge= 51.0 UHinge= 63.0 Step=  12.0
The 5 countries shown are Far-Out in GNPServ(22)
ROUM ALBA BULG TCHE POLO
```

8. By default the number of decimals depends on the setting of the SET DECIMALS=ndec switch; this value is saved with each file, i.e. depending upon the current work area the number of decimals varies.

9. Like all analysis commands, LIST takes the current variable list, if no variable list is present on the command line, i.e. you might use other commands defining variable lists, namely options of the DESCRIBE/DS/VARS family of commands let you build variable lists using searching, statistical and other criteria; this list is then taken up by the LIST command if you do not specify a variable list with it. See the examples at the end of this handout.

Shows all far-out values of variable 1. If you specify serveral variables on the list, all far out values of the first variable are shown together with their values on the other variables.

```
  >SHOW 22 20 21 FAR

  30 countries
Showing :GNPServ(22) %GNP for Services
Median=       59. LHinge=       51. UHinge=       63. Step=       12.
The 5 countries shown are Far-Out in GNPServ(22)
countrie GNPServ  GNPAgr  GNPInd

 25 ROUM     14     16     70
  1 ALBA     15     33     52
  6 BULG     19     13     68
 29 TCHE     21      8     71
 23 POLO     24     14     62
```

Other options let you see OUT values, EXTREME values (OUT and FAR-out), ADJACENT values or specify a condition, like in

```
>SHOW 1-8 IF>10
```

shows all observation, where variable 1 exceeds 10.[10]

Sometimes you need to see what observations show up on top of a variable (largest values, “best”, “highest” etc) or to see the variables with the highest values for particular cases.

```
>SHOW 1-3 TOPCASES

  30 countries
Showing top  14 countries
GNPAgr  :ALBA ISLA HNGR TURQ ROUM GREC POLO BULG IRLA TCHE PORT CHYP F    FI
GNPInd  :TCHE ROUM BULG POLO ALBA D    HNGR E    PORT UK   AND  IRLA A    CH
GNPServ :MONA B    F    DK   CHYP NL   MALT S    I    N    LIE  UK   LUX  CH
```

‘ALBA’ has the highest value for %GNP produced in agricultre, followed by ICELAND. Monaco has the highest value GNP produced in the services. Only the top 14 countries are shown.[11]

```
>SHOW 1-3 TOPRANKS

  30 countries
Showing top (largest)  14 countries on 1st variable; rankings for the others.
GNPAgr  : ALBA ISLA HNGR TURQ ROUM GREC POLO BULG IRLA TCHE PORT CHYP F    FI
GNPInd  :   5   29    7   16    2   25    4    3   12    1    9   28   27   17
GNPServ :  29   23   25   24   30   19   26   28   21   27   20    5    3   16
```

Instead of showing top observations for all variables, you might want to see the top countries on a particular variable (here GNPAgr) and check on the other variables where these countries can be found: The display then shows ranks. For instance Hungary (HNGR) is third on GNPAgr, 7th on GNPInd and 25th on GNPServ.

```
>SHOW 1-3 TOPVARIABLES
```

10. An alternative to this is the general selection mechanism available with the EDA program; it will let you also specify more complex conditions.

11. Options let you show more or less.

```
  30 countries
Showing top   3 variables
ALBA:GNPInd   GNPAgr   GNPServ
D   :GNPServ  GNPInd   GNPAgr
AND :GNPServ  GNPInd   GNPAgr
A   :GNPServ  GNPInd   GNPAgr
B   :GNPServ  GNPInd   GNPAgr
BULG:GNPInd   GNPServ  GNPAgr
CHYP:GNPServ  GNPInd   GNPAgr
DK  :GNPServ  GNPInd   GNPAgr
 (....)
```

If we examine Bulgaria, we see that the most important sector is Industry, followed by the services and agriculture.

Observation oriented analysis often asks questions like “What are the countries closest/smilar to the UK?”, “... the most different countries?”. A number of options help you answer this kind of questions.

```
>SHOW 1-3 CLOSETO=UK

  30 countries
Show countries within a distance of 0.50*midspreads from UK
GNPAgr  :[MONA ]B    LIE  D    LUX  N    A    S    I    CH   NL   MALT E
GNPInd  :[A    AND  IRLA ]PORT CH   TURQ LUX  FI   E    S    LIE  N
GNPServ :LIE  LUX  I    S    N    NL   A    MALT CH   DK   CHYP FI   AND  F
```

The distance criterion used is half a midspread. There are options to used other units (e.g. a full midspread) or standard-deviation instead of midspread. The casids enclosed in [] are cases with exactly the same value as the UK.

This example only shows the obervations closest to the UK, without telling you whether the value is smaller or greater than the value for the UK. This information may be obtained with the DETAILS option, as shown in the next example.

```
>SHOW 1-3 CLOSETO=UK DETAILS

  30 countries
Show countries within a distance of 0.50*midspreads from UK
Legend: (-) below (+) above (=) equal (plus/minus fuzz)
GNPAgr  :(+)B    LIE  D    LUX  N    A    S    I    CH   NL   MALT E    DK
         (=)MONA
GNPInd  :(+)PORT E
         (=)IRLA A    AND
         (-)CH   FI   TURQ LUX  S    N    LIE
GNPServ :(+)N    S    LIE  I    MALT NL   CHYP DK   F
         (-)LUX  CH   A    FI   AND

>SHOW 1-4 AWAYFROM=UK

  30 countries
Show countries outside a distance of 2.00*midspreads from UK
GNPAgr  :ALBA ISLA
GNPInd  :TCHE ROUM BULG POLO MONA ALBA
GNPServ :ROUM ALBA BULG TCHE POLO HNGR
```

This table shows you the most different countries, using a criterion of 2 midspreads distant from the UK. Note that on this list the first country on each row is the most different country, then the second and so on.

Finally there is a SPLIT option, splitting the values of a single variable into two columns according to a criterion (far, out, condition specified by IF), as in

```
>SHOW 1 SPLIT IF>10
```

```
   33 countries
 Showing :InfMor(1) Infant Mortality
 shown  15 countries are greater than 10.00
 Split:left column contains selected cases
 ALBN   3    43.0
 AND    3           10.0
 A      3    10.3
 B      3            9.7
 BLGR   3    14.5
 CYPR   3    12.0
 CZCH   3    13.9
 DK     3            8.4
 FNLD   3            5.8
 F      3            8.0
<more>
```

3 Coded lists

In many situations it is preferable to replace numbers by some well chosen symbol, revealing structural aspects of our data table. Coded lists and other coded displays are frequently used in exploration and many commands have options to produce different forms of coded displays. Note that this text stresses coded lists (using examples of the LIST command), but other coded displays (maps, histrograms, plots) use the same framework.

The following forms of coding are used in EDA:

- *Distributional coding:* The symbols shown reflect the position of each observation within the distribution of the variable, i.e. symbols are used to show whether a case is a far out, out, adjacent or in value. Normally different symbols are used for values below and above the median, e.g. we distinguish low and high out values.

- *Bin coding;* The cases are grouped into a number of bins according to some criterion, and for each bin a different symbol is used. Several criteria are possible: Each bin contains an equal number of observations (Fractile coding); each bin corresponds to an interval of equal width (interval coding) or the bins are defined by the user (indicates the bin boundaries).

- *Reference coding* Each symbol reflects the position of a case with respect to some criterion. A criterion often used is the median; then values below the median are marked differently from cases above the median. Instead of only marking low/high positions it is also possible to indicate the distance of a case from a reference value, by either using different symbols or by using more than one symbol, e.g. a single plus sign for a case close to the median, and 2, 3 etc plus signs for cases farther away.

- *Marking:* All cases corresponding to some criterion are marked with a special symbol, all other cases are left unmarked

In the following section we shall examine the various forms of coding in some detail. Note again that the examples will mainly use the LIST command as an illustration; several other commands work essentially the same way.

3.1 Distributional coding

The symbols shown reflect the position of each observation within the distribution of the variable, i.e. symbols are used to show whether a case is a far out, out, adjacent or in value. Normally different symbols are used for values below and above the median, e.g. we distinguish low and high out values.

```
    >LIST 1-4 DISTCODE
   26 cases
Distributional coding (full); Symbols:@#+*-=&
         ZBLUSONGZFSBBSAASGATTVVNGJ
         HEURZWWLGROSLHRIGRGGIDSEEU
---------------------------------
ICult    |+***----***+**--****++*++*
XCult    |****-+***+*****+*+-*+-&--+
Form     |+****---**++**---***+***++
Sucre    |-****+*-*#-=-******+-++**+
```

The following symbols are used: '@' for a high far out value, '#' for a high out value, '+' for high adjacent, a star for an in value, and '-','=','&' for low adjacent, out and far out values.

Symbols may be changed using the "codes" option on the command line [12]

```
>LIST 1-4 DISTCODE "AB   FG"
```

will use A and B for high far-out and out values, F and G for low far-out and out-values and blank for others.

The DISTCODE option has an additional option SIMPLE, i.e. we do not want to distinguish between 'high' and 'low', i.e. the low and high far-out, out and adjacent values will have the same symbol. Note that it is also possible to use SET GRAPH DISTCODE SIMPLE to produce the same effect for all commands.

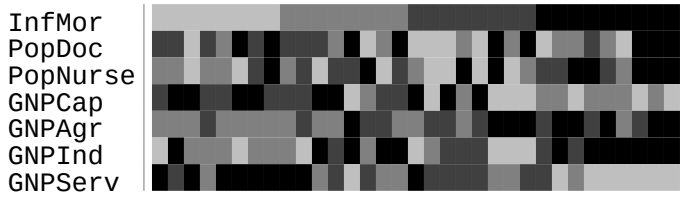
The following example uses the same coding scheme:

```
>SHOW 1-4 CODED

   26 cases
Showing :ICult   (  1) Culture initiative
legend for coded values: (HI far)@ # +   - = &  (LO far)
canton   ICult XCult Form Sucre
 16 AI    7.2   +    -
  6 OW    7.6   +    -    +
  8 GL    9.3        -    -
 15 AR   10.1        -
<more>
```

3.2 Bin coding

The cases are grouped into a number of bins according to some criterion and for each bin a different symbol is used. Several criteria are possible. In the following example the values are divided into four groups containing approximately the same number of cases. (Four groups is the default value). This means that the distribution is broken up into four pieces (fourth, quartiles), i.e. the bin boundaries are the hinges. This is called *fractile coding*.

```
>LIST 1-8 FRACTILES SORT

Bins of equal size (APPROX.); Symbols:
          MILFSCNLFDNFISGUBAIASCGMCBPPHURYA
          OCIN HLU K RRPDK N  aYRLZLRLNSMGL
          NLEL   X   GLAR  D  nPCTCGTNGSNSB
           D D        DN      MREAHRGDRRALN
```



12. The SET GRAPH DISTCODE command (STAT GRAPH DISTCODE shows them) may be used to change these codes for the whole interactive session

Note that in addition to using the FRACTIONAL option to ask for fractional coding the SORT option is used, i.e. we want to concentrate on variable 1.

Note also that the lists look a bit less fancy when ordinary symbols are used; they contain however exactly the same information....

```
   >LIST 1-4 FRACTILES
    26 cases
Bins of equal size (APPROX.); Symbols:.:*#
         ZBLUSONGZFSBBSAASGATTVVNGJ
         HEURZWWLGROSLHRIGRGGIDSEEU
---------------------------------
ICult    |#*::....:**#**..::::##*###
XCult    |:*:#.#*::#***::#:#.##....#
Form     |#::*:...::##**...*::##**##
Sucre    |.**:*#*.:#...::#:*:#.##:*#
```

The symbols used here are the default symbols, '.' for the lowest fourth and '#' highest fourth. On many systems these symbols are replaced by nicer looking graphical symbols. If you want more groups (or less) you will use the "symbols" string to indicate the number of bins, by specifying a series of codes, each character standing for a bin to be defined. Then if you specify three symbols, the variable will be divided into thirds, if you specify 10, tenths and so on.

We call that first form 'fractile coding'. This is usually the default coding used, unless, e.g. in the case of the LIST command, interval coding is used by default (for "historical reasons").

You may request to define the bins by cutting the variable into intervals of equal width (interval length). Again by default it will be divided into four intervals; if you need more or less use the "symbols" string to indicate the number. This is called 'interval coding'.

The READ option can be used to enter your own bin boundaries. The number of bins depend upon the number of codes in the "symbol string" (4 by default). You will then be asked to enter the bin boundaries (one less than bins requested).

3.3 Reference coding

Each symbol reflects the position of a case with respect to some criterion. A criterion often used is the median; then values below the median are marked differently from cases above the median.

```
   >LIST 1-3 REFERENCE

Reference coding below/above the center; Symbols:- +
         ADAABBCDEFFGHIIILLMMNNPPRUSCTT
         L N  UHK I RNRS IUAO LOOOK HCU
         B D  LY    EGLL EXLN  LRU   HR
         A    GP    CRAA   TA  OTM   EQ
-------------------------------------------
GNPAgr   |+-+--++--++++++-------+++---++
GNPInd   |++++-+--+ --++--- ----++++- +
GNPServ  |---++-++--+----+++++++---+++--
```

In this example the symbols express whether a case is below or above the CENTER value, i.e. a reference value stored with each variable. By default this value is the median, but it can be changed to contain other meaningful information. (e.g global percentages and the like).

Values equal or close to the reference value appear in this example as blanks. Equality is not strict equality, but always qualified by a fuzz value, i.e. values lying in the interval [center-fuzz,center+fuzz] are blanked. By default, the EDA system wide fuzz value is used.[13]. LIST REFERENCE has a FUZZ= option used to override the default values.

Changing the FUZZ value, i.e. blanking more or less observations lets you better understand variables in terms of variation around a center value and quickly identify patterns. Again this is meant to be used iteratively, e.g. starting with a large fuzz value and lowering it each time you reproduce the same table.

```
>LIST 1-3 REFERENCE FUZZ=8
```

13. It can be set using the SET FUZZ command

```
  30 countries
Reference coding below/above the center; Symbols:- +
         ADAABBCDEFFGHIIILLMMNNPPRUSCTT
         L N  UHK I RNRS IUAO LOOOK HCU
         B D  LY    EGLL EXLN  LRU   HR
         A    GP    CRAA   TA  OTM   EQ
---------+--------------------------------
GNPAgr   |+          ++ +       + +    +
GNPInd   |++   +-    - + -    -  + +   +
GNPServ  |-   +-    + -      +  - -   --

 >LIST REFERENCE FUZZ=12

Reference coding below/above the center; Symbols:- +
         ADAABBCDEFFGHIIILLMMNNPPRUSCTT
         L N  UHK I RNRS IUAO LOOOK HCU
         B D  LY    EGLL EXLN  LRU   HR
         A    GP    CRAA   TA  OTM   EQ
---------+--------------------------------
GNPAgr   |+           + +
GNPInd   |+    +             -  + +   +
GNPServ  |-    -      -      +  - -   -
```

The example uses the fuzz=2.5 option to tell EDA, that equality is not strict equality by the value plus or minus 2.5. Some commands (e.g. the LIST command) have additional options for the reference value.

Instead of only marking low/high positions it is also possible to indicate the distance of a case from a reference value, by either using different symbols or by using more than one symbol, e.g. a single plus sign for a case close to the median, and 2, 3 etc plus signs for cases farther away.

In the following example the reference is the median, and the distance to the median is expressed using units of 1/2 midspreads, i.e. each + symbol shows a distance of 1/2 midspread.

```
    >LIST 1-4 CODED

   26 cases
variable listing
units of 1/2.0 midspread
case        ICult    XCult     Form    Sucre

  1 ZH          +                +        -
  2 BE                                    +
  5 SZ                   -
  6 OW          -       ++        -        +
  7 NW                            -
  8 GL          -                 -        -
  9 ZG
 10 FR                    +              +++
 11 SO                             +       -
 12 BS         ++                 ++      ---
 13 BL                                      --
```

3.4 Marking

All cases corresponding to some criterion are marked with a special symbols all other cases are not marked.

```
   >LIST 1-4 MARK IF>45
   26 cases
Mark values greater than <val>; Symbols:@
         ZBLUSONGZFSBBSAASGATTVVNGJ
         HEURZWWLGROSLHRIGRGGIDSEEU
----------------------------------
ICult
XCult      @ @   @     @ @ @    @
Form
Sucre        @   @     @   @ @  @
```

In the previous example the cases above 45 are marked with the EDA marking symbol (it might be different in your EDA version, and you can chang GRAPHSYMBOL command.

4 Command syntax

These various forms of coding are available with a number of commands, namely the LIST, MAP, PLOT, CASID and HISTOGRAM command.

Some commands might present slight differences with respect to the forms shown above and the syntax explained below. Namely default values might be different and some default codes might be changed. E.g. It is not always desirable to show blank “symbols” on a plot, as you will see nothing at that particular location...

The syntax chart below is taken from the HISTOGRAM command:

```
<code.opt>
    BINS | [FRAC] | EXACT | READ [“symbols”
    DISTRIBUTIONAL [SIMPLE] [“symbols”]
    REFERENCE=value [“Symbols”] [FUZZ=val]
    MARK|=val | IF>val |IF=val| IF<val | IF~val
       [“symbols”] [FUZZ=val]
```

It shows the various forms of coding. The first line shows the various forms of ‘bin coding’, i.e. FRACTILE (i.e. the default option), BINS requests interval coding, EXACT defines exact fractiles and READ allows you to enter bin boundaries. Optionally “symbols” is used to define alternate symbols.

DISTRIBUTIONAL requests distributional coding. SIMPLE does not distinguish between lower and upper far-out, out and adjacent values. Finally symbols is used to enter alternative symbols.

REFERENCE requests reference coding. You may specify in addition different symbols and a fuzz value.

MARK requests marking. There are four conditions equality (IF=value, or just MARK=value), greater or less than a value (IF<val, IF>val) or inequality (If~val). For the equality /inequality option you may specify a FUZZ value. The “symbols” string is used to specify other symbols than the default symbols.

```
 >HISTOGRAM 1 CODE=7
   33 countries
Histogram:InfMor(1) Infant Mortality
Position#  1 is (  7) GNPServ %GNP for Services
Bins: equal counts (approx.); Symbols:
midpoint
     5.1 |
     7.7 |
    10.3 |
    12.9 |
    15.6 |
    18.2 |
    20.8 |
    23.4 |
    26.0 |
    28.6 |
    31.2 |
    33.9 |
    36.5 |
    39.1 |
    41.7 |
```

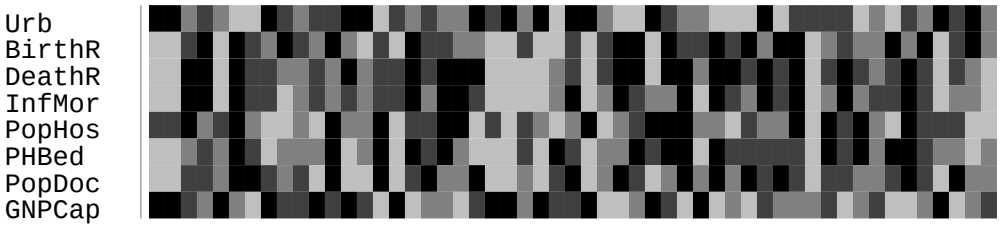
5 Lists are a powerful tools

Despite their technical simplicity, lists are very powerful analytical tools, especially the coded forms.

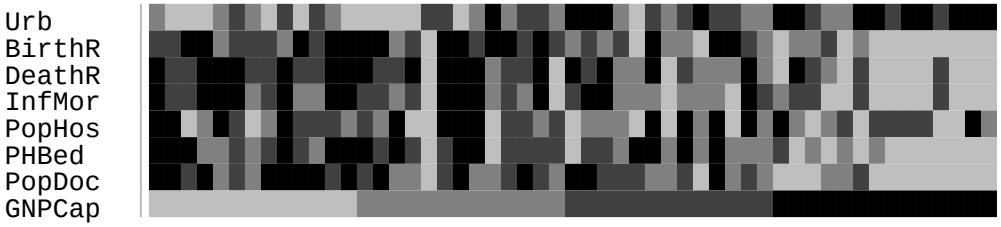
Interactivity is an essential feature of the presented tool. You should never produce only a single list if you are looking for patterns; produce several lists changing symbols (especially using blanks to show structure); used different numbers of bins change the criteria and use sorting.

Sorting, ordering: Date tables become much more readable, patterns can be seen easily if the variables and observations are show in some meaningful order. There are of course the SORT and ORDER options available with the LIST command which let you sort the list on a particular variable.

```
   >LIST 1-8 FRACTILES

 53 countries  out of 183 countries
Bins of equal size (APPROX.); Symbols:
         AAABBBBCCCCCDEEGGGGGGGMSKRLLLMMMMMMNNNOCRSSSSSSTTTTZZZ
         FLNEOUUAAOOOJGTAAHNNNREEEEIIAAAAAOAIIUAWAAEIOWACOUAAI
         RGGNTRRMPMNTIYHBMAEBECYNUSBBDLLRUZMGGGFAHONEUANHGNIMM
         IEOISKUEVOGEBPIOBNEIQECYIOEYAAIORAIRAARNOTERDZZAOIRBB
```




```
   >LIST 1-8 FRACTILES SORT K=8
  53 countries
Bins of equal size (APPROX.); Symbols:
         METMGSMBTOZGNMRSNSSBGZTGBKGLCCLMZSECCASMDCNBCTMAAGSLR
         OTAANIAUCUAAIAWOIAAUNAOHEENIAOEAIEGOANWAJOAOAURLFAEIE
         ZHNLBEDRHGIMGLAUGOHREMGANNEBFMSUMNYTPGARINMTMNCGRBYBU
         AIZAIRAUAARBRINDATOKEBONIYQEROORBEPEVOZOBGISEIEEIOCYI
```



Variables also can be reordered on a particular observation (ORDER option) to change the sequence of the variables. The sequence of the variables could of course be given on a command list

```
>LIST 29 31
12 11 10 1 3
```

i.e. you specify the variables in an order meaningful to you. EDA offers additional possibilities based on statistical criteria.

These possibilites however are not options to the LIST command, but available as a general feature with the DESCRIBE, DS and VARS family of commands. [14] The three commands offer options to change the variable list based on some statistical criterion. There are many options; the purpose of the command line examples below is just to hint at possibilities...

```
   >VARS 1-20 SORT DESCENDING MEDIAN
   >LIST FRACTILES
```

The VARS command changes the variable list, in a way that the first variable has the highest median, followed by the second and so on. The LIST command then takes the current variable list to produce a coded table for the variables.

14. The difference between the three commands is only the presentation (DESCRIBE shows all descriptive information for the variables), DS, for Describe Short shows only variable names, whereas VARS shows only the current variable list (numbers).

```
>VARS 1-20 SORT DESCENDING MAXIMUM IF<20
>LIST FRACTILES
```

The second example uses the MAXIMUM as criterion; the IF option will only keep variables with a maximum value less than 20, i.e. the variables list will not longer contain 20 variables (unless all variables satisfy the condition).

Many criteria are available. These options are meant to be explored....

Table des Matières

1 Numerical lists 1
2 The SHOW command 2
3 Coded lists 5
3.1 Distributional coding 5
3.2 Bin coding 6
3.3 Reference coding 7
3.4 Marking 8
4 Command syntax 9
5 Lists are a powerful tools 9