

A Graphical User Interface for Environmental Statistics

Rudolf Dutter

**Department of Statistics and Probability Theory
Vienna University of Technology, Austria**

Dortmund, Germany

August 13, 2008



Vienna University of Technology

Statistical Data Analysis *Explained*

Applied Environmental Statistics with R

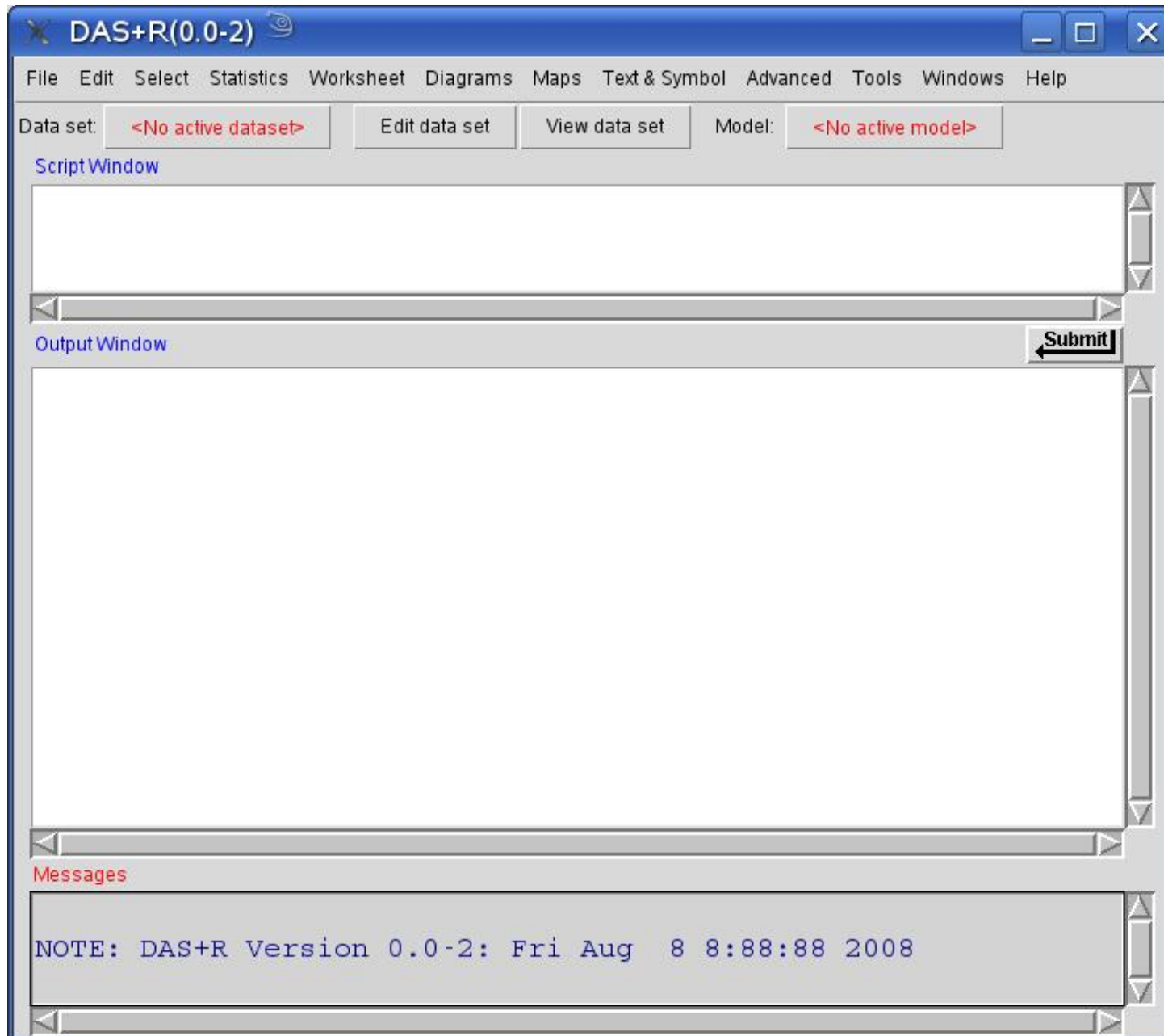
 WILEY

DAS+R a companion of
this book (Wiley, 2008).

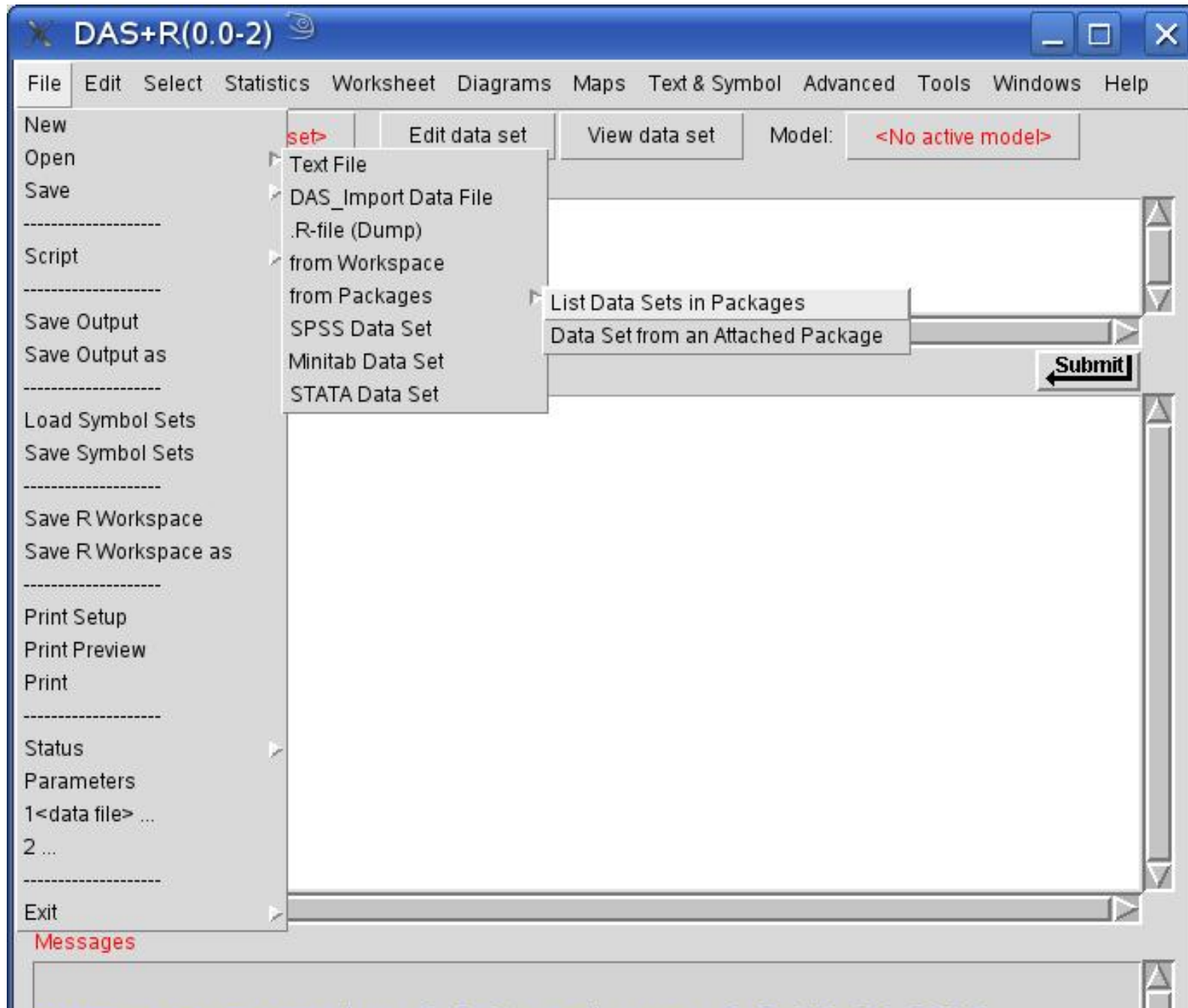
Emphasis:

- Spatial Data
- Strong Graphical Tools
- Data from
Geochemistry
- Little Mathematics
- Provided Software:
DAS+R₁

- Basis: DAS, (all) functions of R, Tcl/Tk
- Emphasis:
 - easy to use (e.g. Rcommander)
 - fast reproducibility (with small changes) of the results
 - interactive definition of subsets (numerically and graphically)
 - strong interaction between statistical data and spatial information
 - graphical analysis



Open Data File



Read DAS/Imp (Preview)

Read Data From DAS/Imp File: /home/dutt/TEACHING/2[...]_prak_05/schwind/new/

Enter name for data set:

Missing data indicator:

Field Separator

White space

Commas

Tabs

Other Specify:

Decimal-Point Character

Period [.]

Comma [,]

Preview: /home/dutt/TEACHING/2004-05/proj_prak_05/schwind/new/new/DASplusR7/KOLA95_C2MM.csv

```

HEADER                                KOLA PROJECT, regional sampling 1995 (Finland (FIN), Norway (
COMMENT DATASET                        C-Horizon of Podsol profiles, air dried, fraction <2 mm
SAMPLE IDENTIFIER  ID
COORDINATES          XC00  YC00
COMMENT VARIABLES
EXTRACTION                                Aqua Regia
METHOD                                GF-AAS
UDL
LDL                                0.001
UNIT          m east  m north  m a.s.l.          cm          mg/kg
VARIABLE          ID XC00  YC00  ELEV  COUN ASP  TOPC  LITO  Ag
1  547960 7693790 135  FIN  NW  35  20  0.01
2  770025 7679170 140  RUS  SW  52  4  0.01
3  498651 7668150 255  FIN  N  52  31  0.021
4  795152 7569390 240  RUS  NE  40  20  0.022

```

OK Cancel Help

Read DAS/Imp (Types of Variables)

Read Data From DAS/Imp File: /hon

ID	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
XCOO	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
YCOO	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
ELEV	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
COUN	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input type="radio"/> factor	<input checked="" type="radio"/> character
ASP	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input type="radio"/> factor	<input checked="" type="radio"/> character
TOPC	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
LITO	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
Ag	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
Ag_INAA	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
Al	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
Al_XRF	<input type="radio"/> logical	<input type="radio"/> integer	<input checked="" type="radio"/> double	<input type="radio"/> factor	<input type="radio"/> character
Al2O3	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
As	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
As_INAA	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
Au	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
Au_INAA	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character
P	<input type="radio"/> logical	<input type="radio"/> integer	<input type="radio"/> double	<input checked="" type="radio"/> factor	<input type="radio"/> character

Make DAS Data

Make DAS Data

Active Data Set: KOLA95_C2MM

Header: KOLA PROJECT, regional sampling 1995 (Finland (FIN), Norway (NOR) and F

Comment: <empty>

Variable containing sample identification (pick one)

ID
Ir_INAA
K
K20

X-coordinate (pick one) Y-coordinate (pick one)

XCOO Yb_INAA
Y YCOO
Yb_INAA Zn
YCOO Zn_INAA

Ranges of

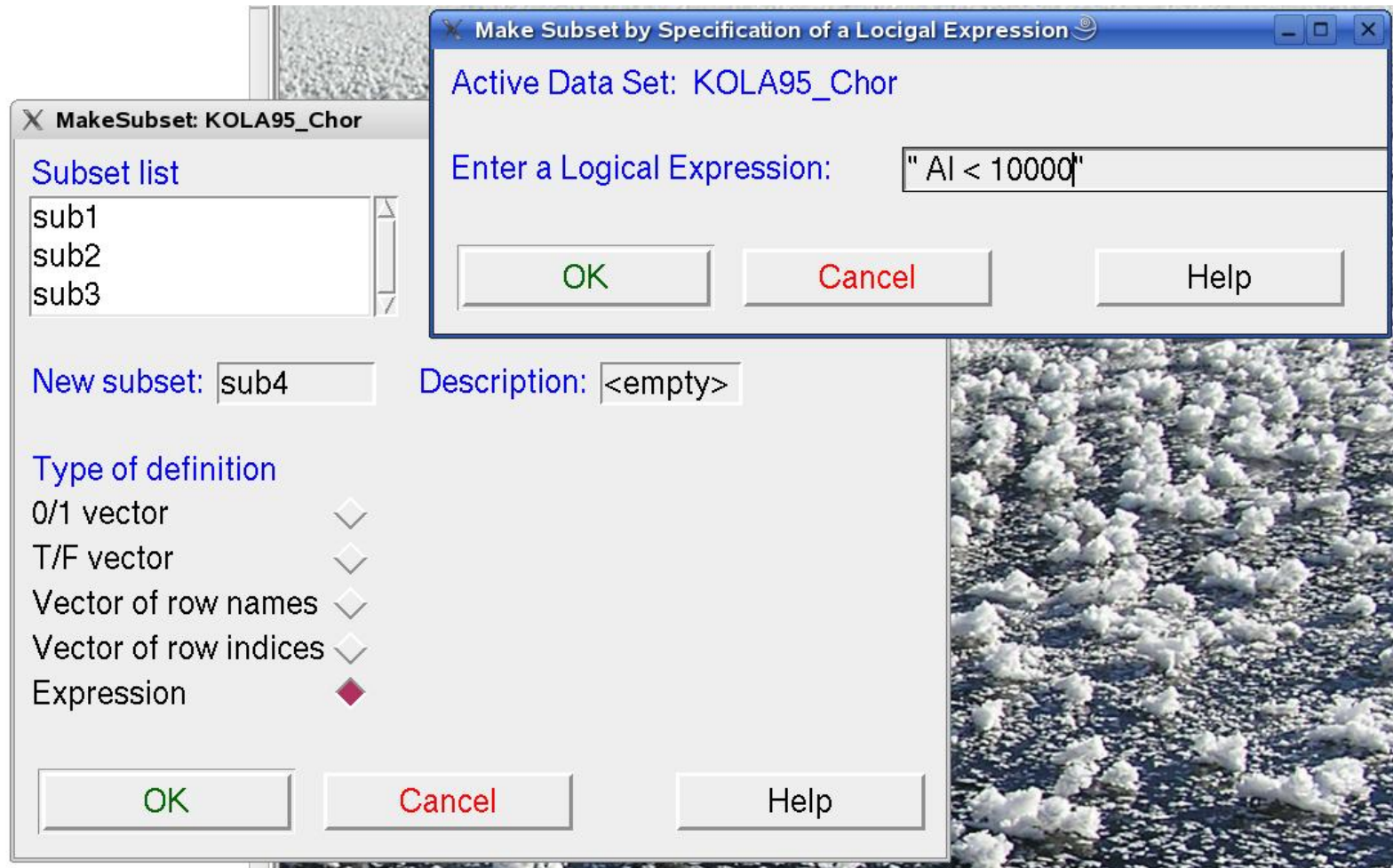
X-coordinate: Y-coordinate:

min: <empty> max: <empty> min: <empty> max: <empty>

Variable Specific Labelling

	Units	Extrac. Meth.	Meth. of Anal.	Lo. Detec. Lim.	Up. Detec. Lim.	Comments
ID						
XCOO	m east					
YCOO	m north					
ELEV	m a.s.l.					
COUN						
ASP						
TOPC	cm					
LITO						

Make Subset



(Pre-)Select Variables

File Edit Select Statistics Worksheet Diagrams Maps Text & Symbol Ad

Data set:

Script Win

```

invis
invis
summar
f.hist
makeSu
makeSu
select

```

Output Win

```

[1] "
> make
Subset
> make

```

Select Variables: KOLA95_Chor

Variables (pick one to select resp. to deselect)

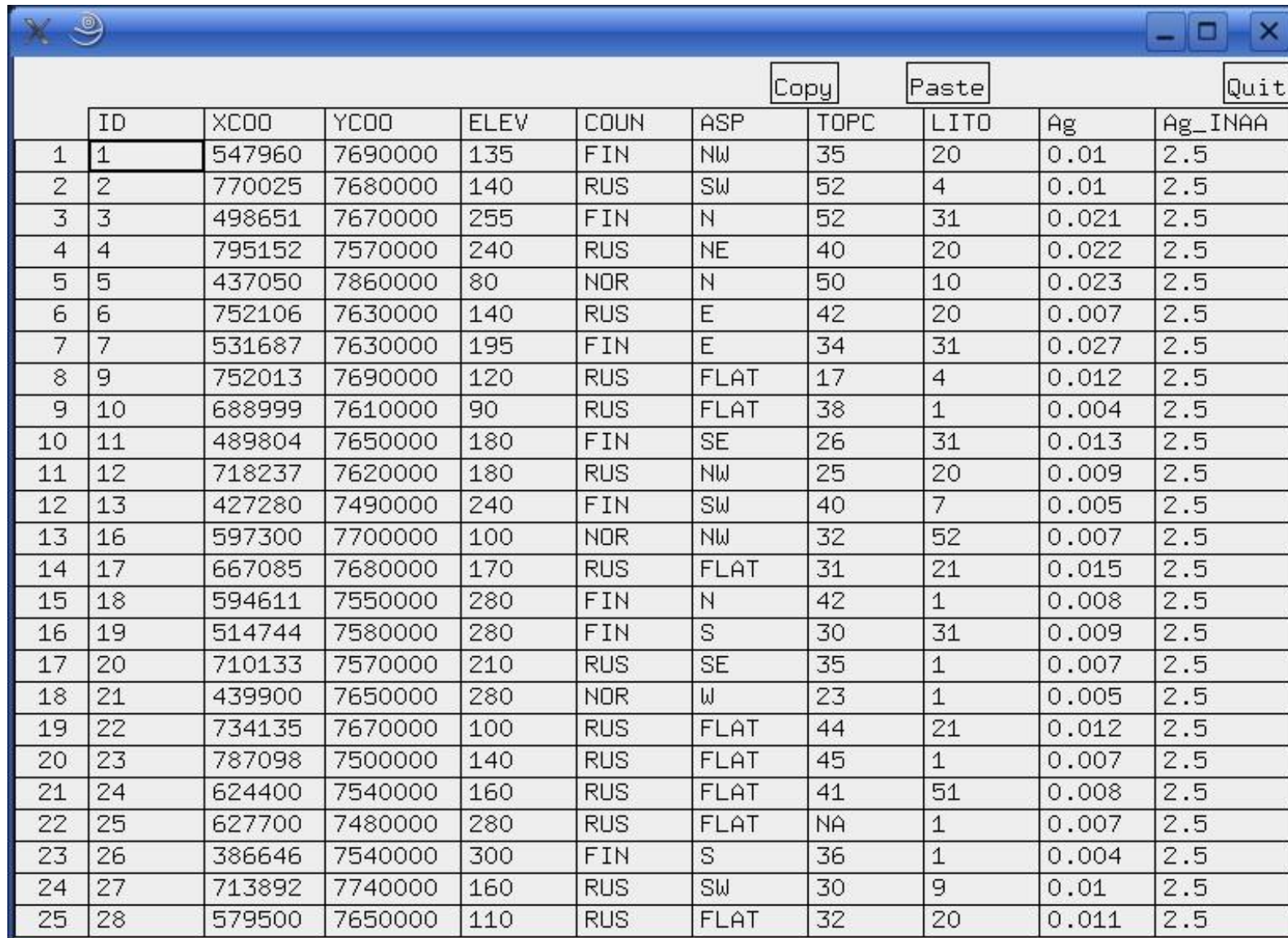
1: Ag
Ag_INAA
2: Al
Al2O3
Al_XRF
5: As
As_INAA
ASP
4: Au
Au_INAA
B
3: Ba
Ba_INAA
Be
Bi
Br_IC
Br_INAA
Ca
Ca_INAA
CaO

Select all

Deselect all

- Numerically
- Graphically
 - Histogram
 - Boxplots
 - xy-plot
 - Ternary Plot
 - Scattermatrix (Draftman)

View Data Set



	ID	XCOO	YCOO	ELEV	COUN	ASP	TOPC	LITO	Ag	Ag_INAA
1	1	547960	7690000	135	FIN	NW	35	20	0.01	2.5
2	2	770025	7680000	140	RUS	SW	52	4	0.01	2.5
3	3	498651	7670000	255	FIN	N	52	31	0.021	2.5
4	4	795152	7570000	240	RUS	NE	40	20	0.022	2.5
5	5	437050	7860000	80	NDR	N	50	10	0.023	2.5
6	6	752106	7630000	140	RUS	E	42	20	0.007	2.5
7	7	531687	7630000	195	FIN	E	34	31	0.027	2.5
8	9	752013	7690000	120	RUS	FLAT	17	4	0.012	2.5
9	10	688999	7610000	90	RUS	FLAT	38	1	0.004	2.5
10	11	489804	7650000	180	FIN	SE	26	31	0.013	2.5
11	12	718237	7620000	180	RUS	NW	25	20	0.009	2.5
12	13	427280	7490000	240	FIN	SW	40	7	0.005	2.5
13	16	597300	7700000	100	NDR	NW	32	52	0.007	2.5
14	17	667085	7680000	170	RUS	FLAT	31	21	0.015	2.5
15	18	594611	7550000	280	FIN	N	42	1	0.008	2.5
16	19	514744	7580000	280	FIN	S	30	31	0.009	2.5
17	20	710133	7570000	210	RUS	SE	35	1	0.007	2.5
18	21	439900	7650000	280	NDR	W	23	1	0.005	2.5
19	22	734135	7670000	100	RUS	FLAT	44	21	0.012	2.5
20	23	787098	7500000	140	RUS	FLAT	45	1	0.007	2.5
21	24	624400	7540000	160	RUS	FLAT	41	51	0.008	2.5
22	25	627700	7480000	280	RUS	FLAT	NA	1	0.007	2.5
23	26	386646	7540000	300	FIN	S	36	1	0.004	2.5
24	27	713892	7740000	160	RUS	SW	30	9	0.01	2.5
25	28	579500	7650000	110	RUS	FLAT	32	20	0.011	2.5

Histogram

DAS+R(0.0-2)

File Edit Select Statistics Worksheet Diagrams Maps Text & Symbol Advanced Tools Windows Help

Data set: Edit d

Script Window

```
activeDataSet("KOLA95_
importDASImpData("/hon
```

Model:

ING/2004-05/proj_prak_05

Submit

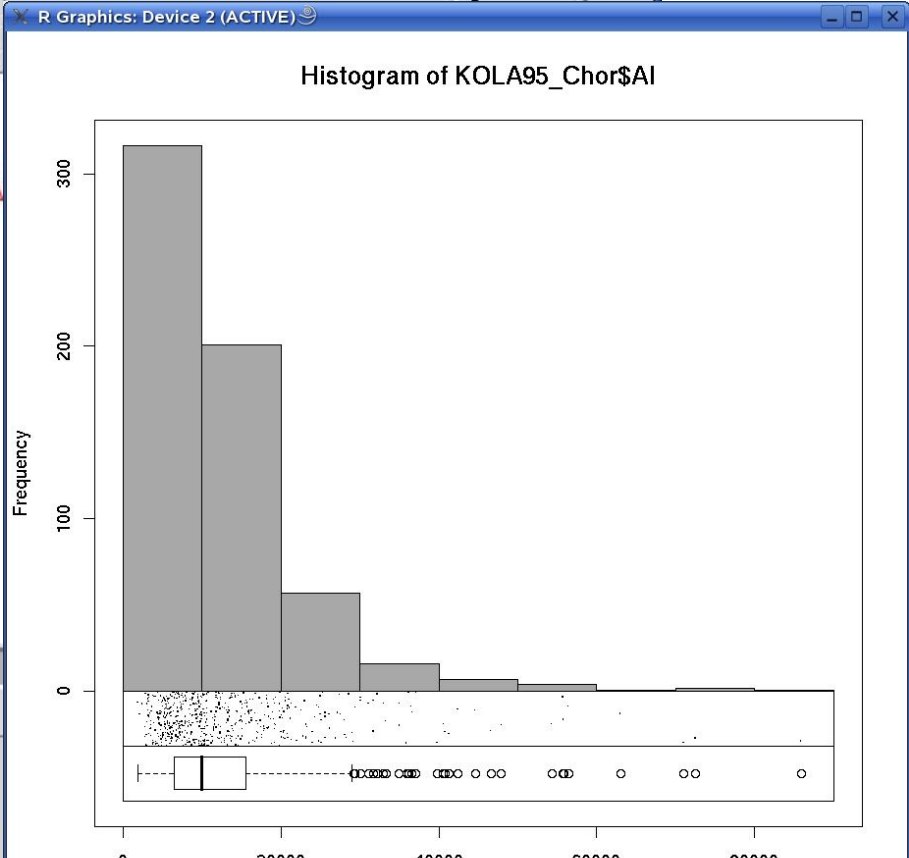
Diagram menu items: Histogram, Density, Cdf, Boxplot, Xyplot, Ternary Plot, Draftman, Smooth, Save Graph to File

Output Window

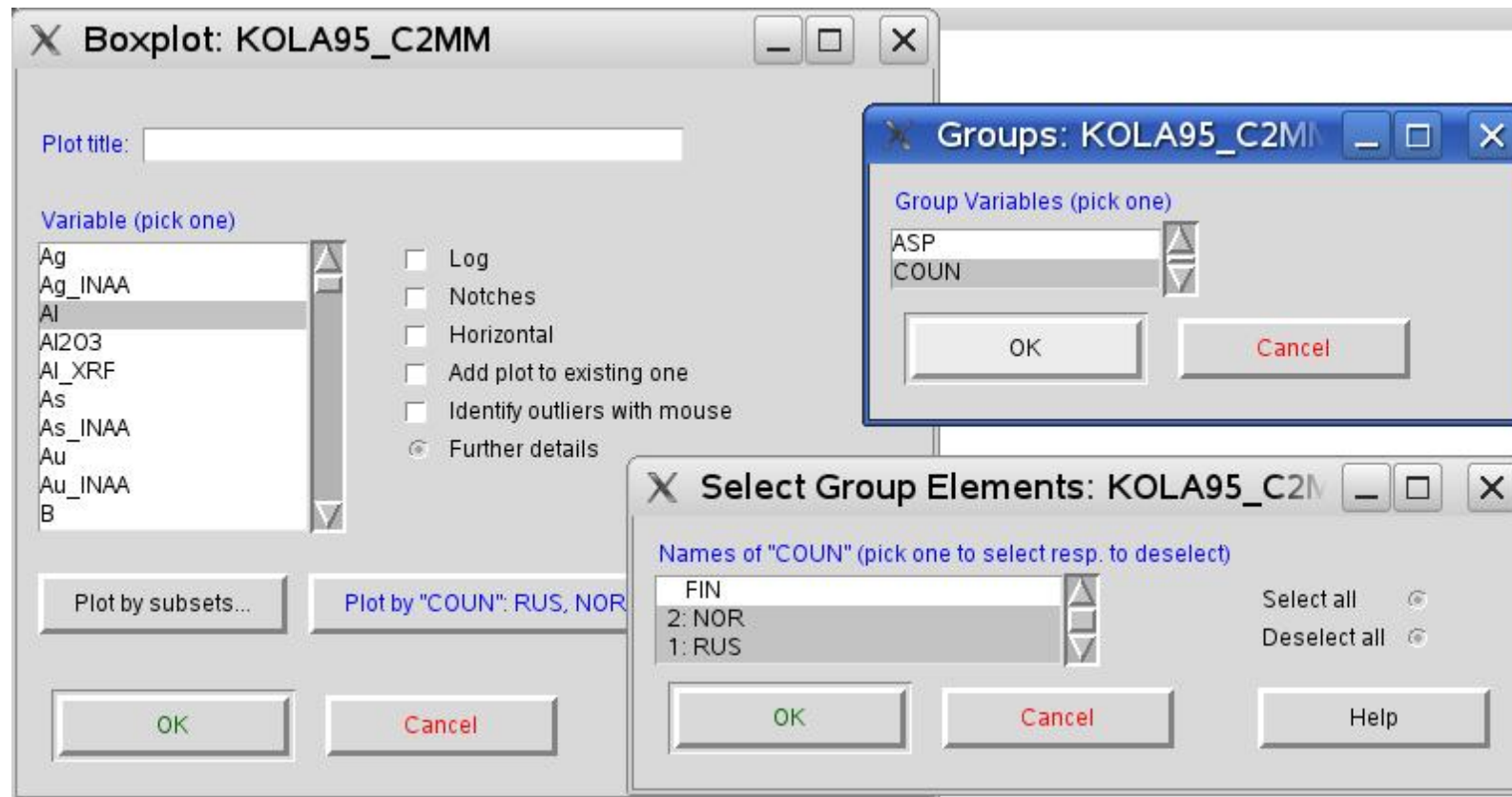
```
> activeDataSet("KOLA95_Chor")
> importDASImpData("/home/dutt/TEA
```

Messages

NOTE: The dataset KOLA95_C2MM has

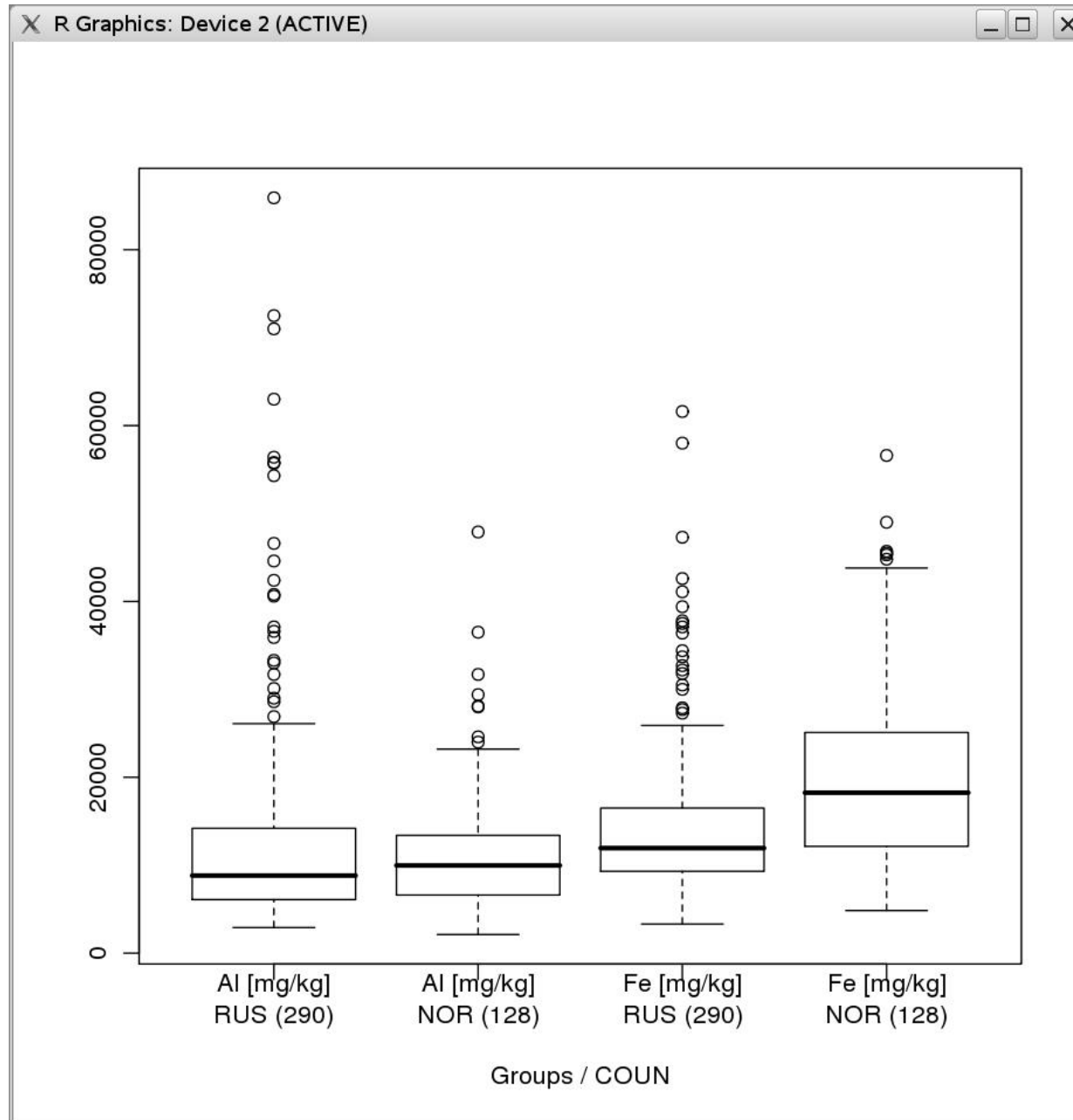


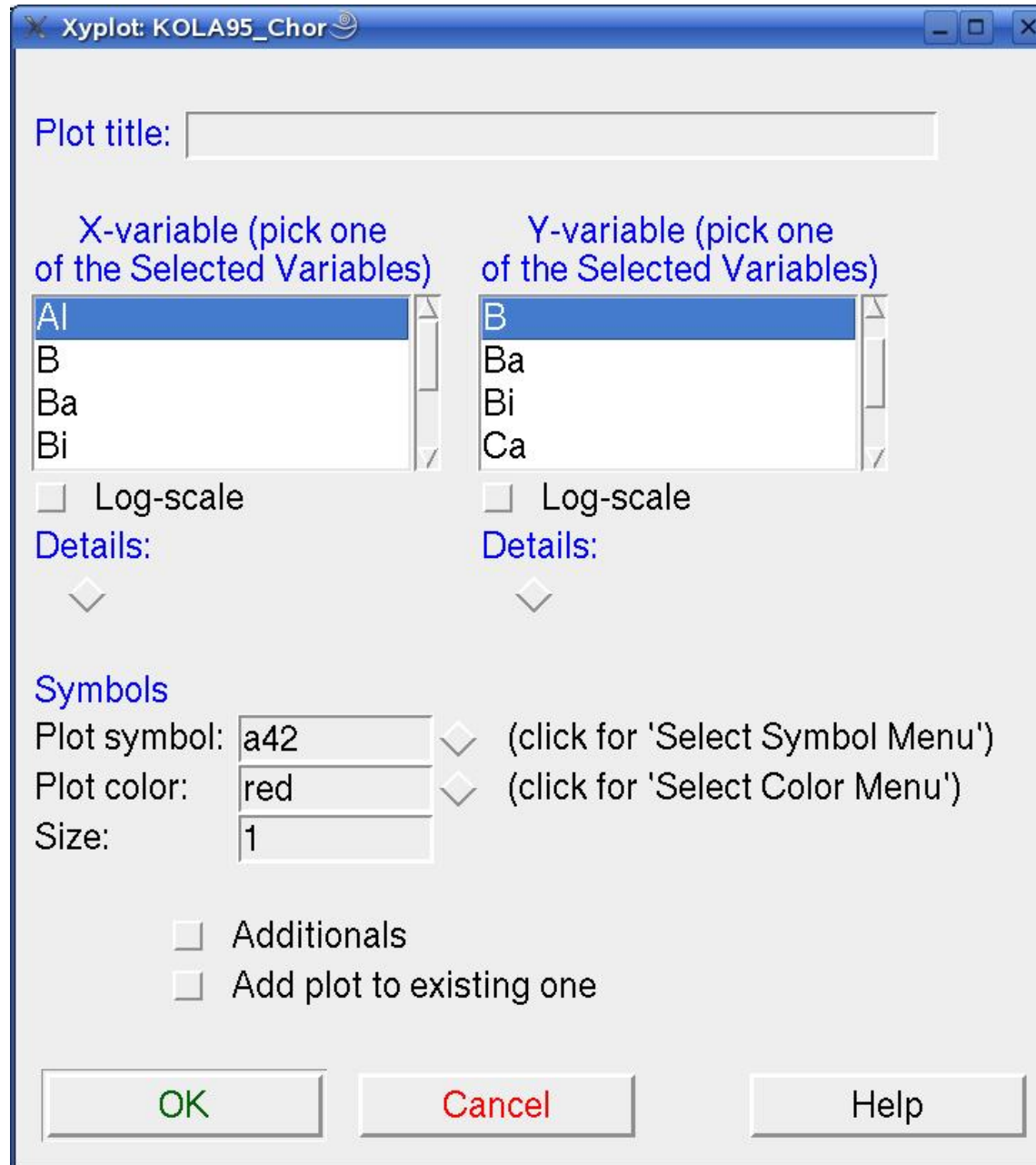
Boxplots



The image shows three overlapping dialog boxes from the Minitab software interface:

- Boxplot: KOLA95_C2MM**: The main dialog box. It has a "Plot title" field, a "Variable (pick one)" list with "AI" selected, and several checkboxes: "Log", "Notches", "Horizontal", "Add plot to existing one", "Identify outliers with mouse", and "Further details" (which is selected). At the bottom, there are "Plot by subsets..." and "Plot by 'COUN': RUS, NOR" buttons, along with "OK" and "Cancel" buttons.
- Groups: KOLA95_C2MM**: A smaller dialog box titled "Group Variables (pick one)". It has a list with "ASP" and "COUN" (selected). It includes "OK" and "Cancel" buttons.
- Select Group Elements: KOLA95_C2MM**: A dialog box titled "Names of 'COUN' (pick one to select resp. to deselect)". It has a list with "FIN", "2: NOR", and "1: RUS" (selected). It includes "Select all", "Deselect all", "OK", "Cancel", and "Help" buttons.





Xyplot: Specifications

The image shows two overlapping dialog boxes from the Xyplot software. The background window is titled "Xyplot: KOLA95_Chor <2>". It features a "Plot title:" field, two lists for "X-variable" and "Y-variable" (with "Al" and "Ca" selected), checkboxes for "Log-scale", and a "Symbols" section with fields for "Plot symbol: s11", "Plot color: red", and "Size: 1". There are also checkboxes for "Additional" and "Add plot to existing one", and "OK", "Cancel", and "Help" buttons.

The foreground window is titled "Plot Symbols With Identification Numbers". It displays a 5x5 grid of 25 numbered symbols (0-25) for selection. The symbol in cell 11 (a red star) is highlighted with a green background. The symbols include various shapes like circles, triangles, crosses, diamonds, squares, and stars in different colors (red, yellow, black). "OK", "Cancel", and "Help" buttons are at the bottom.

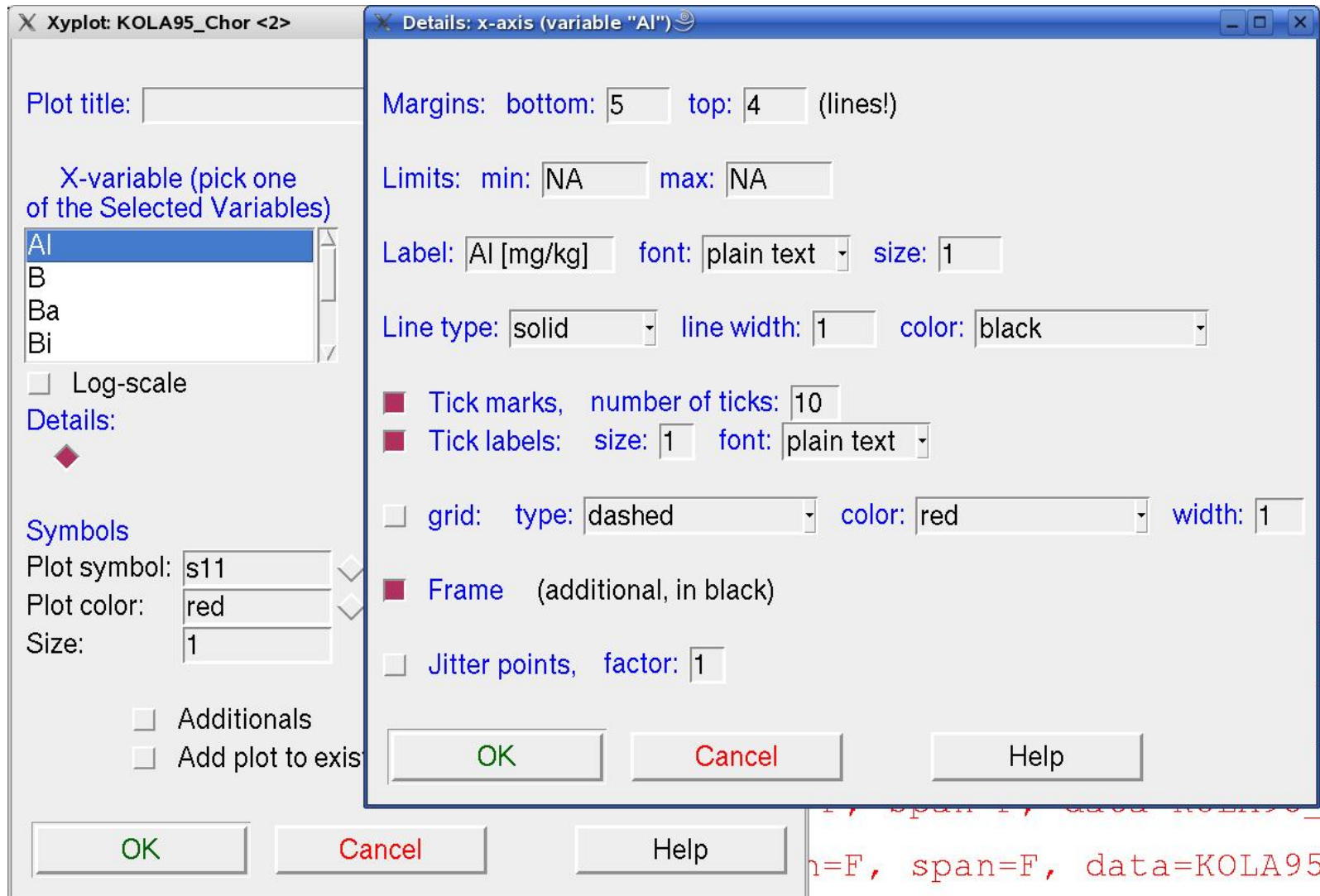
Below the dialog boxes, a snippet of R code is visible:

```

g] ", xlab="COUN", data=KOLA95_Cho
g] ", xlab="COUN", data=KOLA95_Cho
g] ", xlab="COUN", data=KOLA95_Cho
=F, span=F, data=KOLA95_Chor, l

```

Xyplot: Specifications



Xyplot: KOLA95_Chor <2>

Plot title:

X-variable (pick one of the Selected Variables)

- Al
- B
- Ba
- Bi

Log-scale

Details:

Symbols

Plot symbol:

Plot color:

Size:

Additional

Add plot to exist

Details: x-axis (variable "Al")

Margins: bottom: top: (lines!)

Limits: min: max:

Label: font: size:

Line type: line width: color:

Tick marks, number of ticks:

Tick labels: size: font:

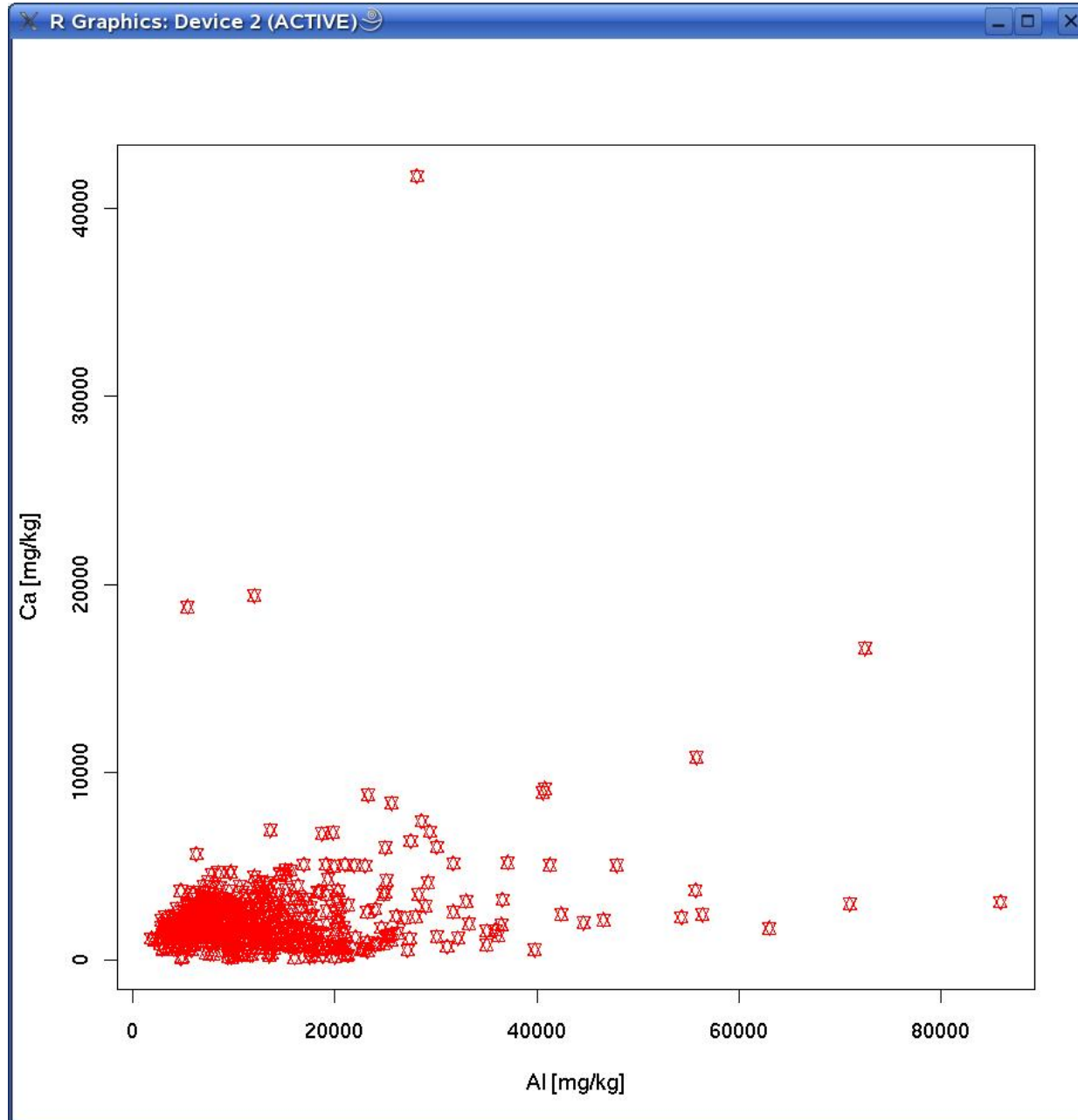
grid: type: color: width:

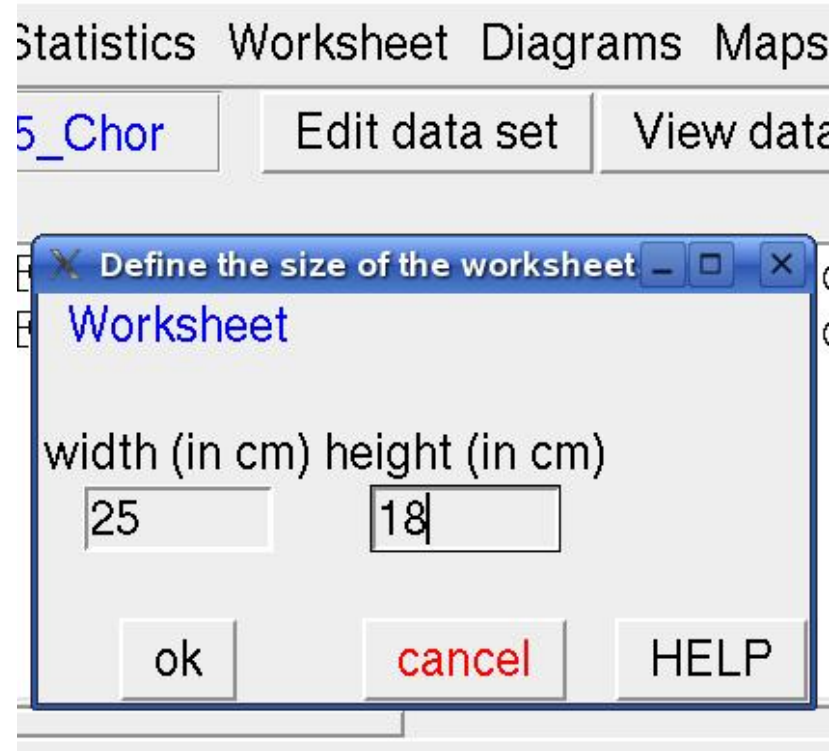
Frame (additional, in black)

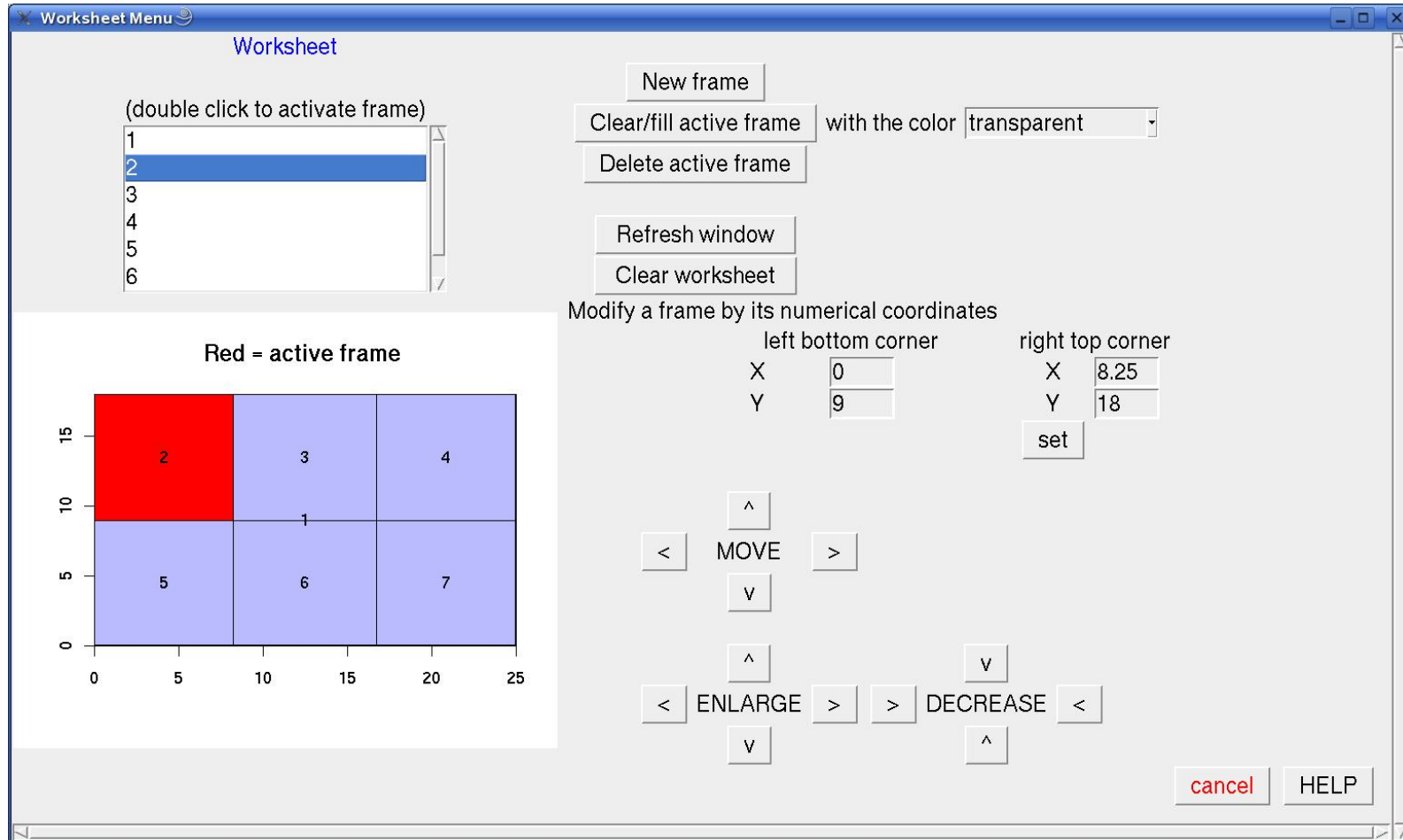
Jitter points, factor:

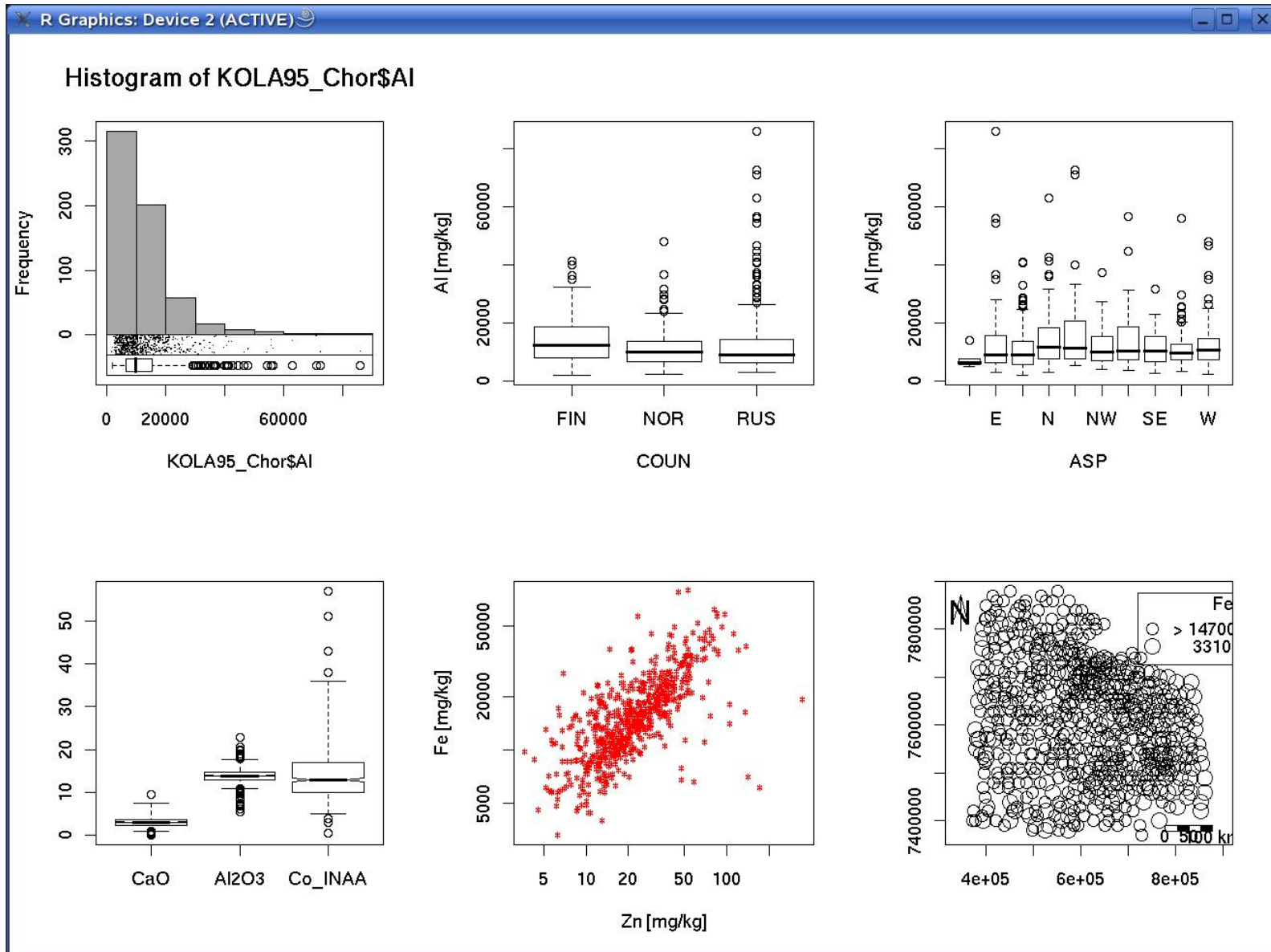
OK Cancel Help

OK Cancel Help

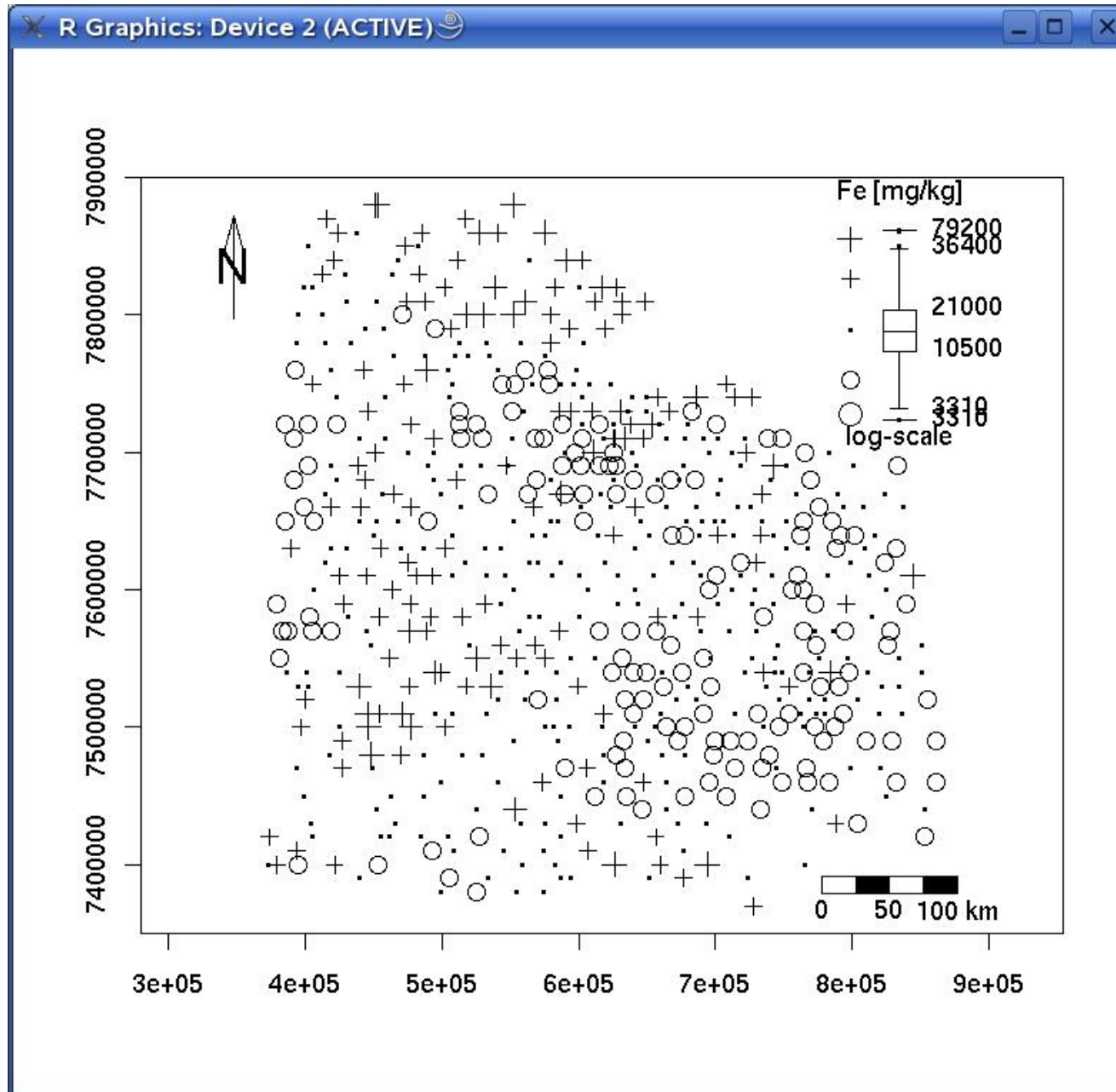




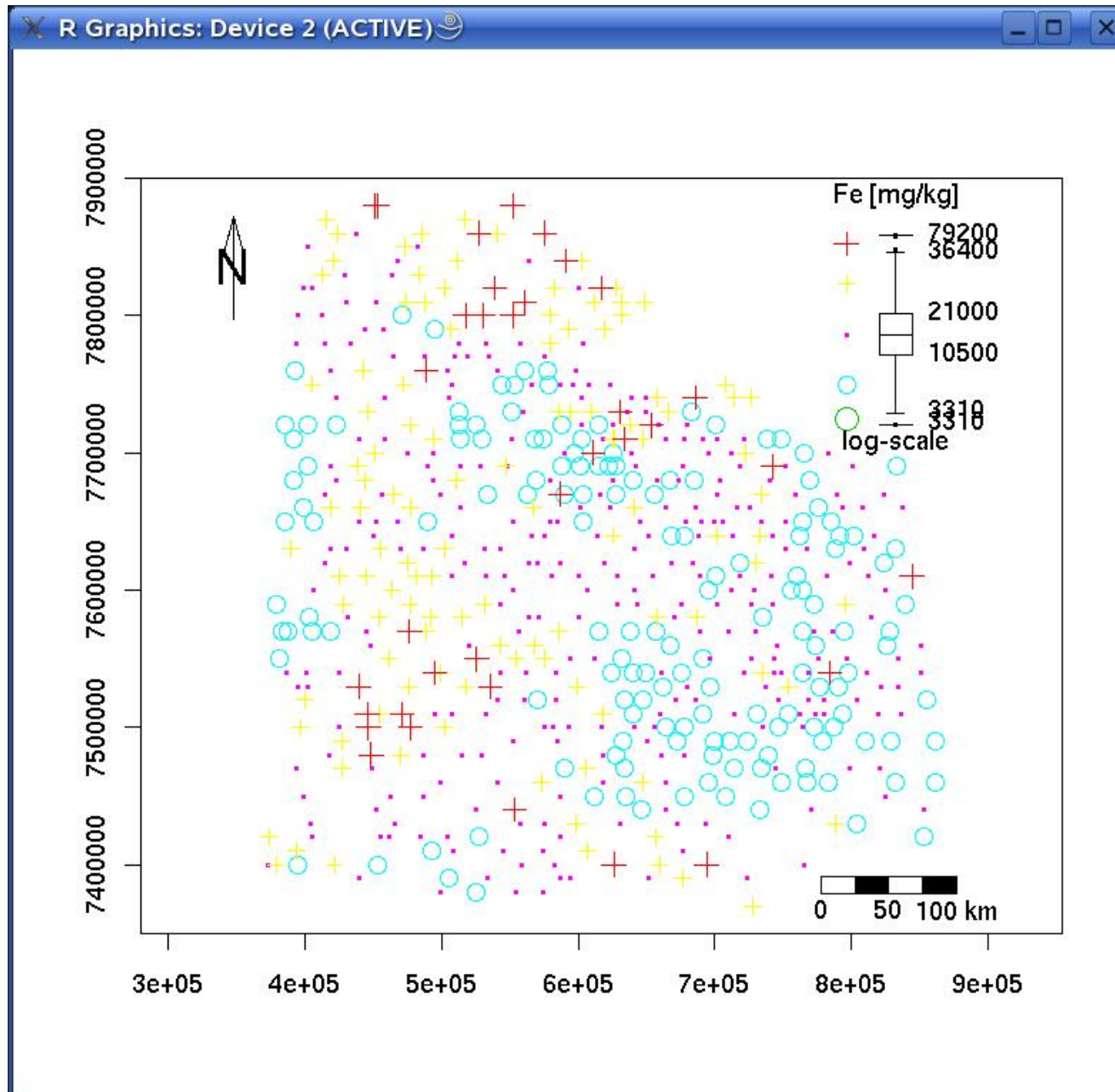




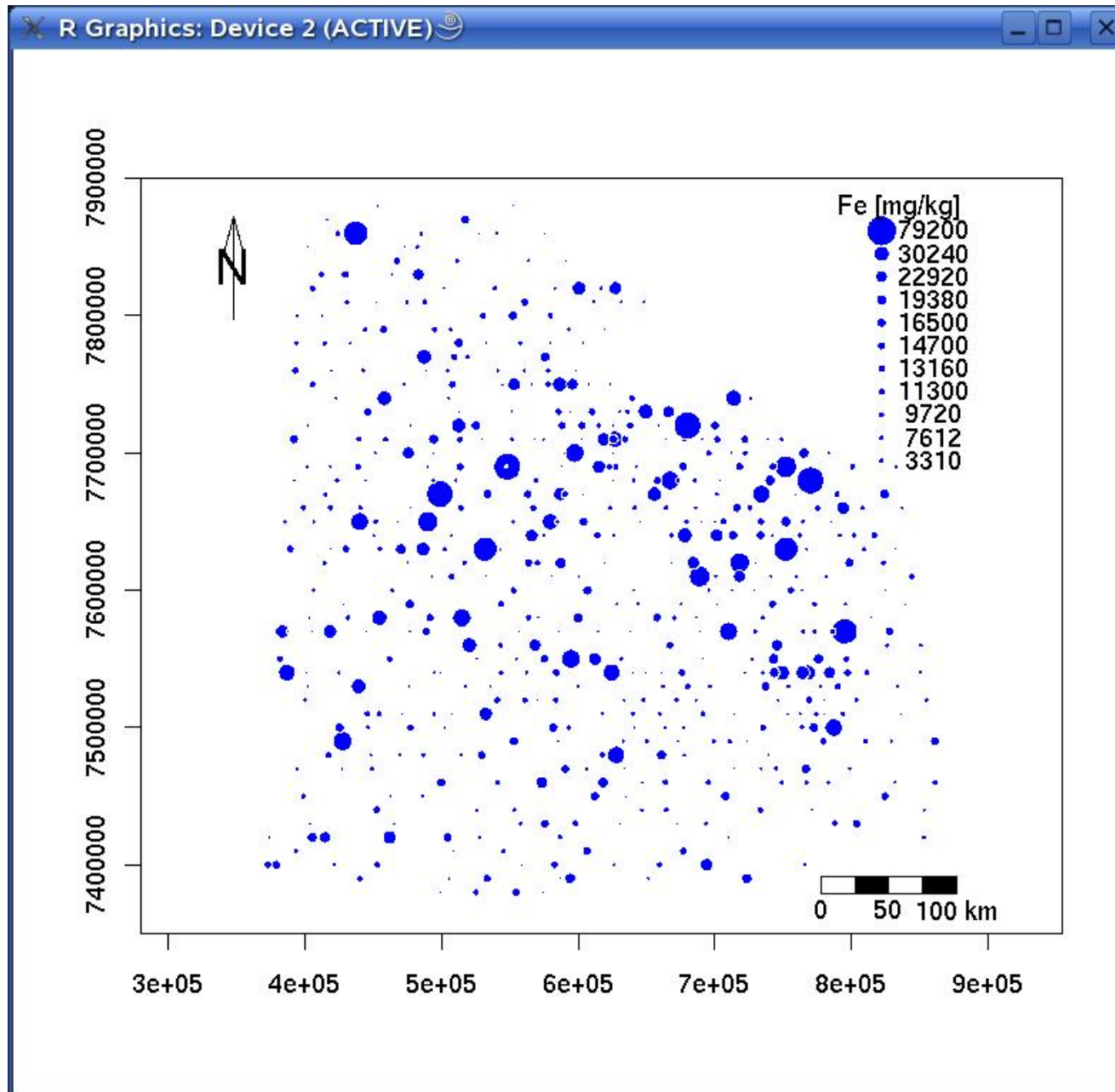
Mapping: Boxplot



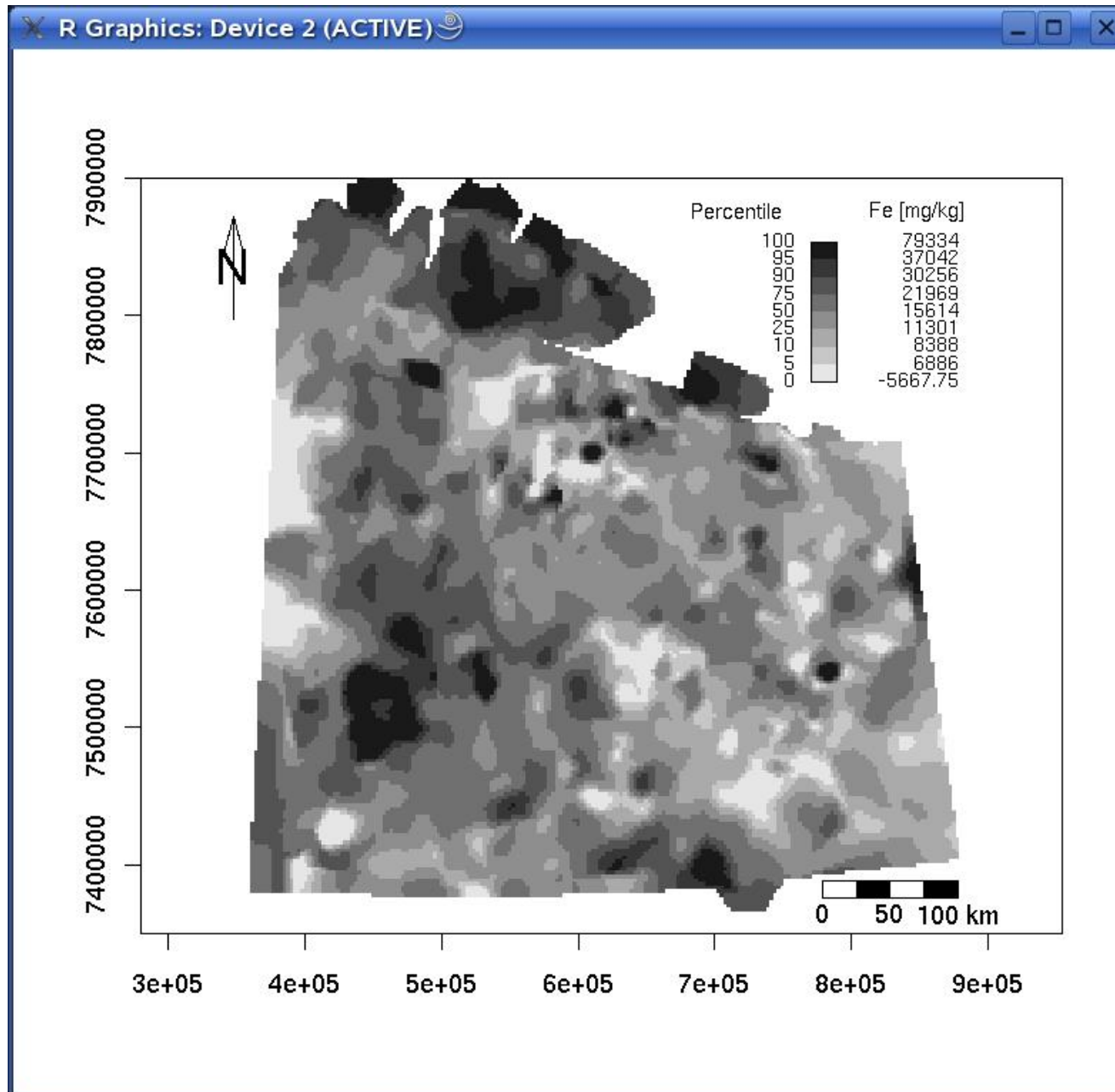
Mapping: Boxplot



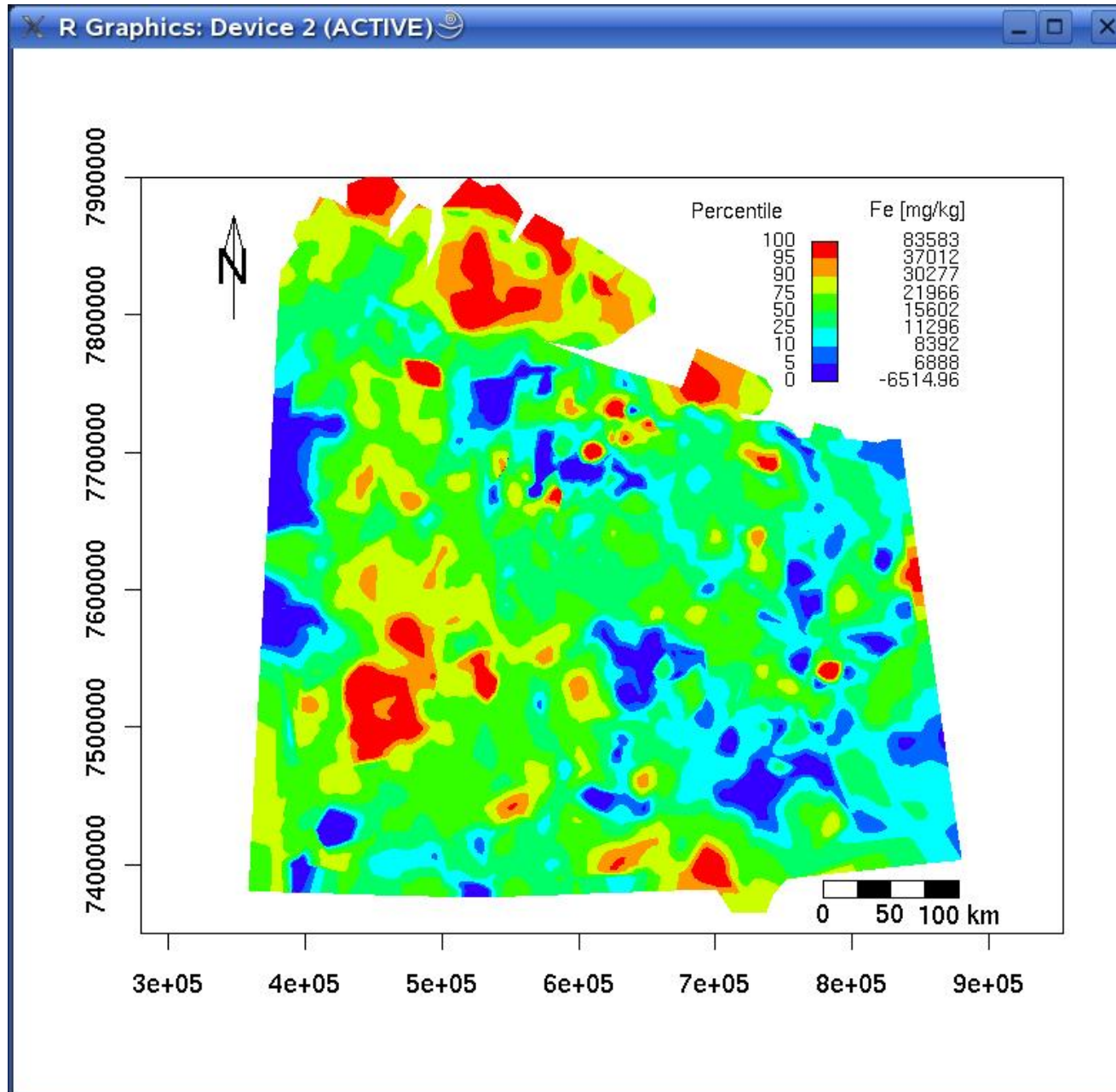
Mapping: Proportional Plot



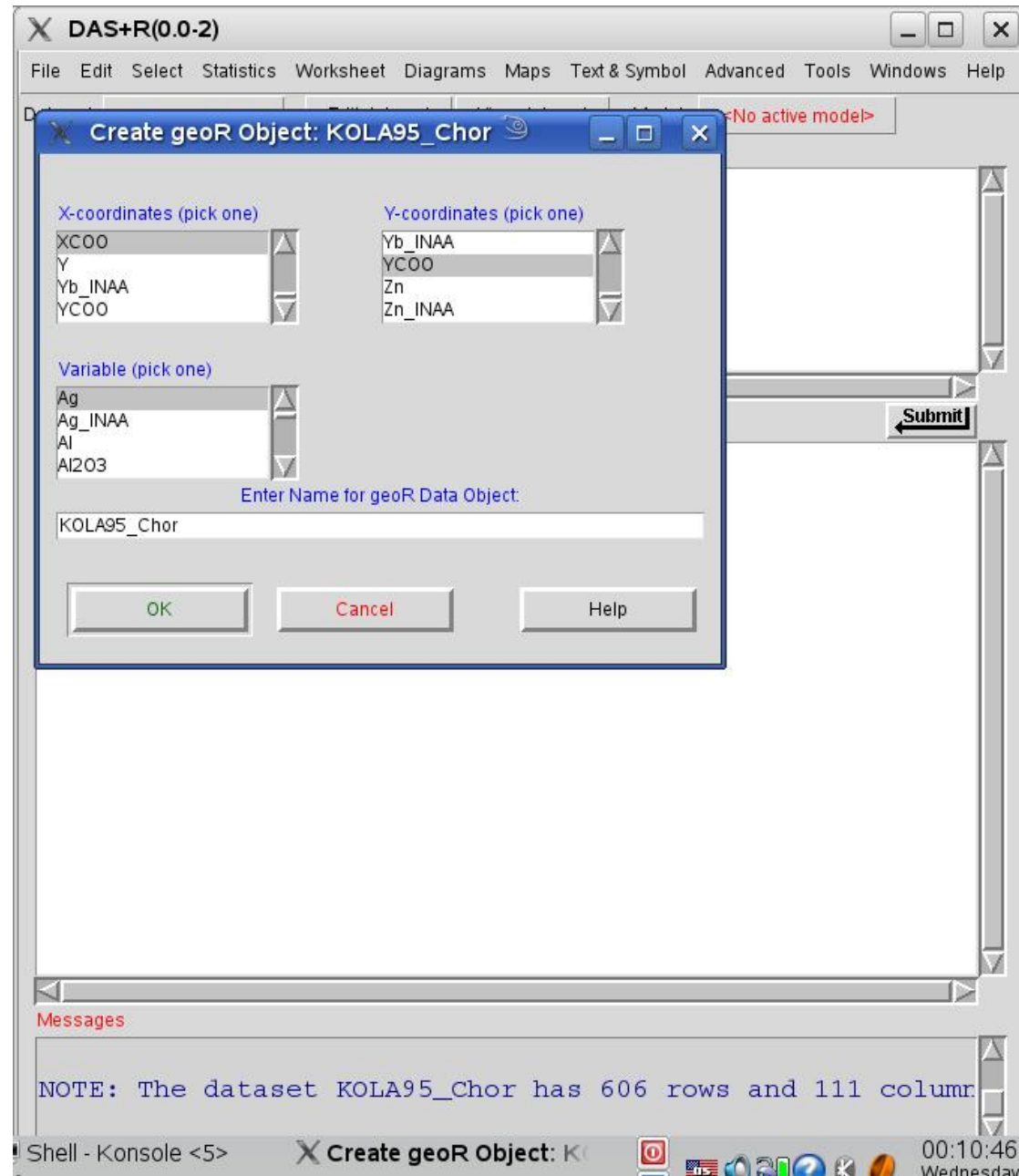
Mapping: Simple Smoothing



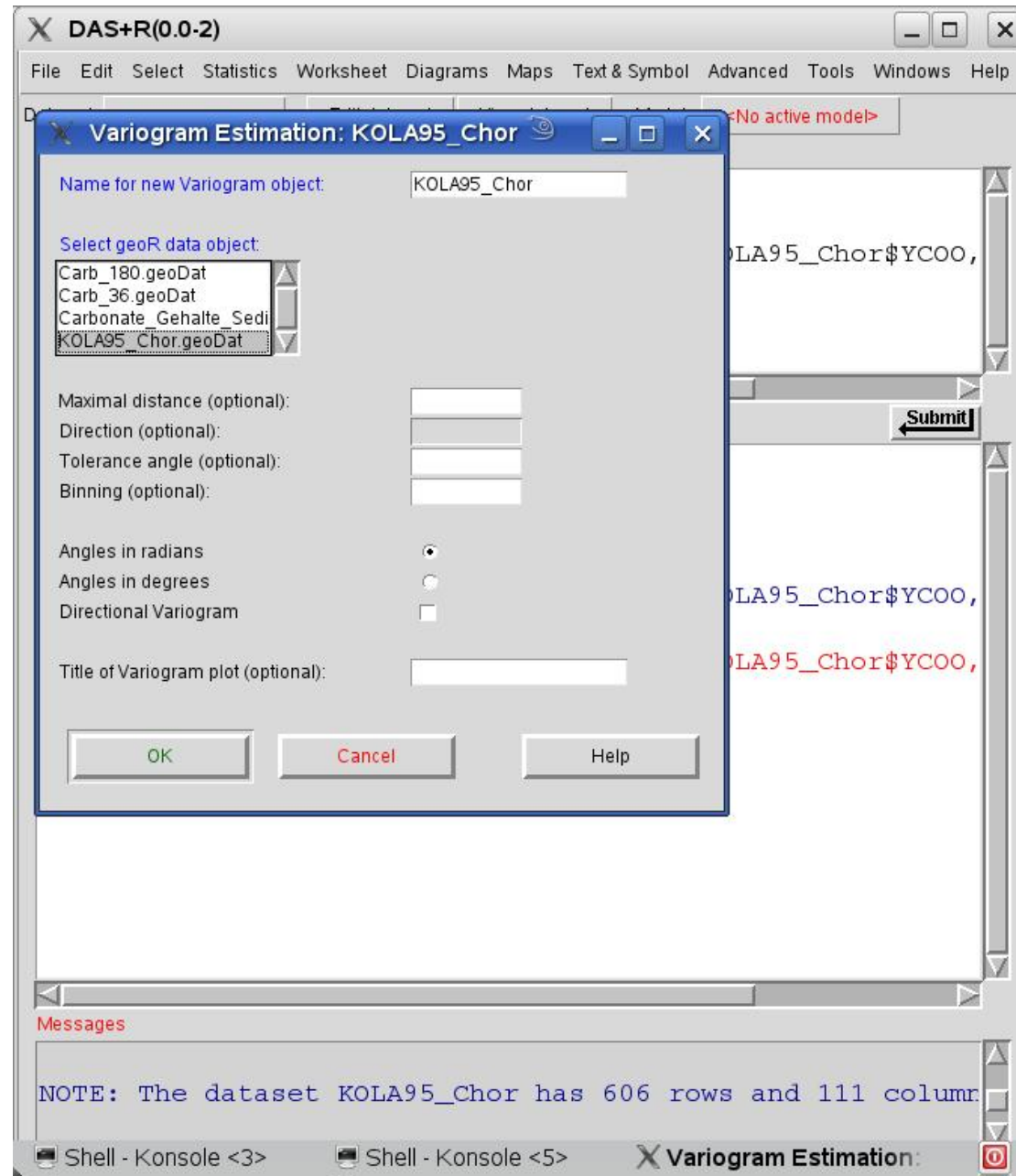
Mapping: Simple Smoothing



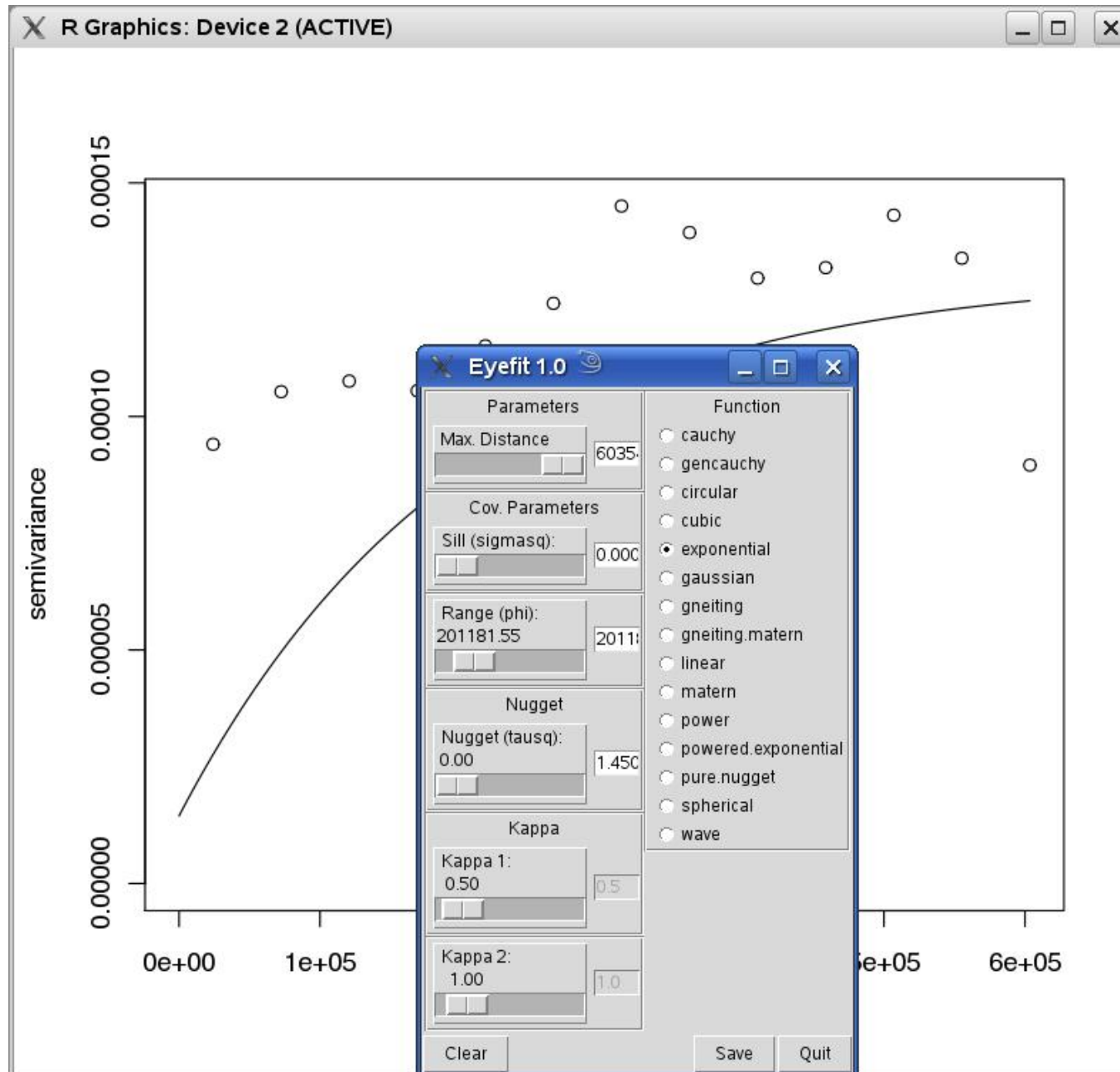
Mapping: Kriging



Mapping: Kriging



Mapping: Kriging



Principal Components

Principal Component Analysis: KOLA95_Chor

Variables (pick several of the Selected Variables)

B
Ba
Co_INAA
Zn
Bi
Ca

Select all
 Deselect all

Log-transform

Use of

Previously Calculated Principal Components
 Correlation (Standardization)
 Robust Covariance Matrix Estimation

Plot Type

Screeplot
 Biplot

Name for the object to be returned

princomp.obj

Default

OK Cancel Help

Principal Components

```

Output Window
> princomp.obj <- princompDAS(data=KOLA95_Chor, vars=c("Al", "CaO", "Fe", "Al2O3", "
> summary(princomp.obj)
Importance of components:
              PC1          PC2          PC3          PC4          PC5
Standard deviation  1.9886845  1.5158522  1.0274942  0.9298833  0.79809531
Proportion of Variance 0.3954866  0.2297808  0.1055744  0.0864683  0.06369561
Cumulative Proportion 0.3954866  0.6252674  0.7308418  0.8173101  0.88100572
              PC6          PC7          PC8          PC9          PC10
Standard deviation  0.65051344  0.53951065  0.44207959  0.38933541  0.35872949
Proportion of Variance 0.04231677  0.02910717  0.01954344  0.01515821  0.01286868
Cumulative Proportion 0.92332250  0.95242967  0.97197311  0.98713132  1.00000000

```


Principal Components

Screeplot: KOLA95_Chor

Plot Title: KOLA95_Chor

X-axis Label: Principal Component Number

Eigenvalues Details

Axis Label: Eigenvalue

Plot Color: gray32 (click for 'Select Color' menu)

Plot Symbol: s1 (click for 'Select Symbol' menu)

Plot Symbol Size: 1

Line Type: dashed

Line Width: 1

Cumulative % Details

Axis Label: Cumulative %

Plot Color: red (click for 'Select Color' menu)

Plot Symbol: s4 (click for 'Select Symbol' menu)

Plot Symbol Size: 1

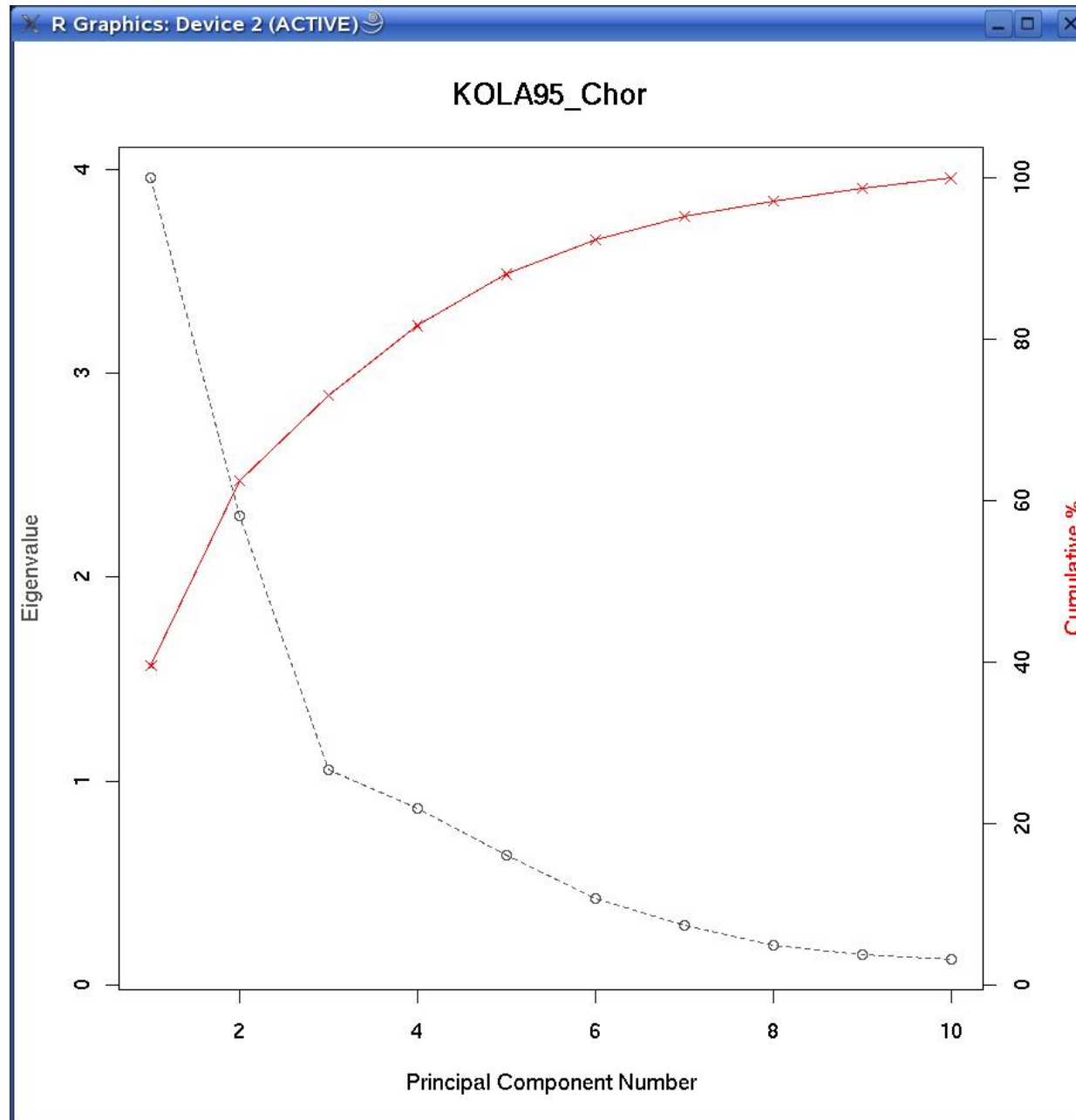
Line Type: solid

Line Width: 1

Default

OK Cancel Help

Principal Components



Biplot: KOLA95_Chor

Plot Title:

X-variable (pick one)

- PC1
- PC2
- PC3
- PC4
- PC5

Y-variable (pick one)

- PC2
- PC3
- PC4
- PC5
- PC6

Scores Details

Plot Color: (click for 'Select Color' menu)

Plot String:

Plot String Size:

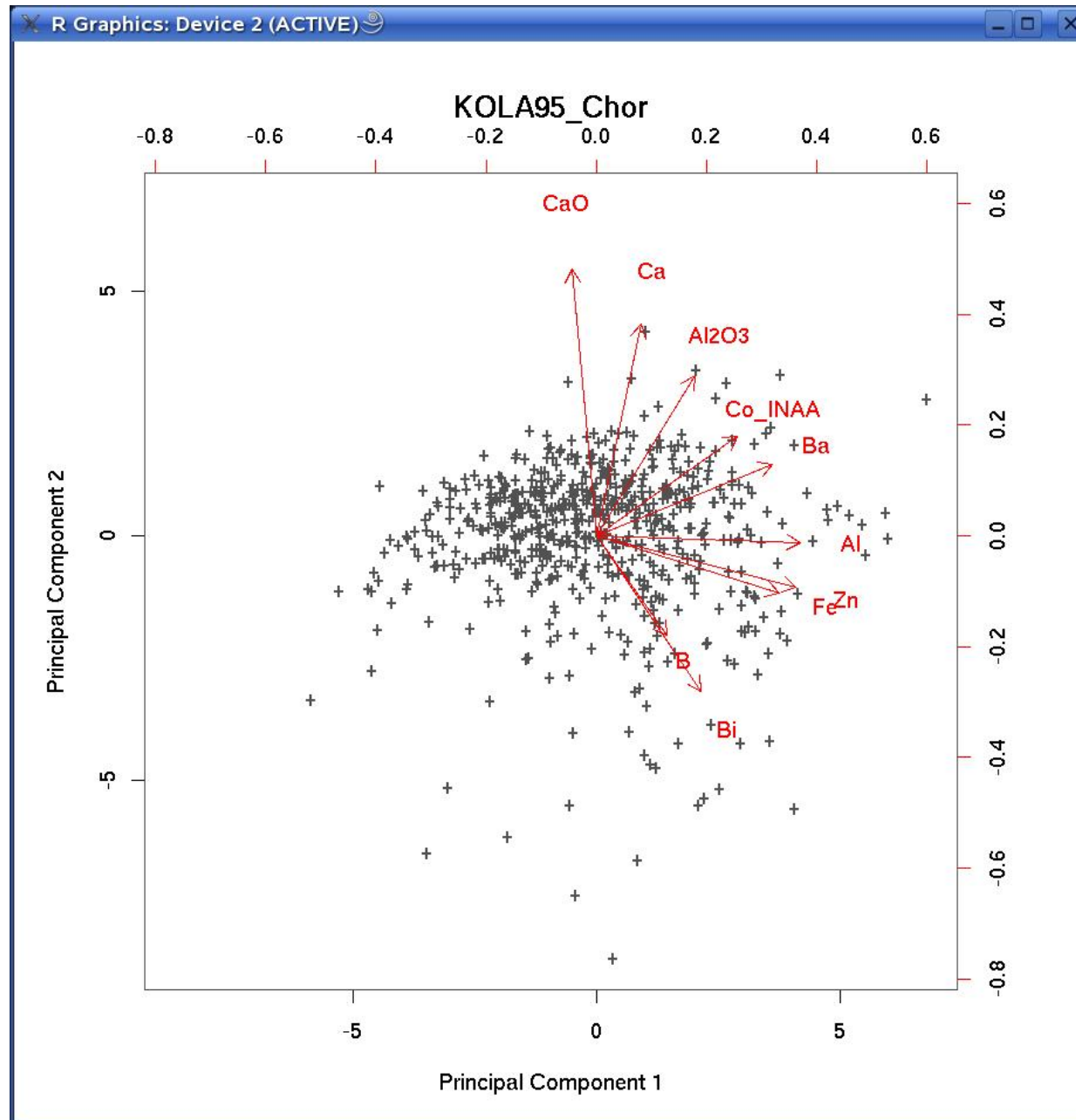
Loadings Details

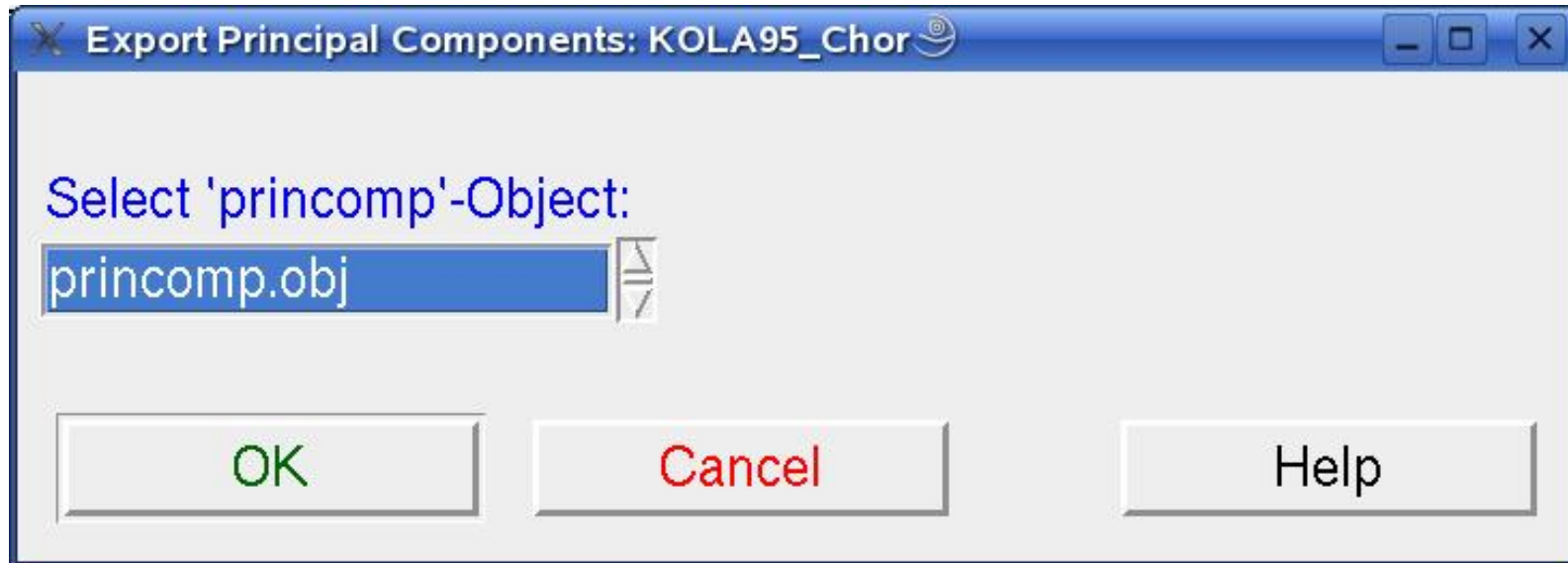
Arrow Color: (click for 'Select Color' menu)

Arrow Head Length:

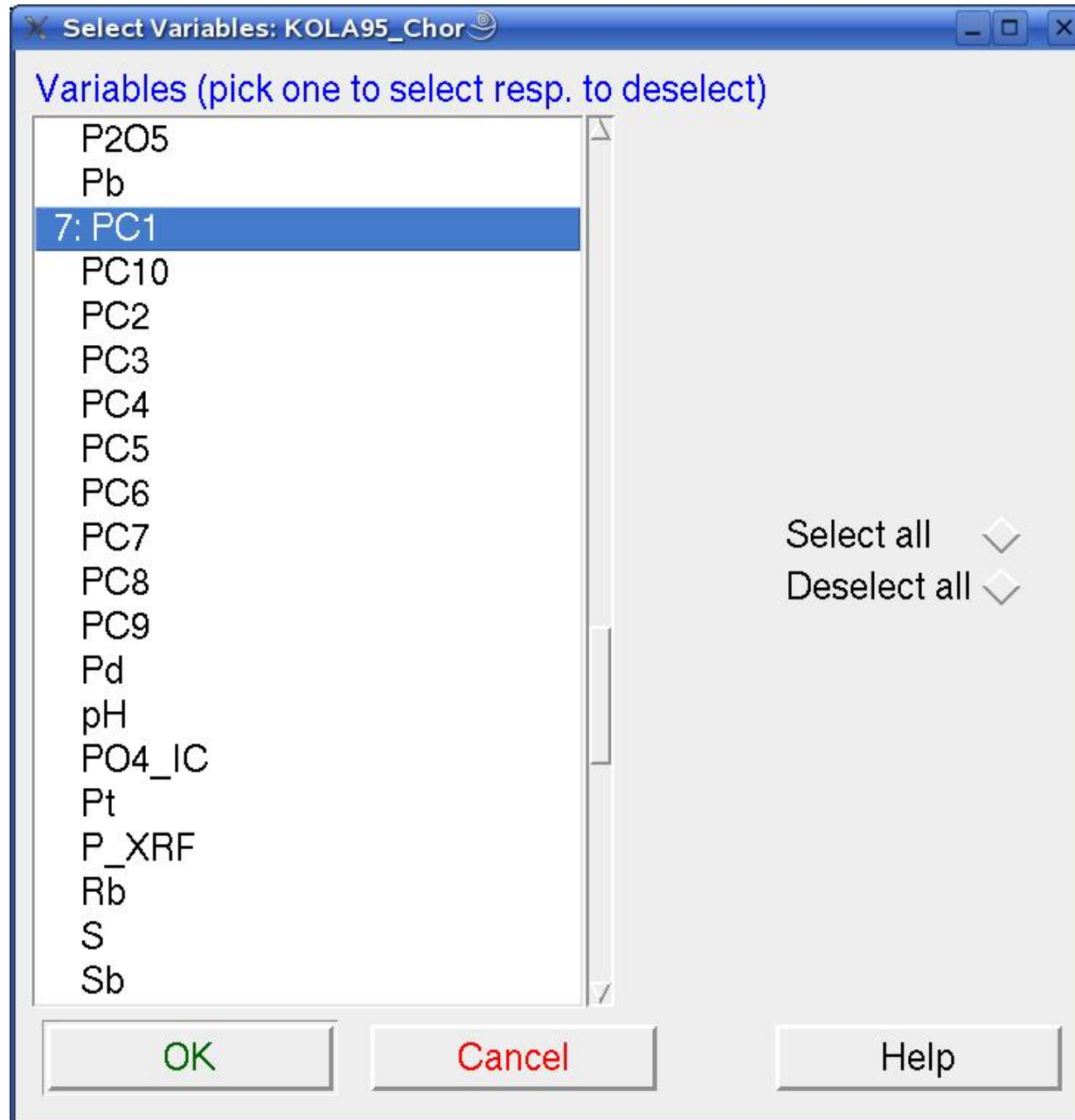
Plot Symbol Size:

Principal Components

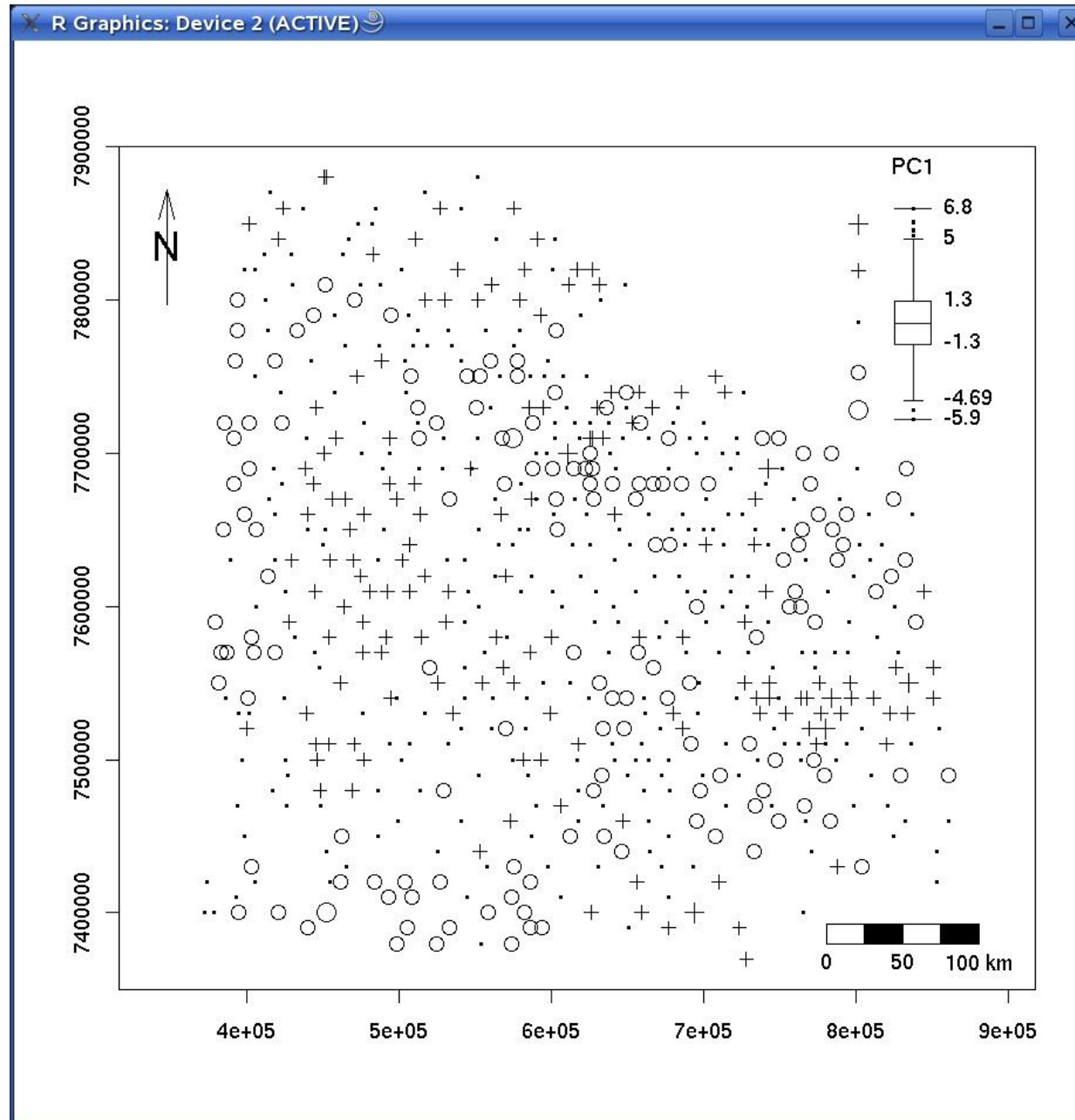




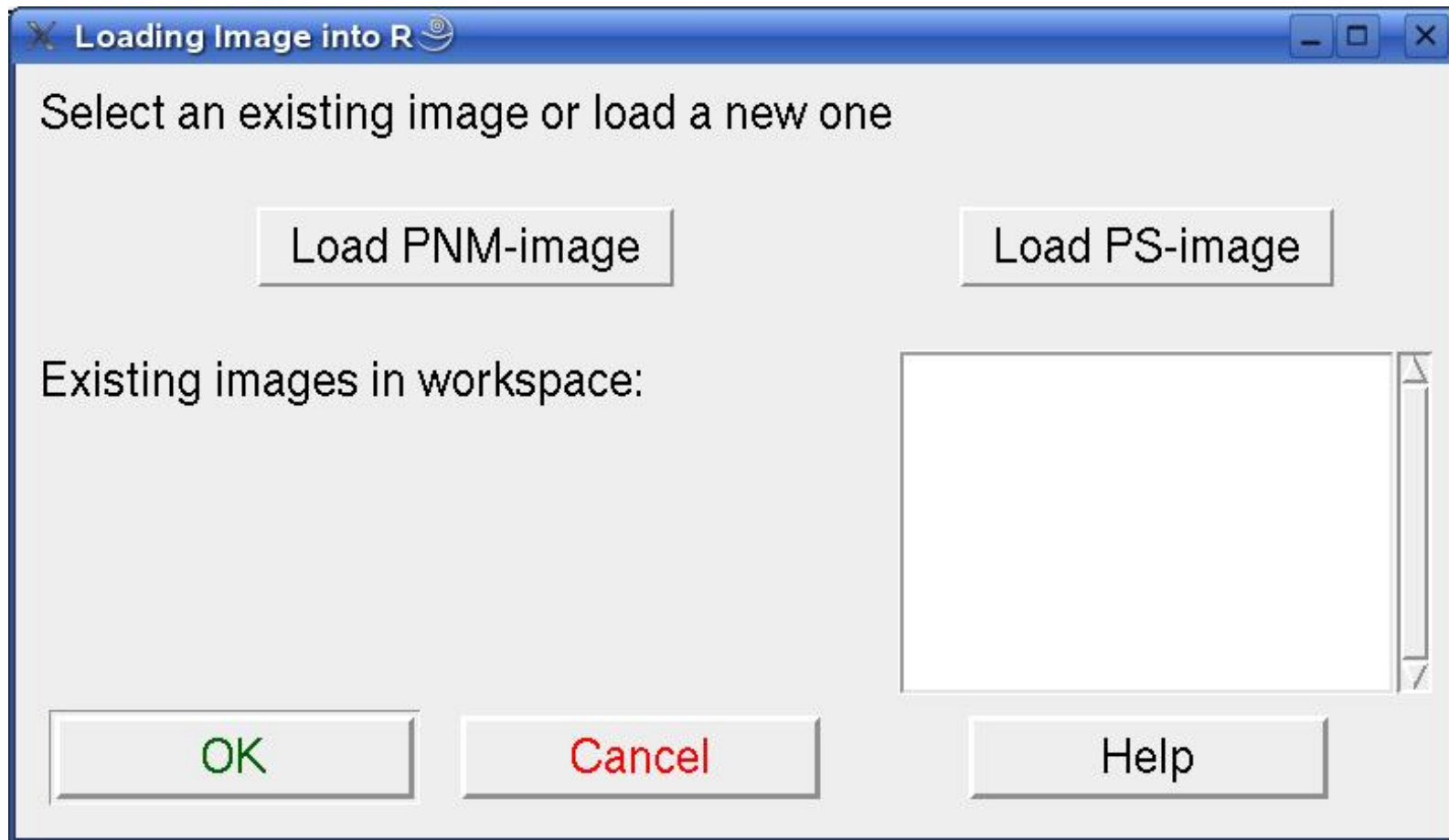
Principal Components: Mapping



Principal Components: Mapping



Background: Convert



Background: Convert

The screenshot shows the R Graphics interface with a map titled "R Graphics: Device 3 (ACTIVE)". The map displays a grid of reference points (black crosses) and a set of data points (blue crosses) overlaid on a geographical map. A dialog box titled "Selecting Reference Points: KOLA95_Chor" is open, providing instructions and options for the transformation.

Selecting Reference Points: KOLA95_Chor

Click for: Add/Identify a Reference Point
(First click=image, second click=data points)

Transformation Type:

- Similarity Transformation (min : 2)
- Affine Transformation (min : 3)
- Nonlinear Transformation (min : 4)
- Robust Similarity Transformation (min : 5)

Buttons: Reset, Save Background Map, OK, Cancel, Help

```

x= 2000 , y= 1000 , var=
ddle="50",max="100",un:
871910,347446,7834874)
tname="KOLA95_Chor",va:
EACHING/2004-05/proj_p:
5/proj_prak_05/schwind:
EACHING/2004-05/proj_p:

```

Background: Convert

Selecting Reference Points: KOLA95_Chor

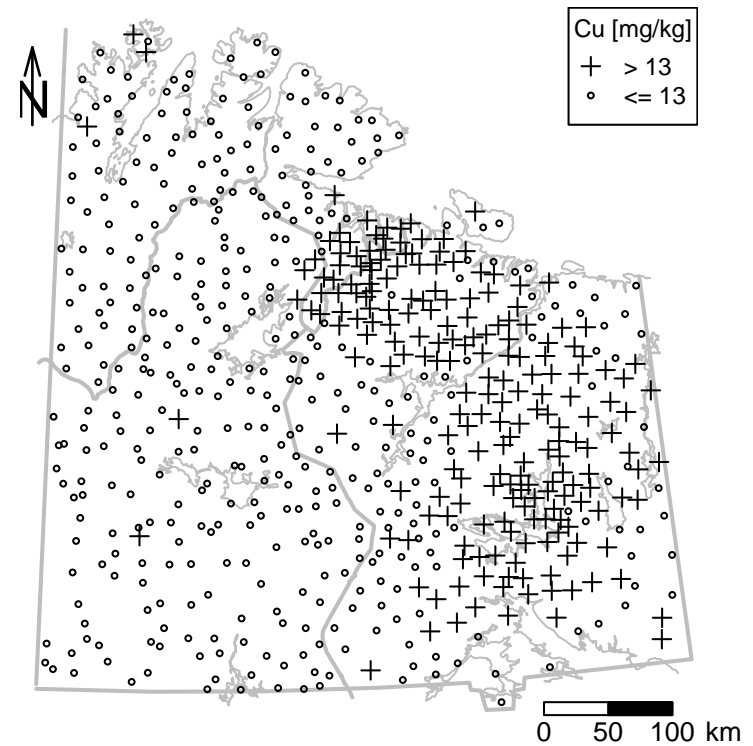
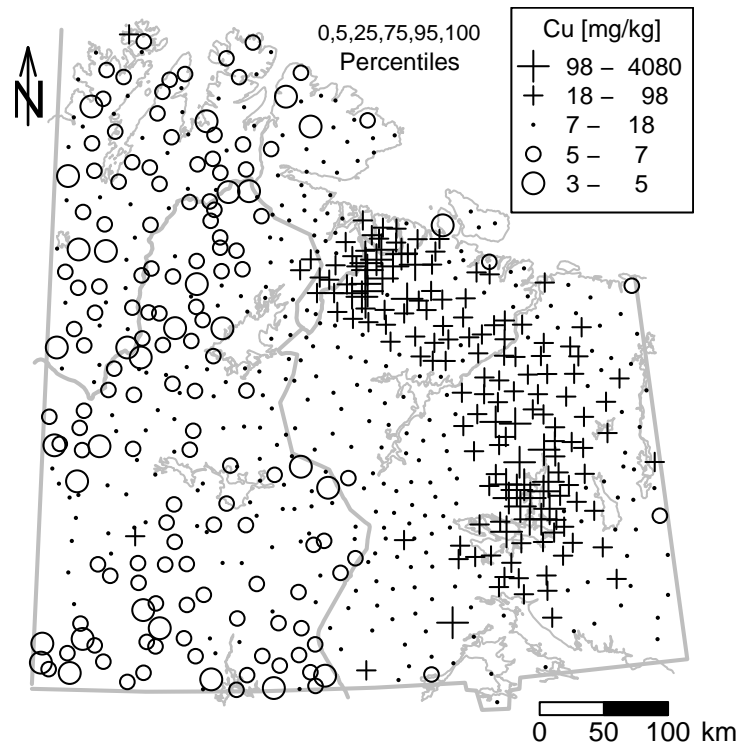
Click for: Add/Identify a Reference Point
(First click=image, second click=data points)

#	Image Coordinates		Data Coordinates		Delete	Move
	X	Y	X	Y		
1	1271.43	6009.21	401650	7850000	◇	◇
2	875.96	1168.07	372602	7400000	◇	◇
3	6565.86	1416.34	853146	7420000	◇	◇
4	6162.33	4312.74	833075	7690000	◇	◇

Transformation Type:

- Similarity Transformation (min : 2)
- Affine Transformation (min : 3)
- Nonlinear Transformation (min : 4)
- Robust Similarity Transformation (min : 5)

Result: Spatial Distrib. of Cu



- DAS+R is user friendly
- enables to construct complicated R-commands via the GUI
- high repeatability (fast prototyping)
- using of subsets (of variables/observations)
- emphasis on graphical analysis.
- Finally: Embedding of new functions should not be difficult!!!!

To do:

- Complete intended methodology
- Enter all the subsets functionalities in the procedures where applicable
- Remove most of the bugs
- Fill all help buttons
- Submit the package to CRAN.