

DAS+R: A graphical user interface for multivariate data analysis in R

R. Dutter, A. Zainzinger and A. Alfons

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

Chemometrics Workshop, Vienna
June 28, 2010



Vienna University of Technology

- 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

DAS+R(0.0-3)

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

Data set: **<No active dataset>** Edit data set View data set Model: **<No active model>**

Script Window

```
makeSubsetDAS(KOLA95_Chor, datname="KOLA95_Chor", subname="subn", ind=c(2,5,6,9,10,
deleteSubsetsDAS("KOLA95_Chor", "sub01")
boxplotDAS(datname="KOLA95_Chor", varnames=c("A1"), names=c("A1 [mg/kg] "))
|
```

Output Window Submit

```
> activeDataSet("KOLA95_Chor")

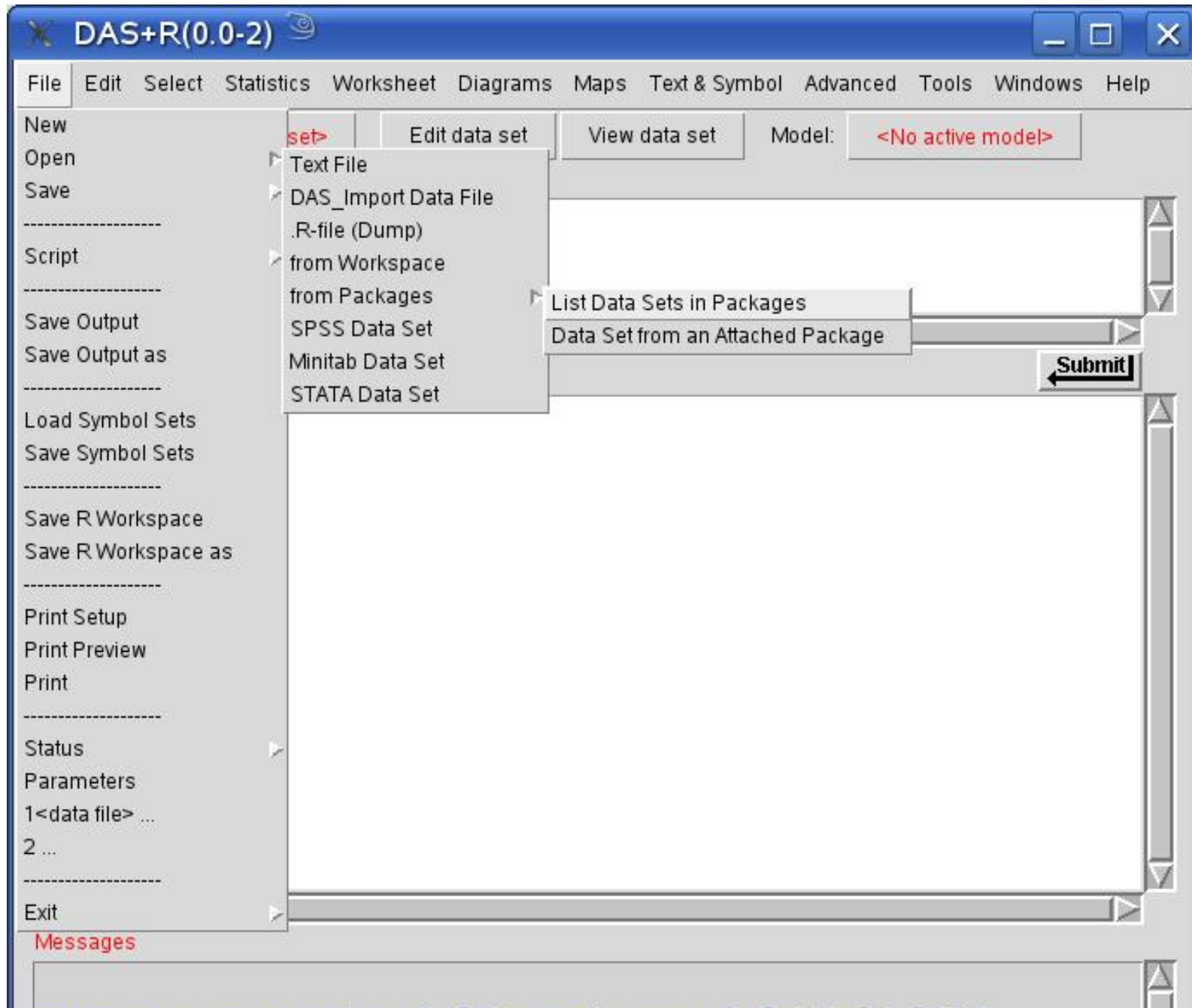
> makeSubsetDAS(KOLA95_Chor, datname="KOLA95_Chor", subname="subn", ind=c(2,5,6,9,1
[1] TRUE

> deleteSubsetsDAS("KOLA95_Chor", "sub01")
Subset sub01 successfully deleted!
[1] TRUE

> boxplotDAS(datname="KOLA95_Chor", varnames=c("A1"), names=c("A1 [mg/kg] "))
```

Messages

NOTE: DAS+R Version 0.0-3: Wed Feb 24 14:00:19 2010



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

Enter name for data set: KOLA95_C2MM

Missing data indicator: NA

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 (NOR))

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

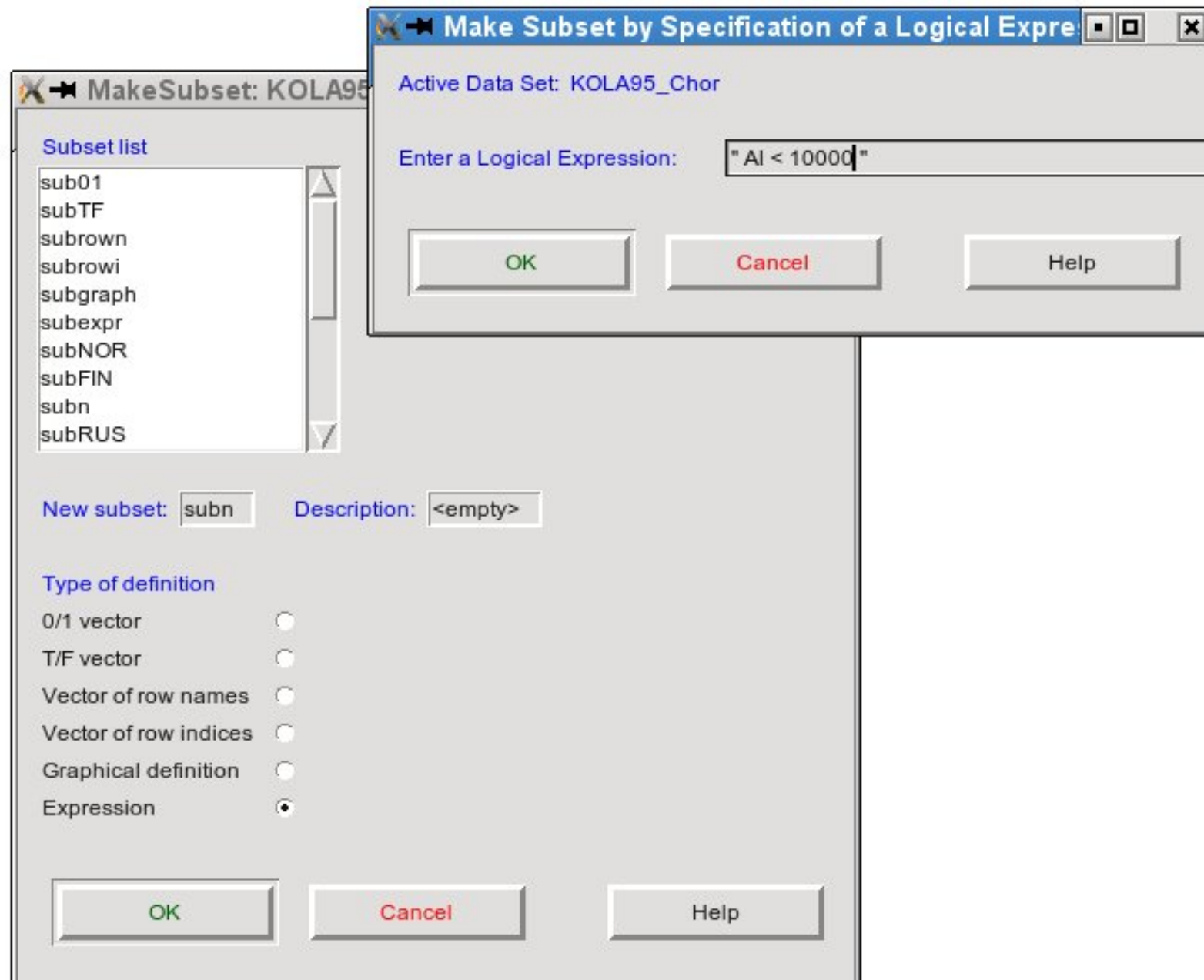
UNIT m east m north m a.s.l. cm mg/kg

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	

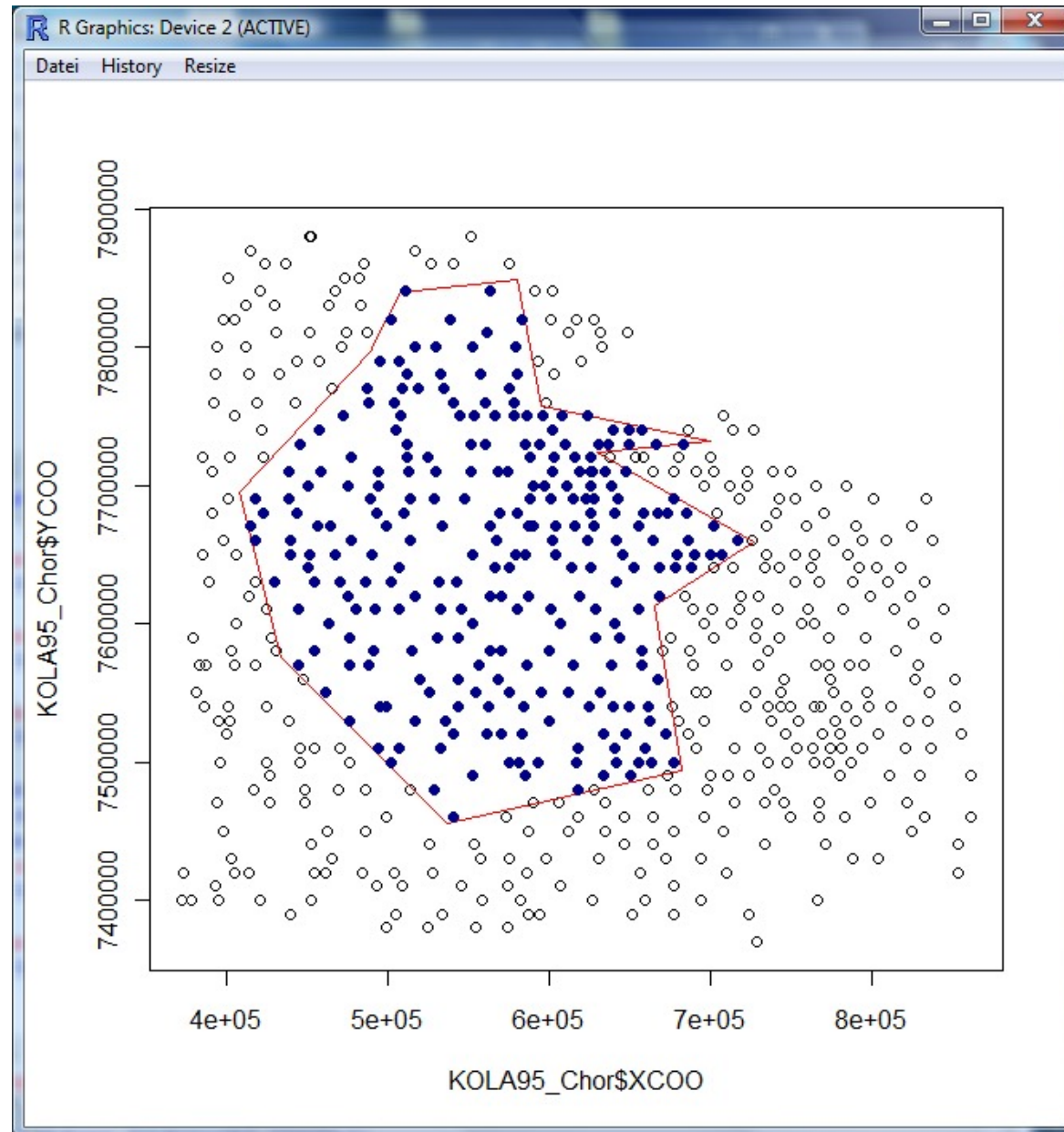
OK Cancel Help

5

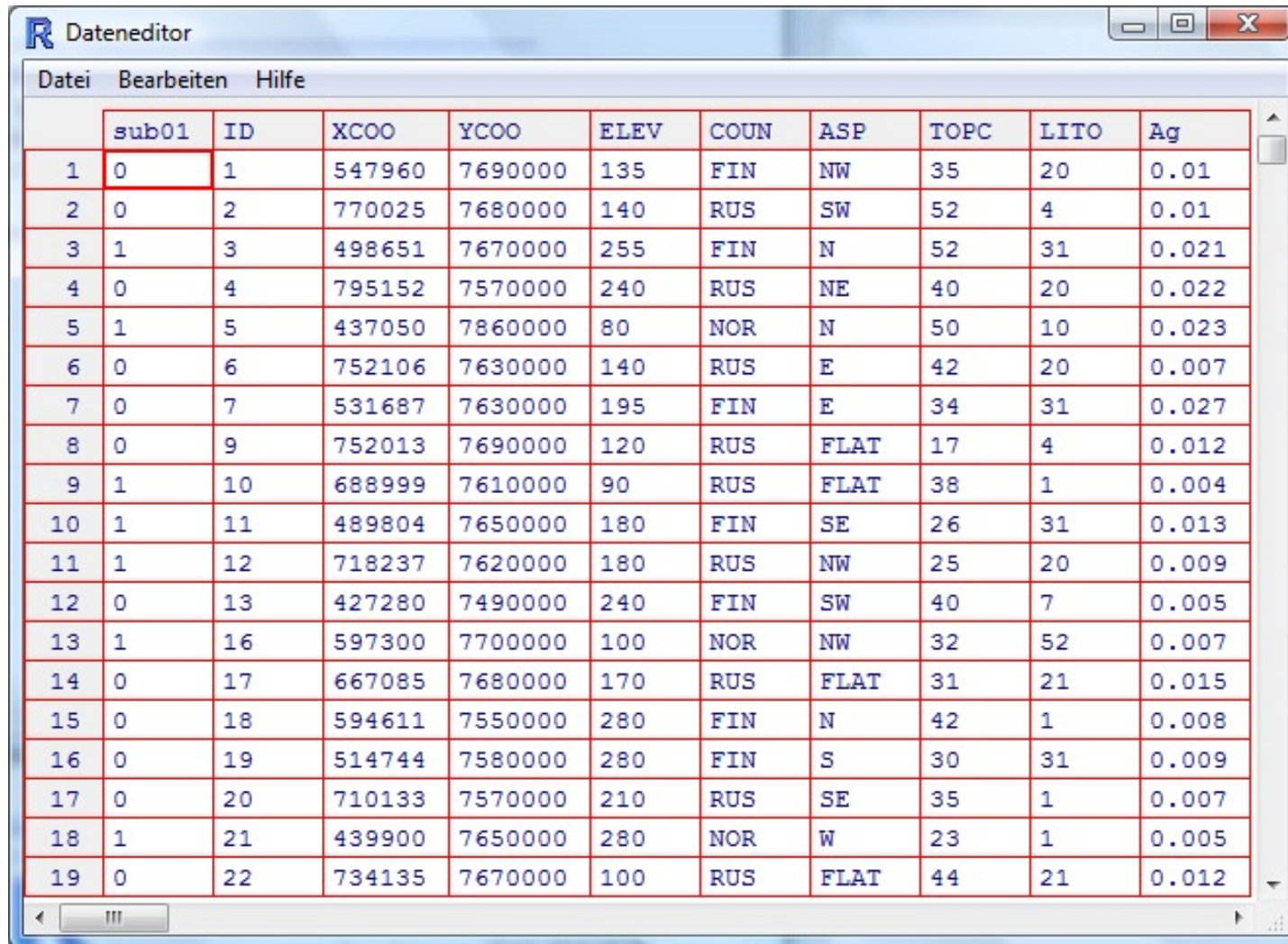
Make Subset: Formula



Make Subset: Graphical definition



Make Subset: Data editor



R Dateneditor

Datei Bearbeiten Hilfe

	sub01	ID	XCOO	YCOO	ELEV	COUN	ASP	TOPC	LITO	Ag
1	0	1	547960	7690000	135	FIN	NW	35	20	0.01
2	0	2	770025	7680000	140	RUS	SW	52	4	0.01
3	1	3	498651	7670000	255	FIN	N	52	31	0.021
4	0	4	795152	7570000	240	RUS	NE	40	20	0.022
5	1	5	437050	7860000	80	NOR	N	50	10	0.023
6	0	6	752106	7630000	140	RUS	E	42	20	0.007
7	0	7	531687	7630000	195	FIN	E	34	31	0.027
8	0	9	752013	7690000	120	RUS	FLAT	17	4	0.012
9	1	10	688999	7610000	90	RUS	FLAT	38	1	0.004
10	1	11	489804	7650000	180	FIN	SE	26	31	0.013
11	1	12	718237	7620000	180	RUS	NW	25	20	0.009
12	0	13	427280	7490000	240	FIN	SW	40	7	0.005
13	1	16	597300	7700000	100	NOR	NW	32	52	0.007
14	0	17	667085	7680000	170	RUS	FLAT	31	21	0.015
15	0	18	594611	7550000	280	FIN	N	42	1	0.008
16	0	19	514744	7580000	280	FIN	S	30	31	0.009
17	0	20	710133	7570000	210	RUS	SE	35	1	0.007
18	1	21	439900	7650000	280	NOR	W	23	1	0.005
19	0	22	734135	7670000	100	RUS	FLAT	44	21	0.012

(Pre-)Select Variables

File Edit Select Statistics Worksheet Diagrams Maps Text & Symbol Ad

Data set:

Script Win

```
invis  
invis  
summa  
f.hist  
makeSu  
makeSu  
select
```


Output Win


```
[1] "1  
  
> make  
Subset  
  
> make
```

Select Variables: KOLA95_Chor

Variables (pick one to select resp. to deselect)

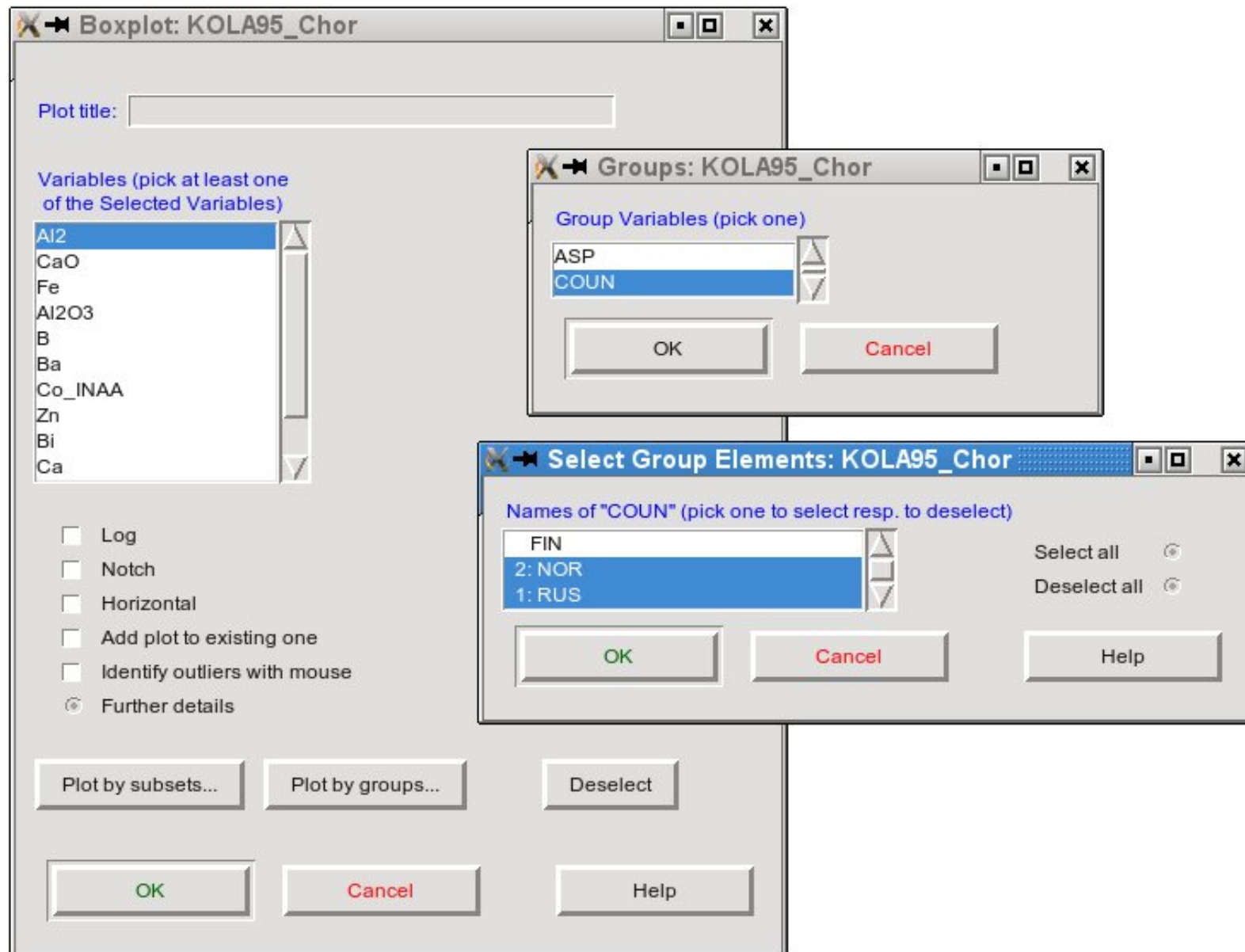
- 1: Ag
Ag_INAA
- 2: Al
Al2O3
Al_XRF
- 5: As
As_INAA
ASP
- 4: Au
Au_INAA
B
- 3: Ba
Ba_INAA
Be
Bi
Br_IC
Br_INAA
Ca
Ca_INAA
CaO

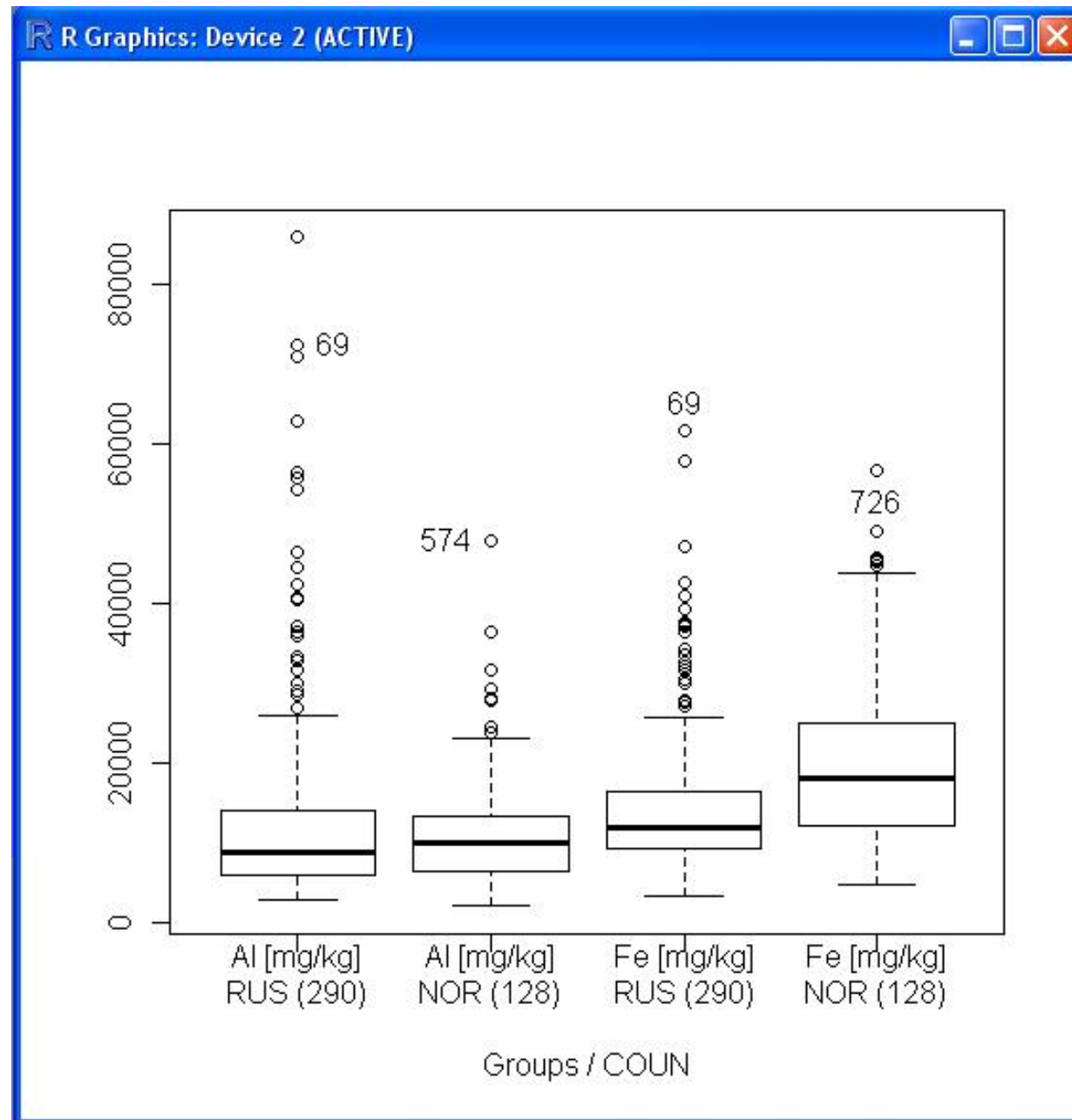
Select all 

Deselect all 

- Numerically
- Graphically
 - Boxplot
 - Density Trace
 - Xy-plot
 - Ternary Plot
 - Histogram
 - Scattermatrix (Draftman)

Copy Paste Quit										
	ID	XC00	YC00	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	NOR	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	NOR	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	NOR	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





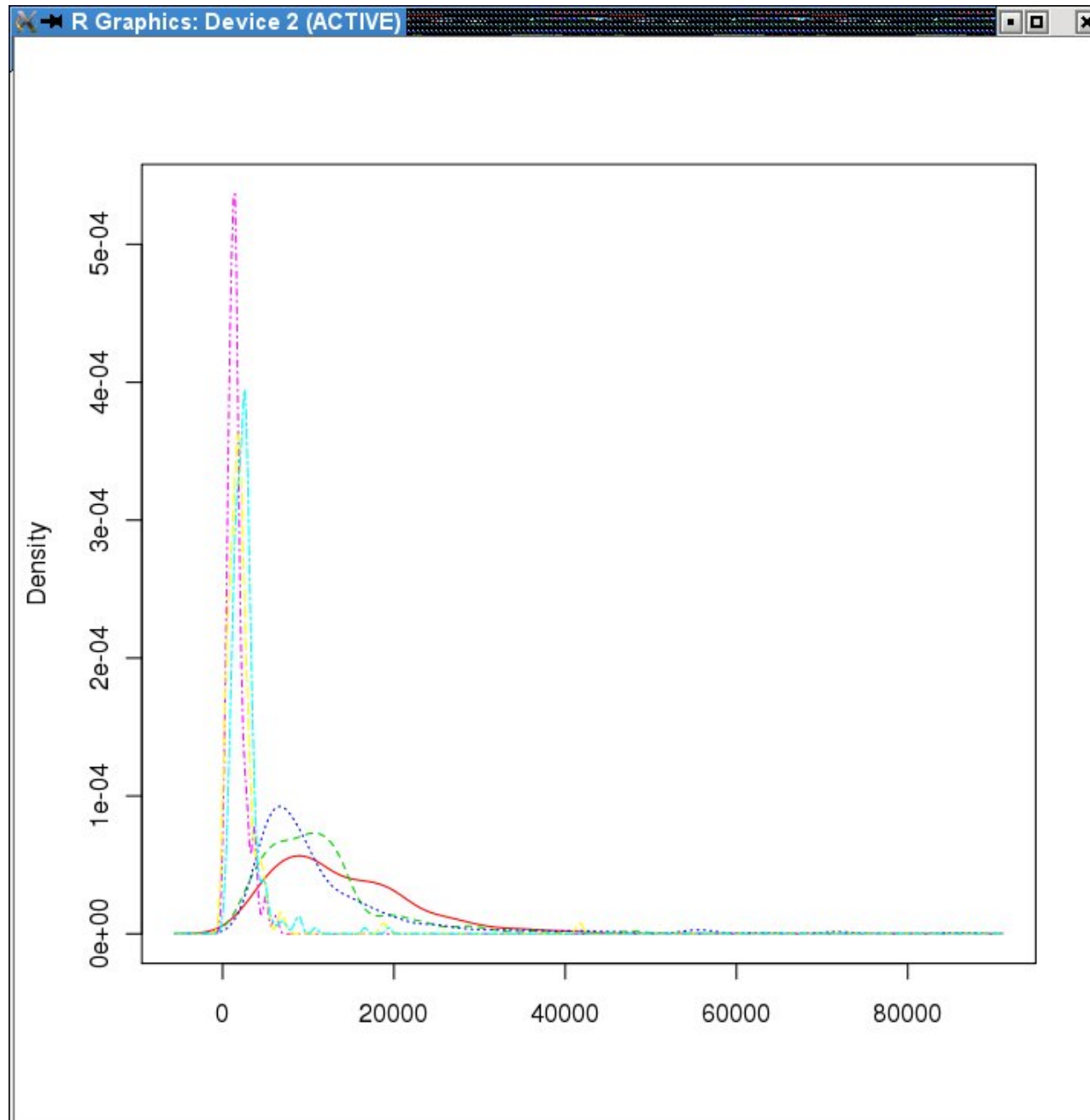
Plot title:

Variables (pick at least one of the Selected Variables)

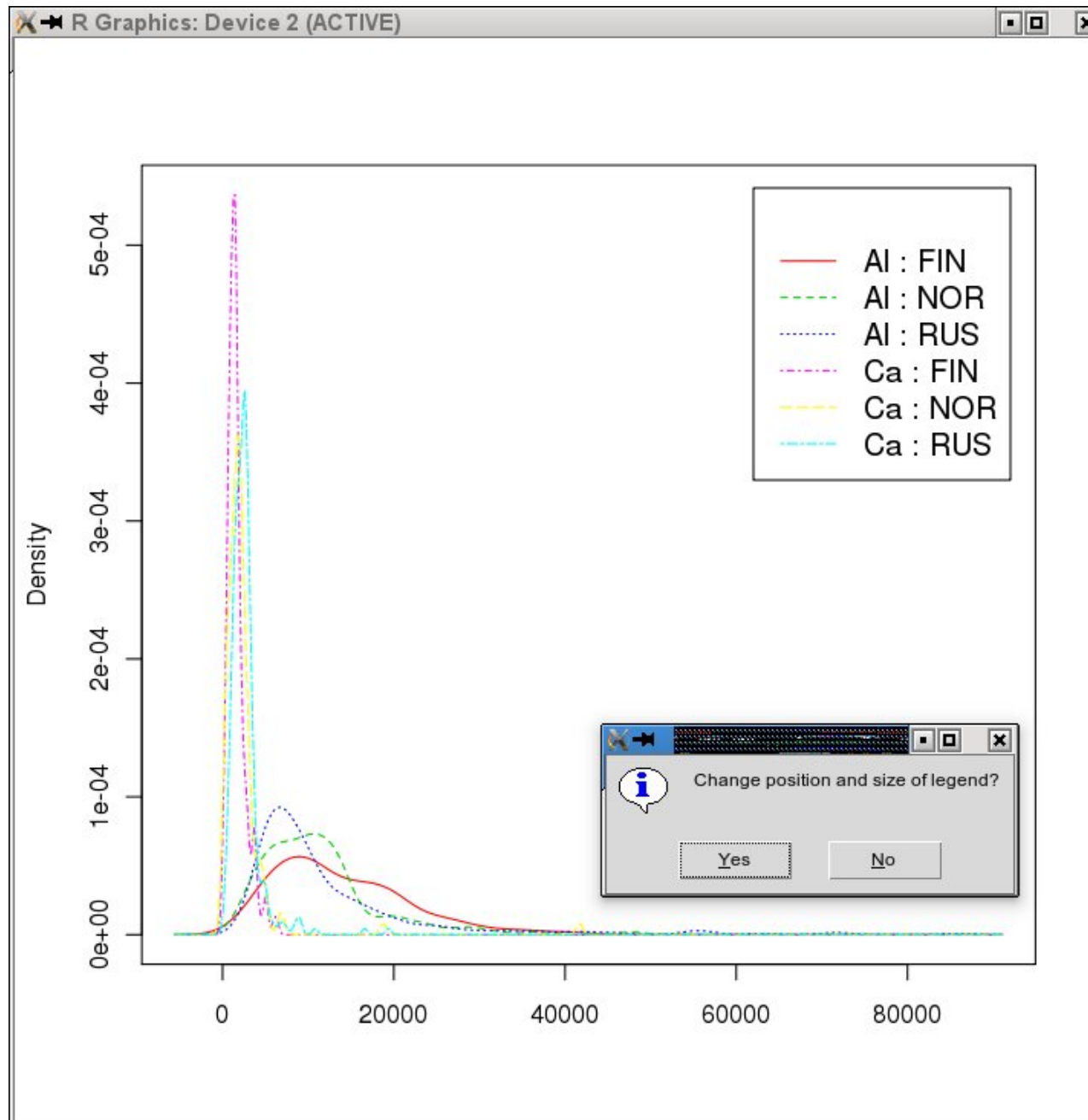
Al
CaO
Fe
Al ₂ O ₃
B
Ba
Co_INAA
Zn
Bi
Ca

☐ Log-scale

☒ x-Axis
☒ Lines
☐ Legend ON
☒ Additional
☒ Generate new plot







Plot title:

X-variable (pick one of the Selected Variables)

- Al2
- CaO
- Fe
- Al2O3
- B
- Ba
- Co_INAA
- Zn
- Bi
- Ca

☐ Log-scale

Y-variable (pick one of the Selected Variables)

- Al2
- CaO
- Fe
- Al2O3
- B
- Ba
- Co_INAA
- Zn
- Bi
- Ca

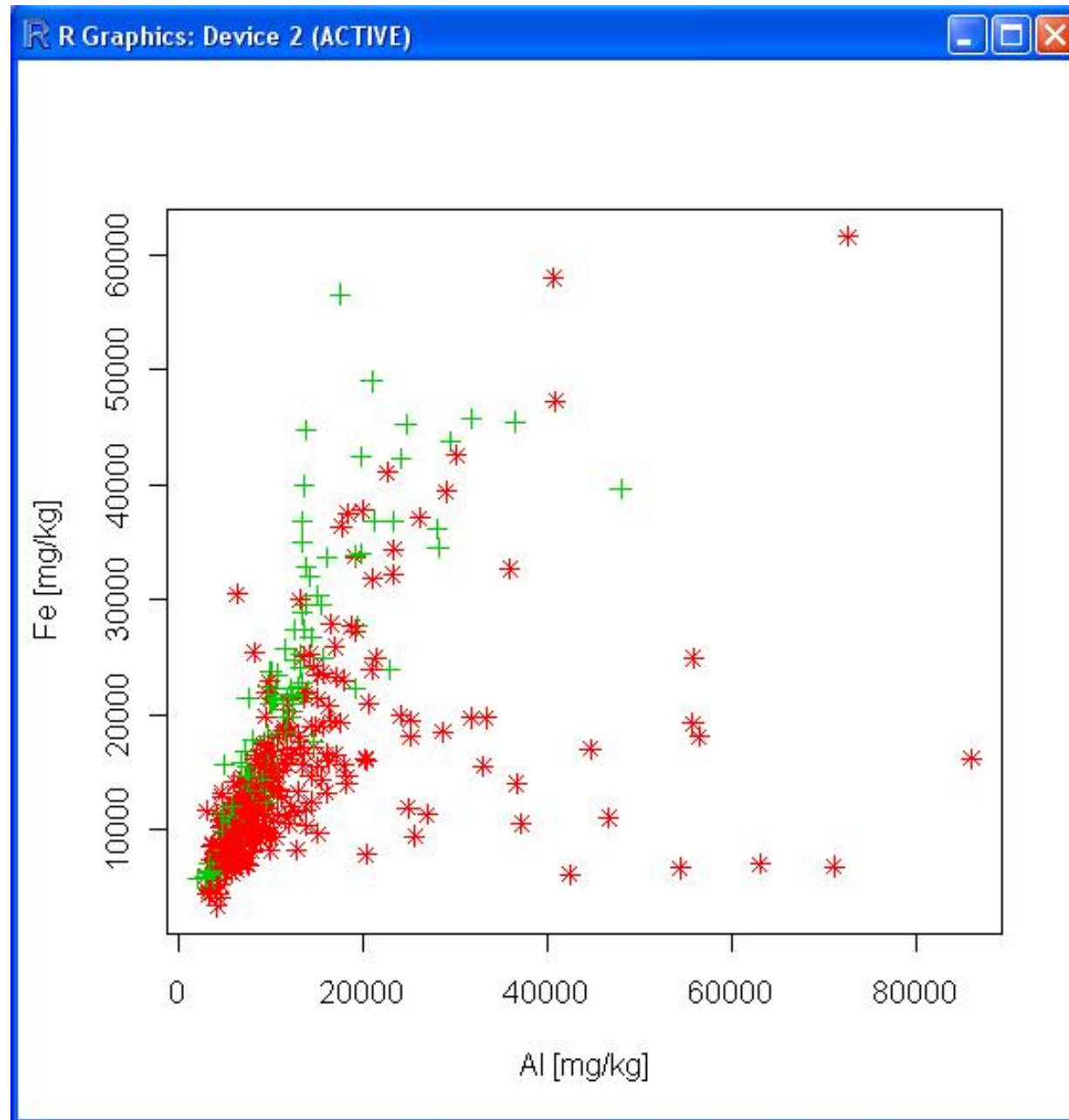
☐ Log-scale

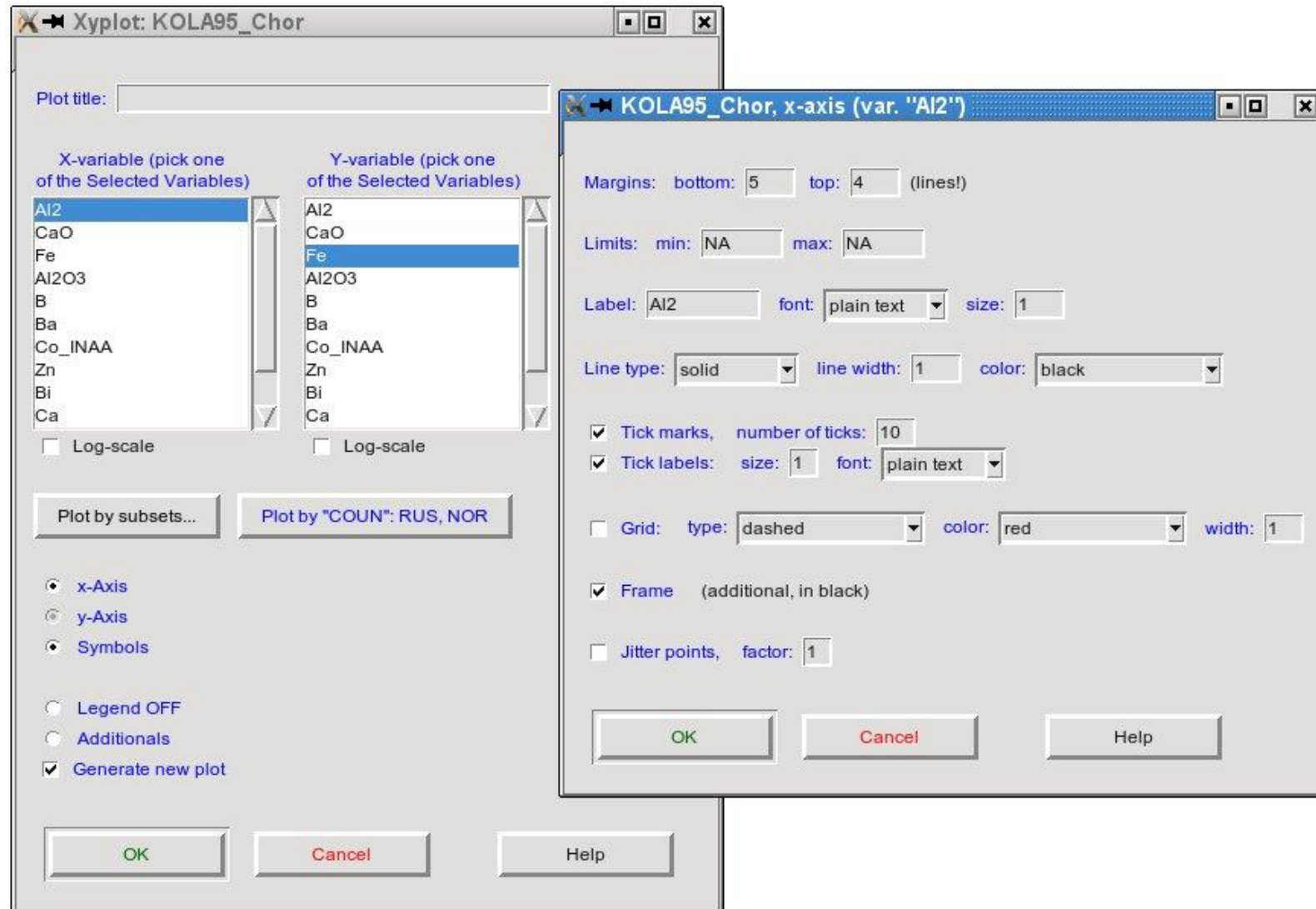
Plot by subsets... Plot by "COUN": RUS, NOR Deselect

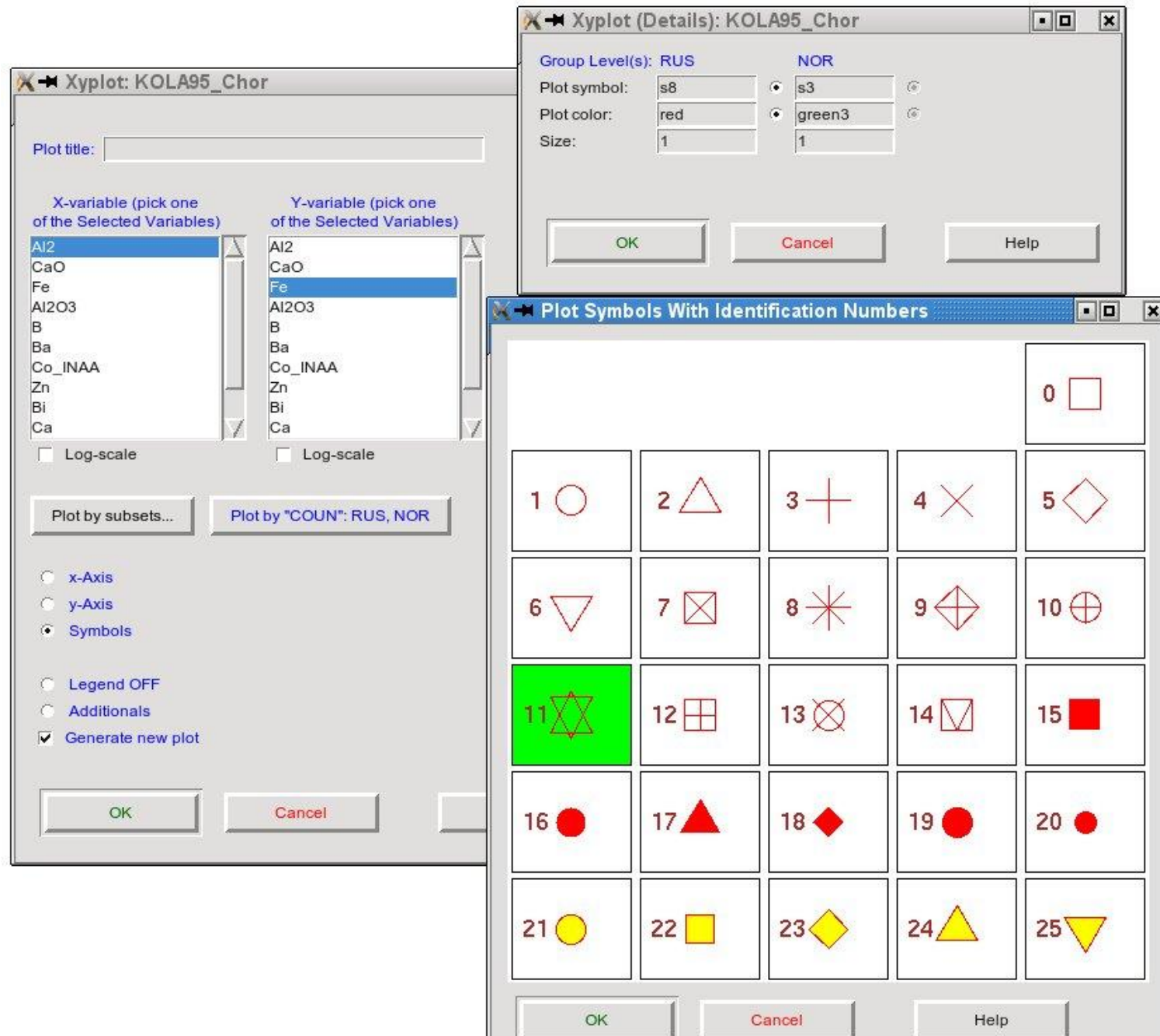
☒ x-Axis
☐ y-Axis
☒ Symbols

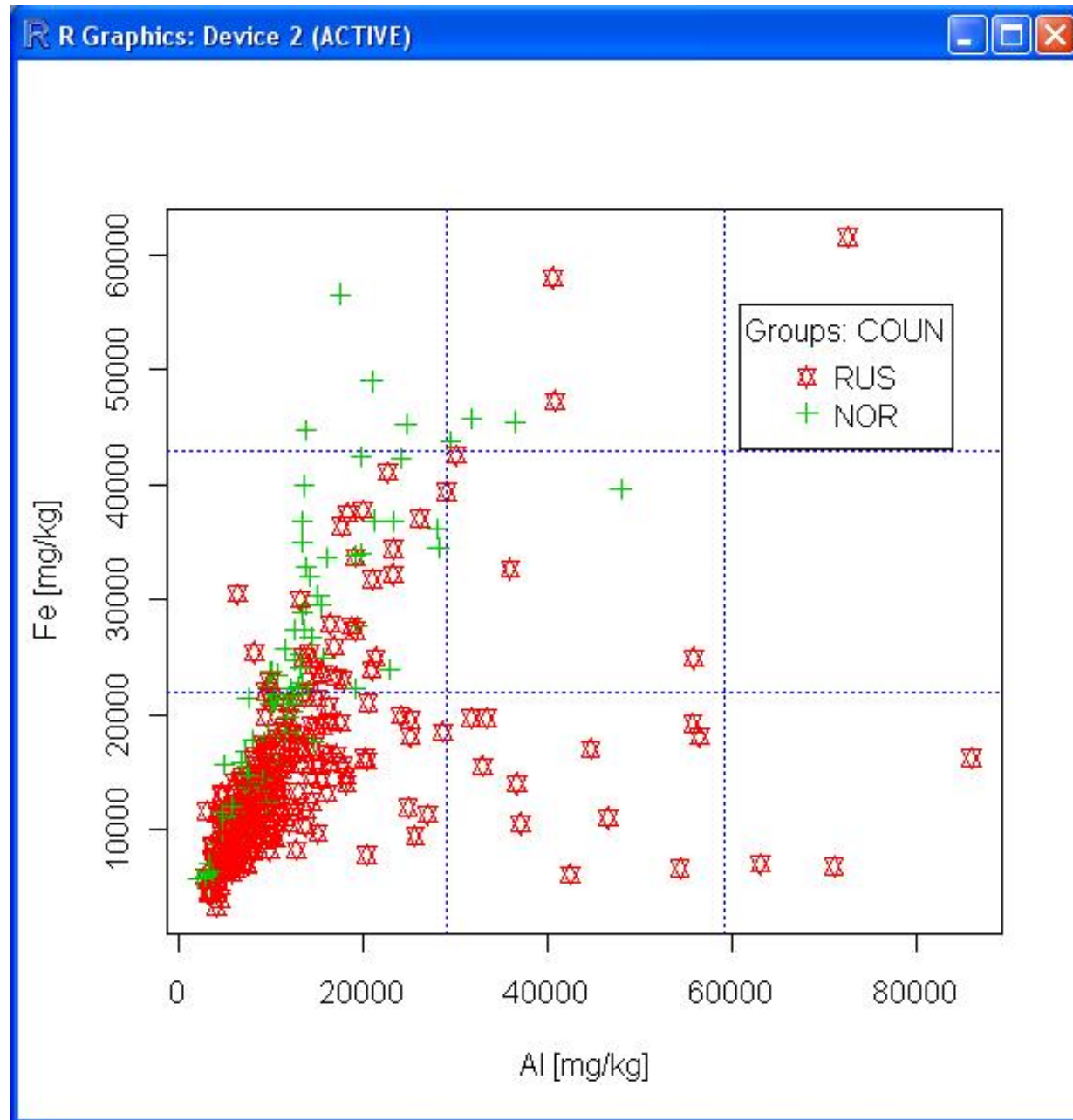
☐ Legend OFF
☐ Additional
☒ Generate new plot

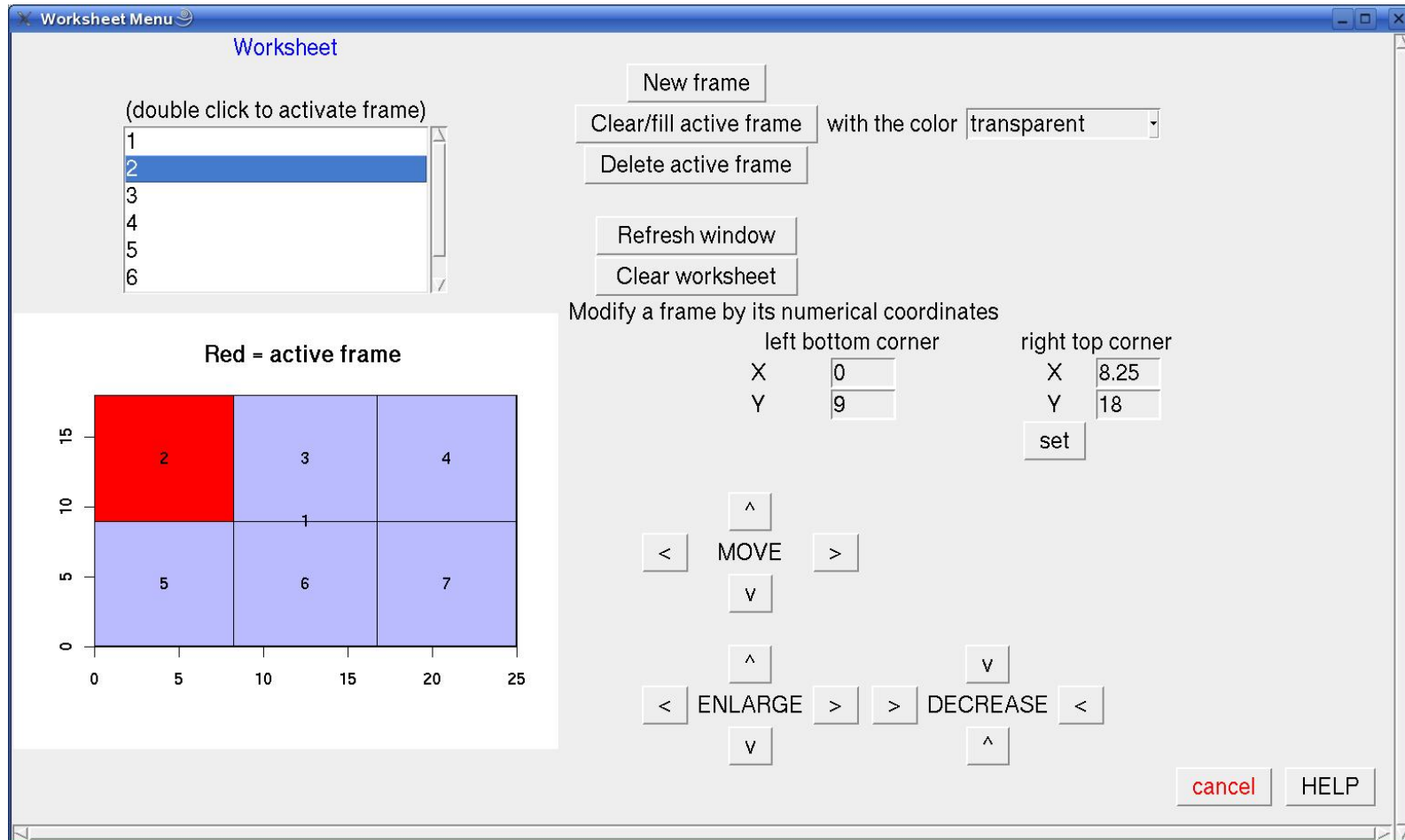
OK Cancel Help

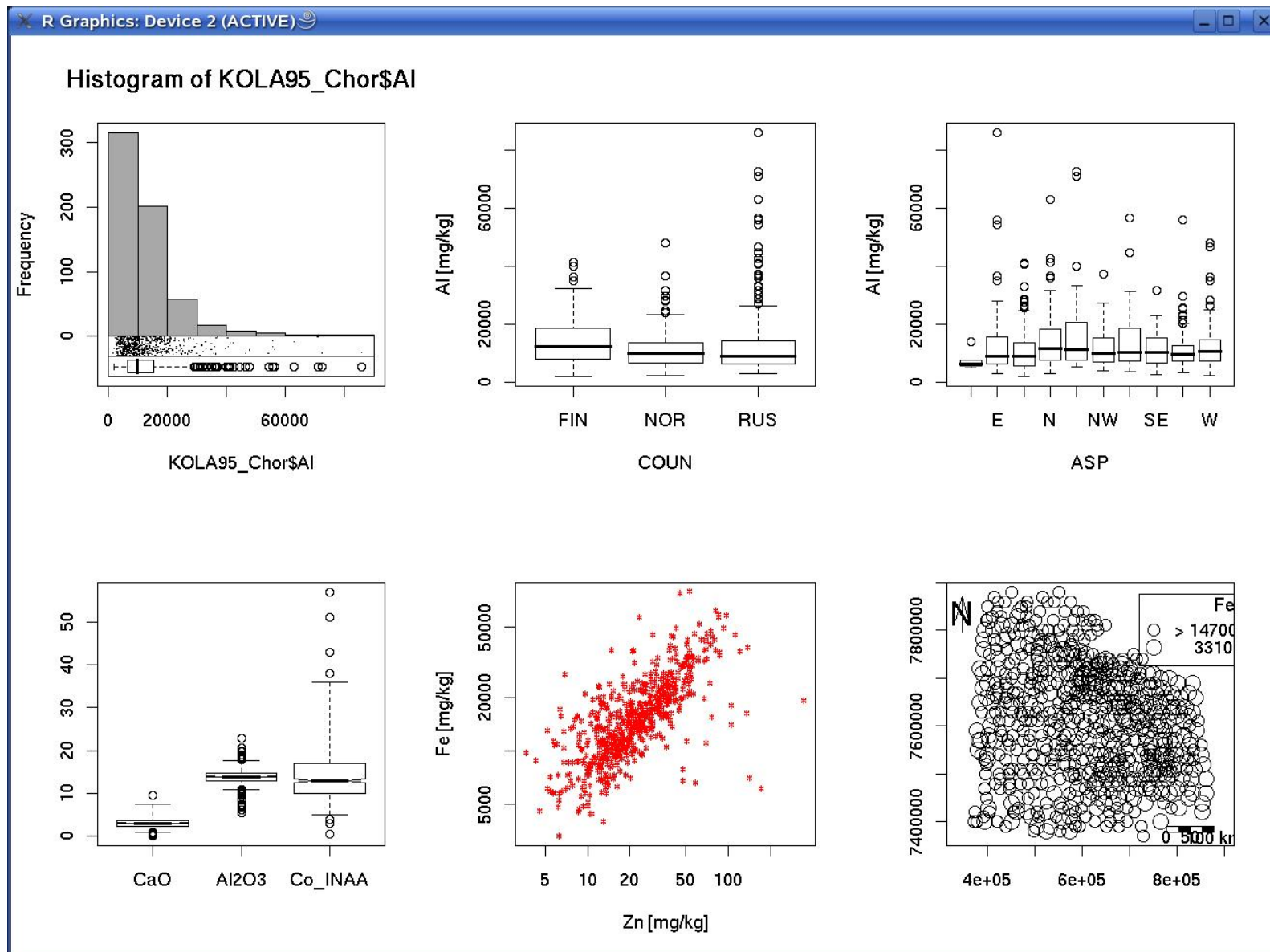




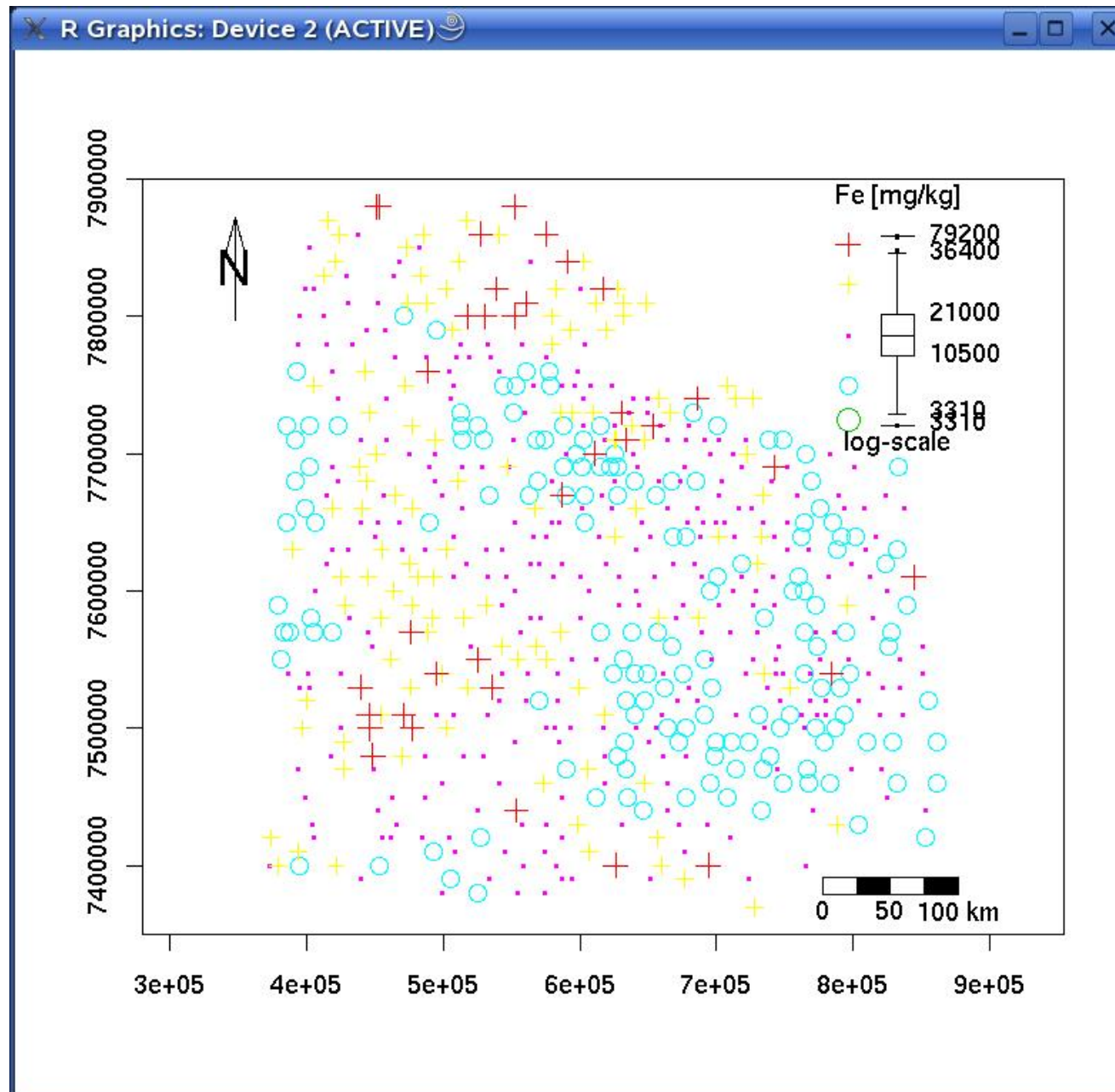




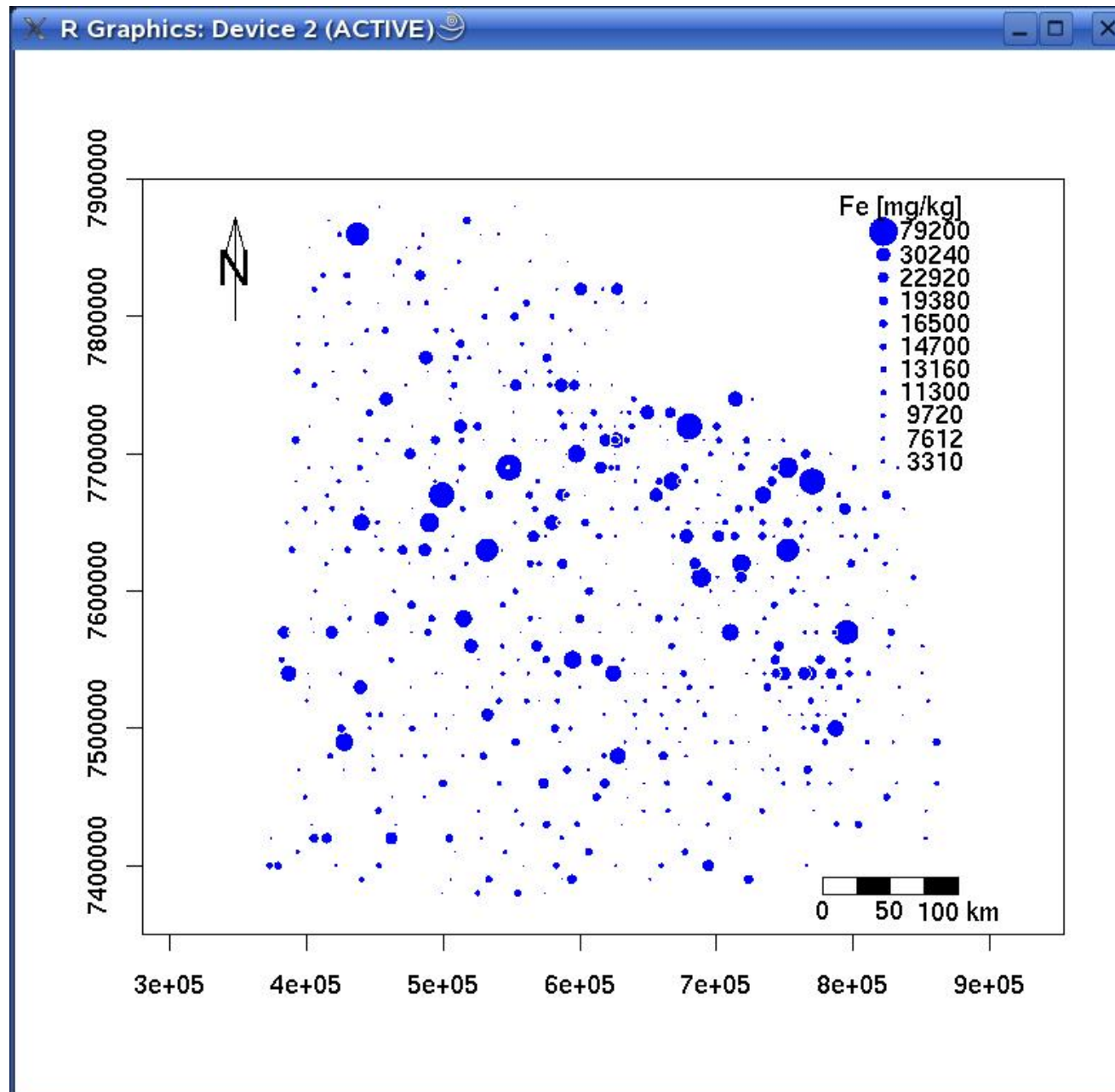


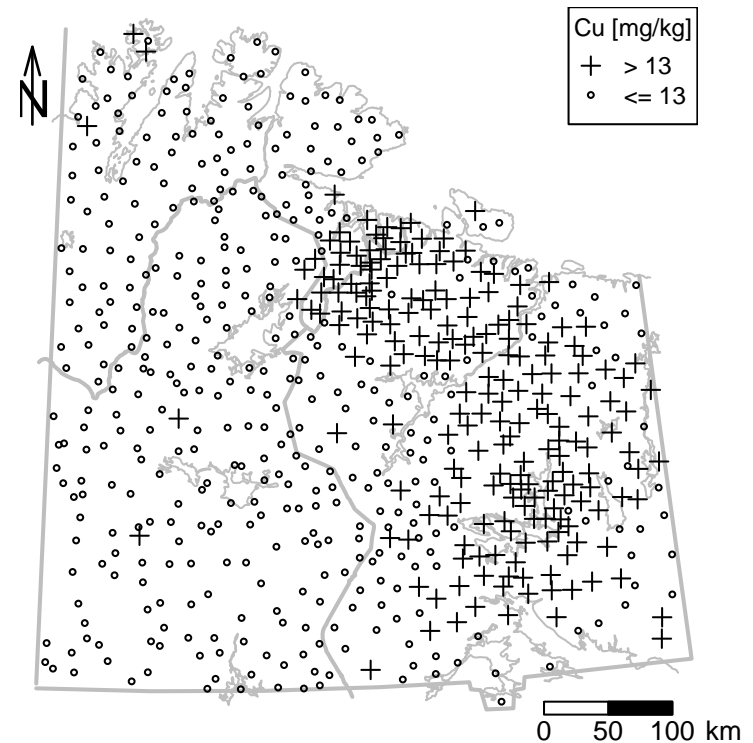
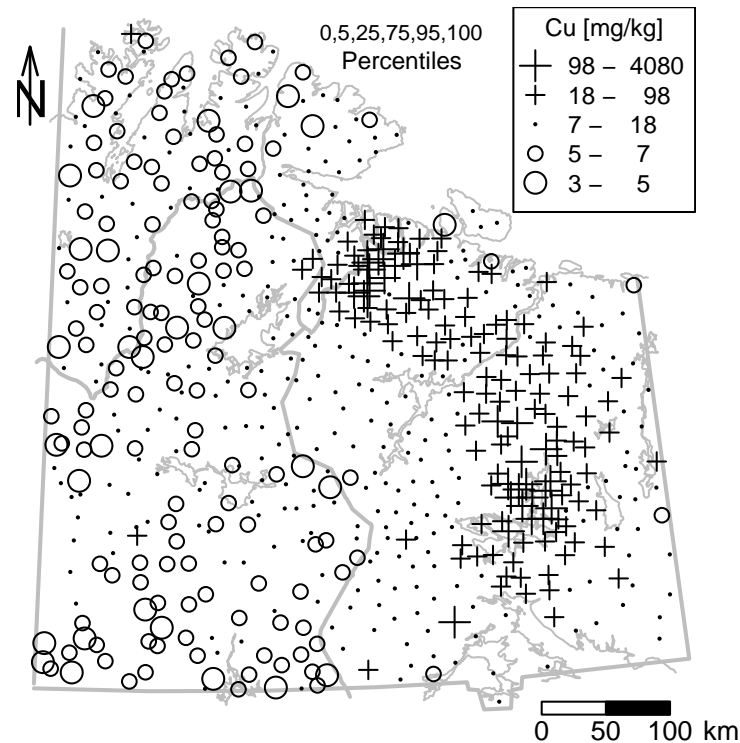


Mapping: Boxplot



Mapping: Proportional Plot





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

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

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)

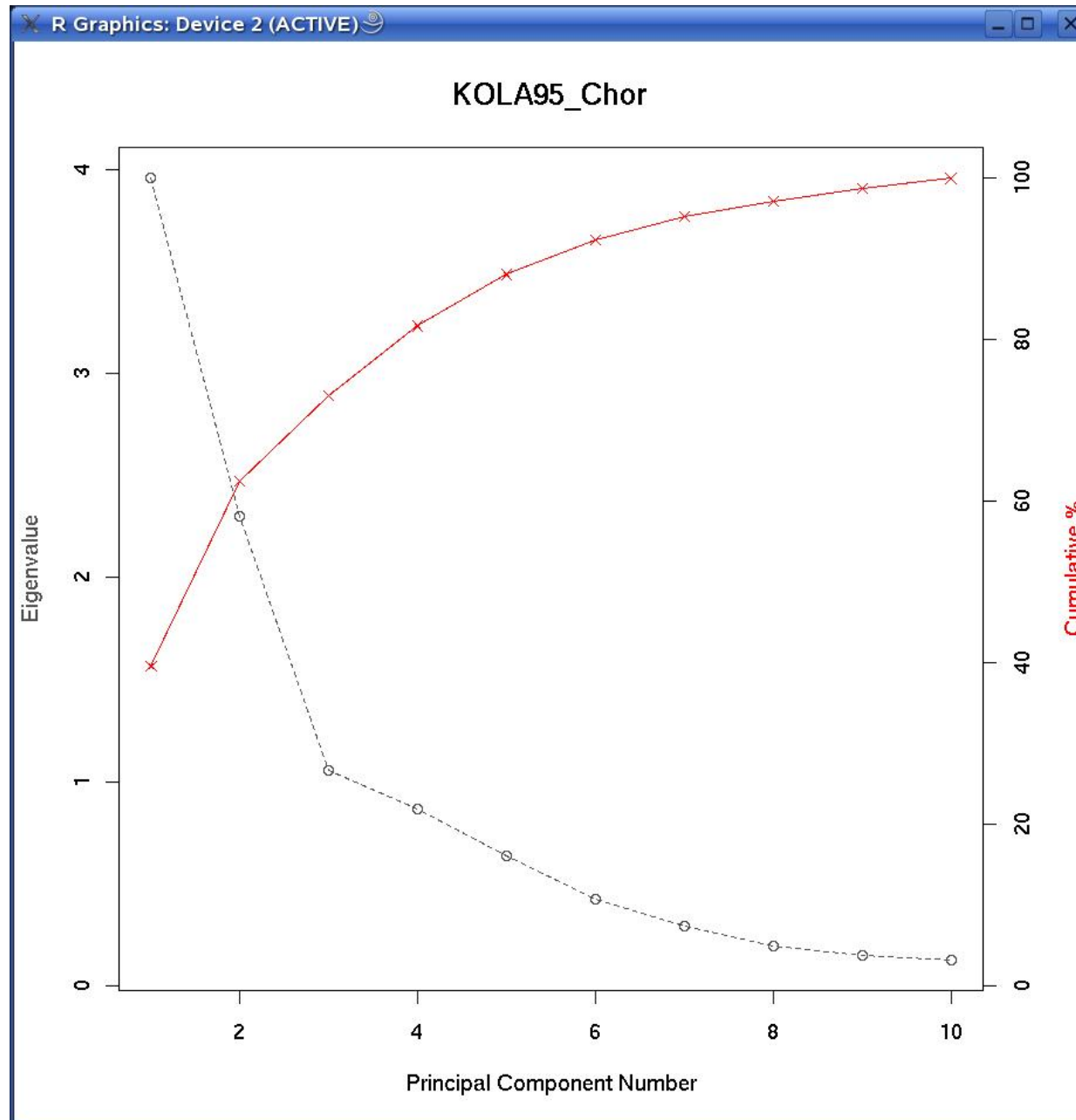
Plot Symbol Size: 1

Line Type: solid

Line Width: 1

Default

OK Cancel Help



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:

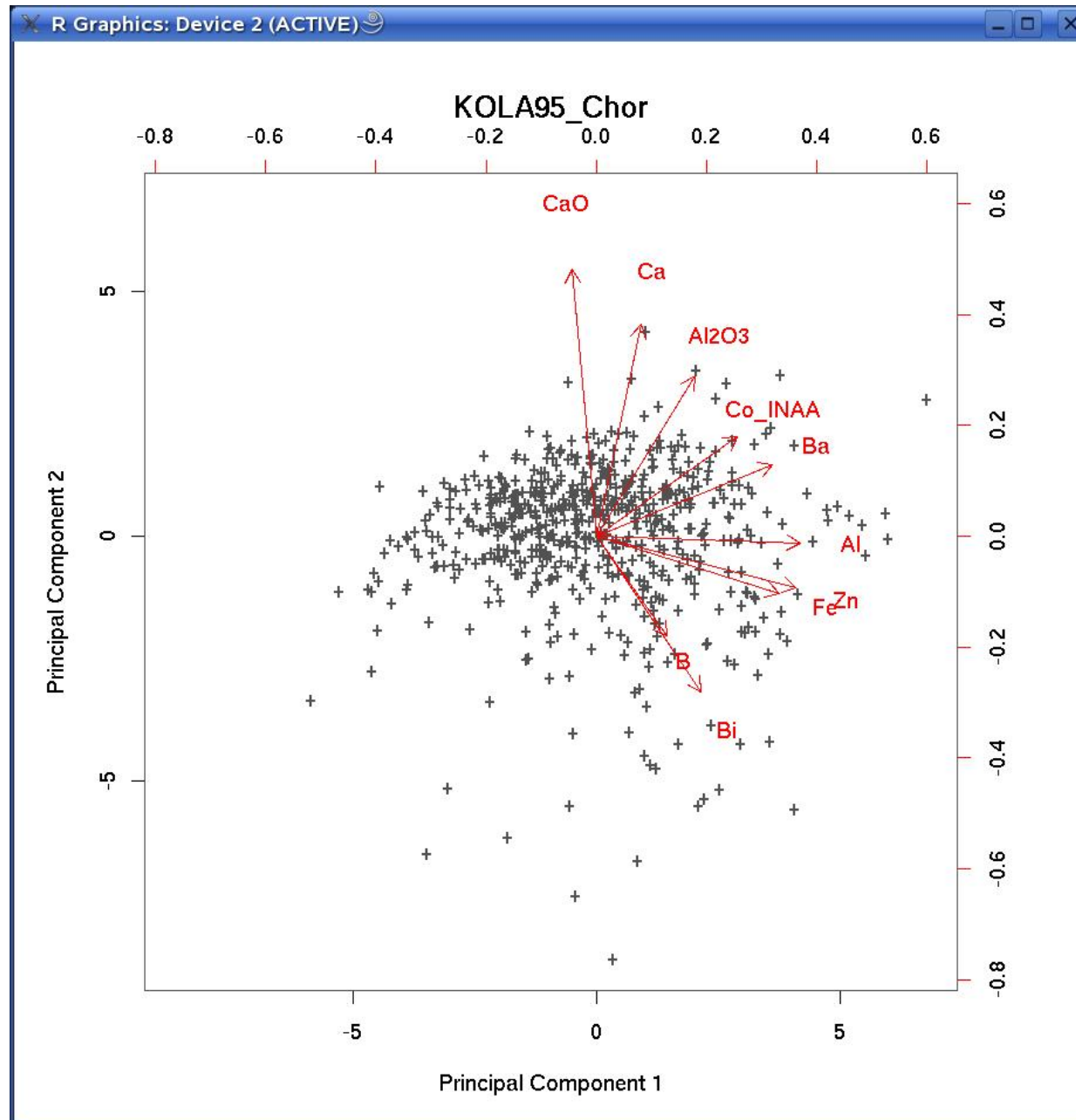
Plot String Size:

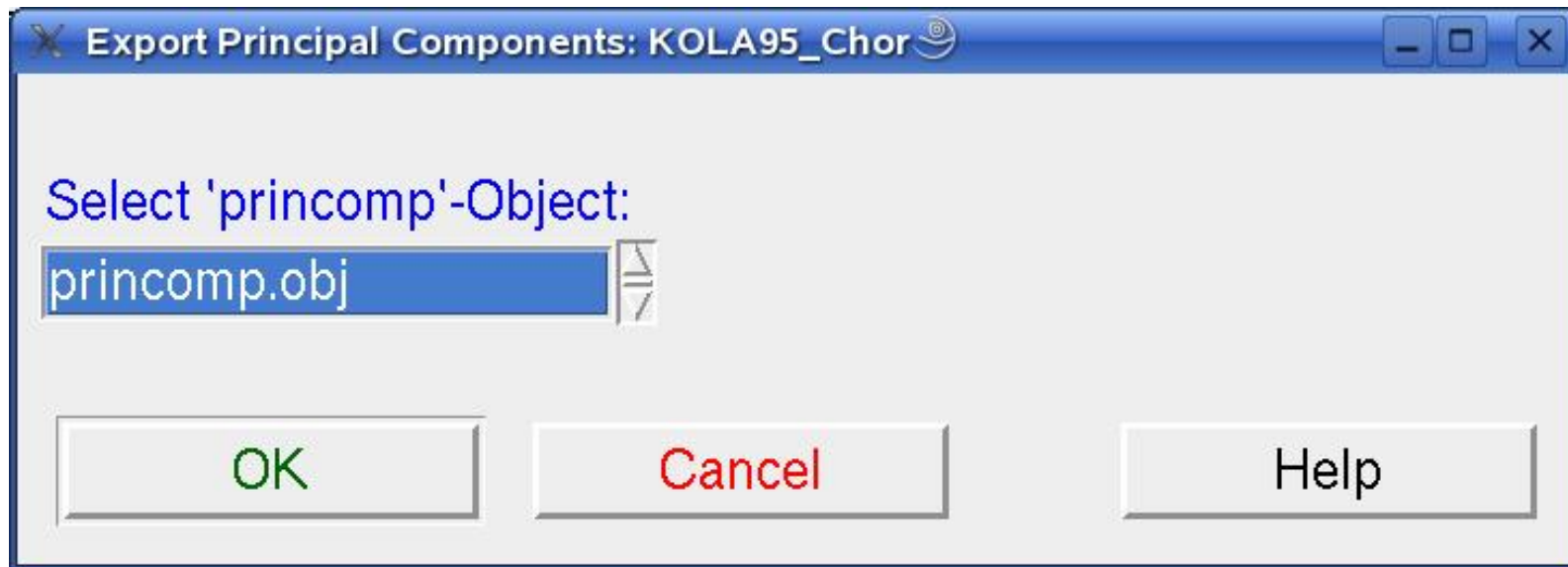
Loadings Details

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

Arrow Head Length:

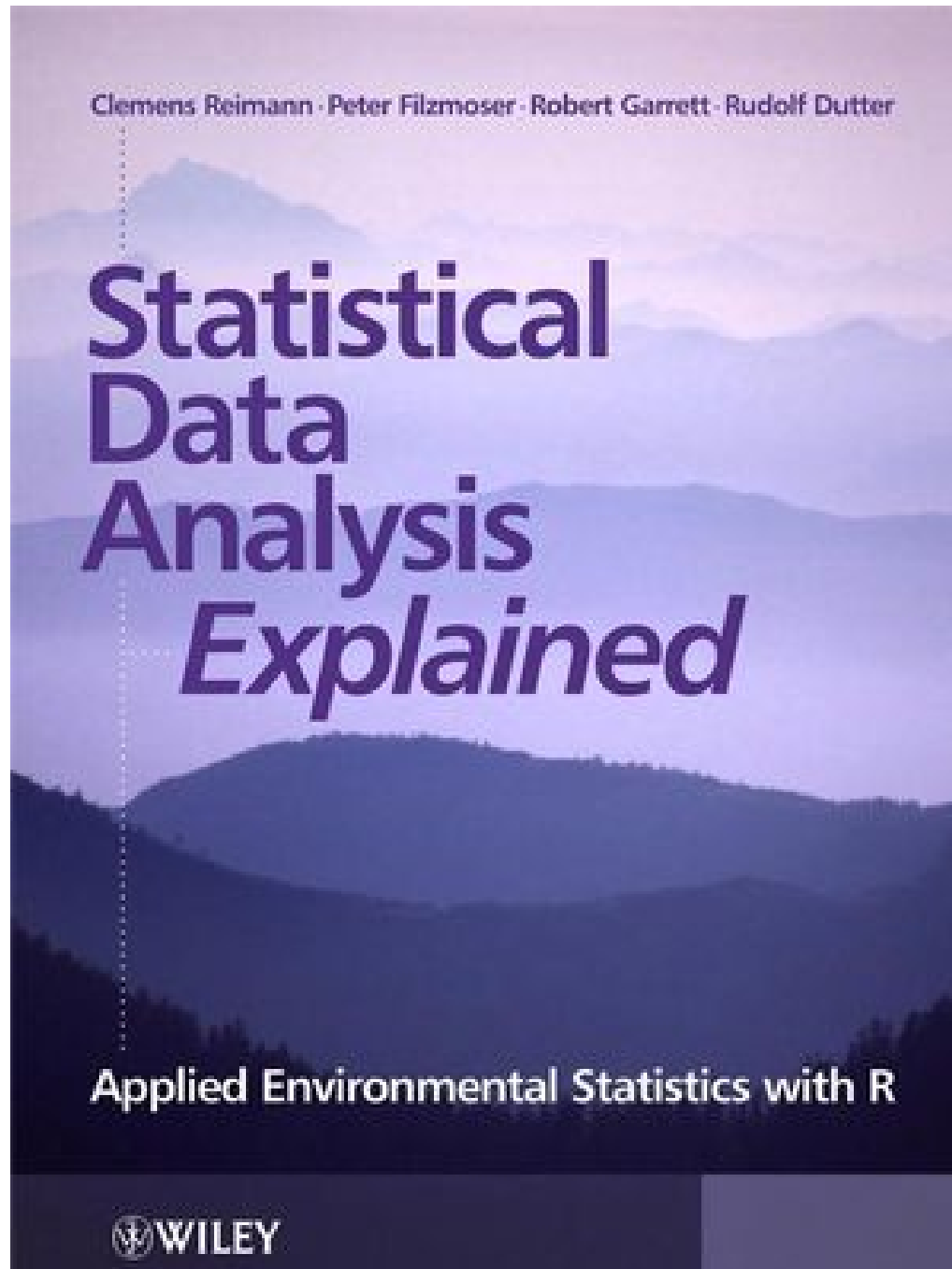
Plot Symbol Size:





- 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!!!!

- See <http://www.statistik.tuwien.ac.at/StatDA/DASplusR>



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

Emphasis:

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