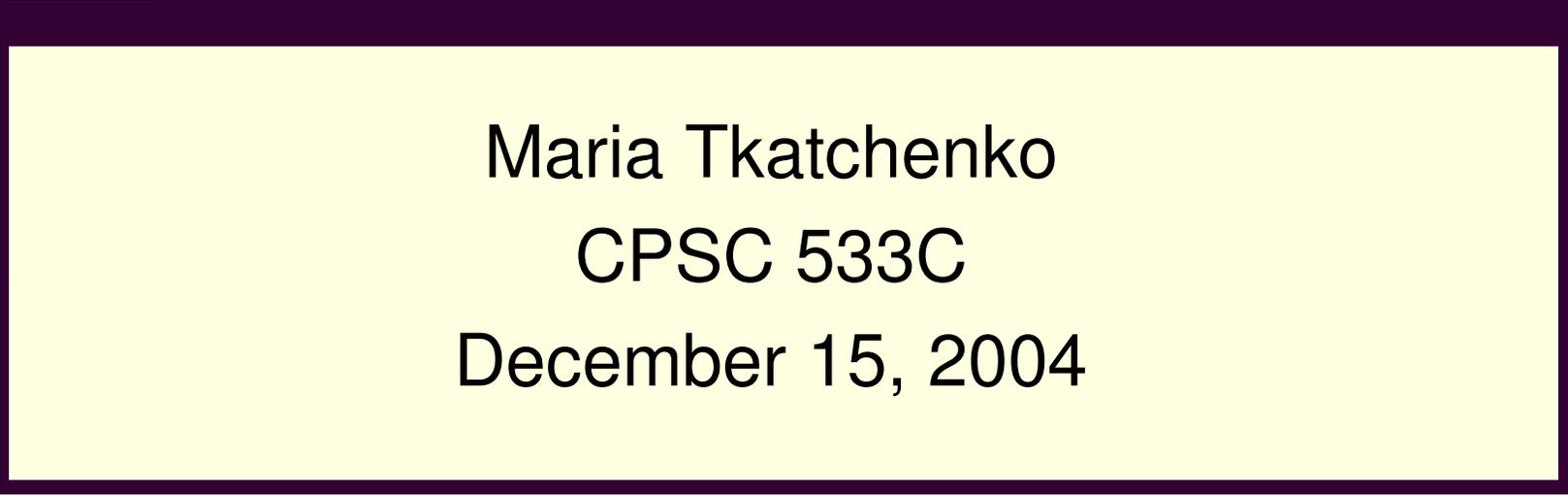




Survey Visualization



Maria Tkatchenko
CPSC 533C
December 15, 2004

Problem

- Survey composed of
 - Questionnaire
 - Collected data
- Want to explore relationships between responses to various questions

Data set

- National Election Studies center's pre- and post-election surveys
- National Center for Chronic Disease Prevention and Health Promotion Behavioral Risk Factor survey
- 1K - 100K respondents
- 100s of questions (dimensions)

Statistics

■ Correlation

- measure of “closeness” of two sets of data
- degree to which the association of two variables approaches a linear function
- ranges from -1 to 1
- higher (absolute) value is better



Main elements of solution

Textual View

Questionnaire

V023009x

B2x. Summary - Approve Congress Jo
Numeric
Missing eq 0, ge 8

B2x.

Do you APPROVE or DISAPPROVE of the way the U.S. Congress has been handling its job?
(Do you [approve/disapprove] STRONGLY or NOT STRONGLY?)

SUMMARY: APPROVAL OF CONGRESS HANDLING ITS JOB

Built from B2 and B2a.

1. Approve strongly
2. Approve not strongly
4. Disapprove not strongly
5. Disapprove strongly

8. Don't know (B2 or B2a)
9. Refused (B2 or B2a)
0. NA (B2 or B2a)

V023009X	Frequency
0	53
1	294
2	587
4	234
5	290
8	37
9	16

C1

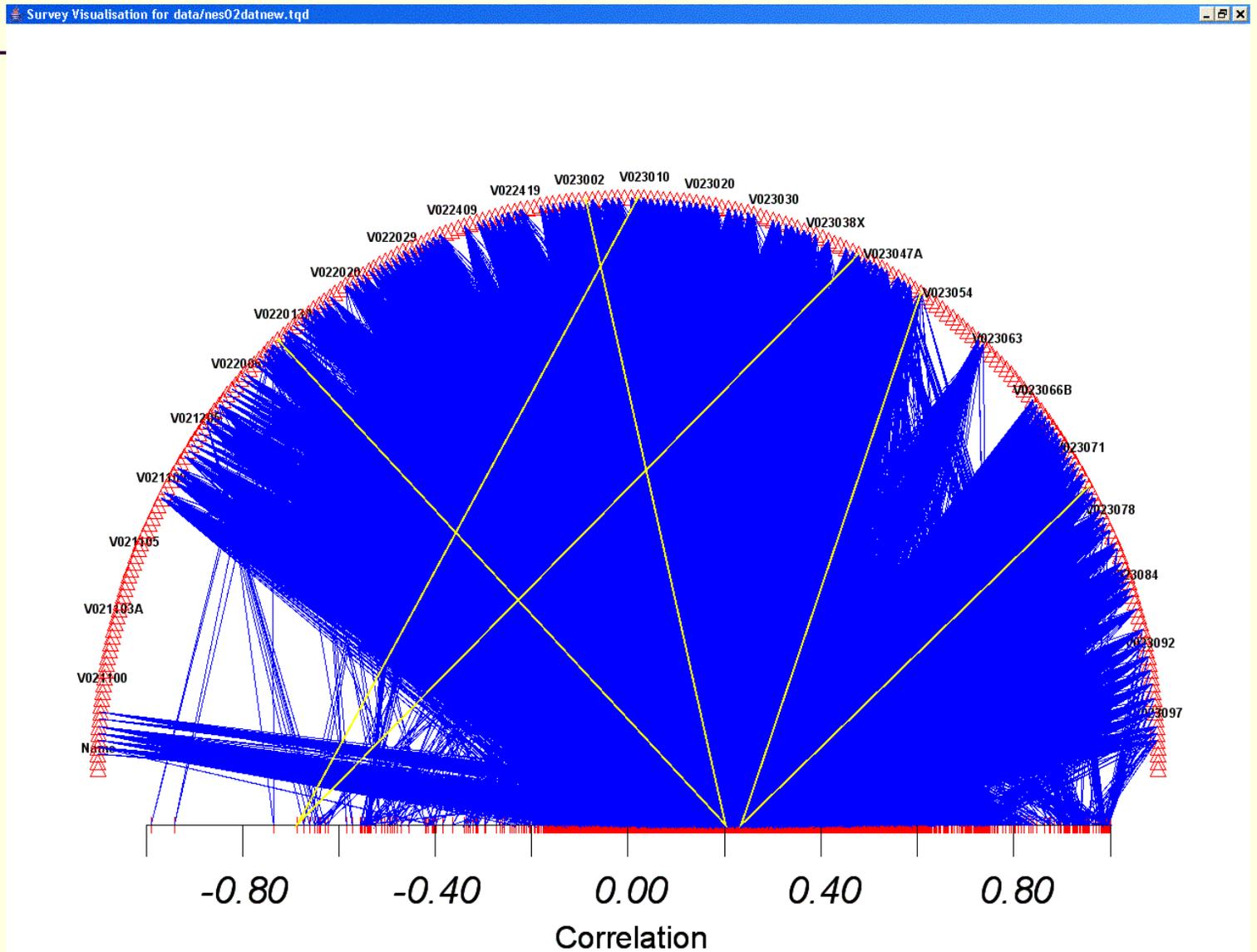
I'd like to get your feelings toward some people in the news these days. I'll read the name of a person and I'll ask you to rate that person on a thermometer that runs from 0 to 100 degrees. Rating above 50 means that you feel favorable and warm toward the person. Rating below 50 means that you feel unfavorable and cool toward the person. Rating right at the 50 degree mark means you don't feel particularly warm or cold. You may use any number from 0 to 100 to tell me how favorable or unfavorable your feelings are. If we come to a person whose name you don't recognize, just tell me and we'll move on to the next one.

V023010

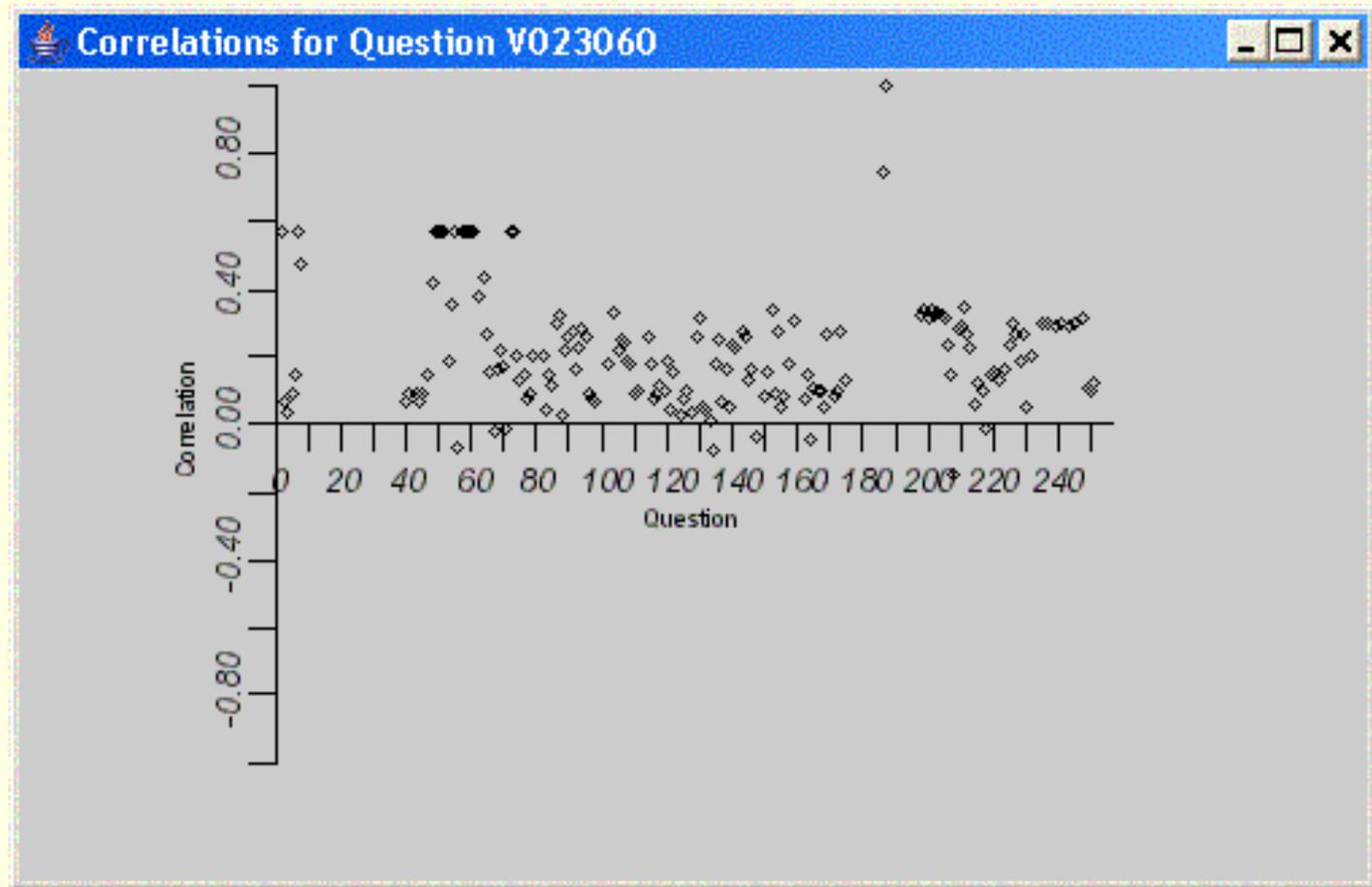
C1a. George W. Bush Thermomete
Numeric
Missing eq 887, ge 888

C1a.

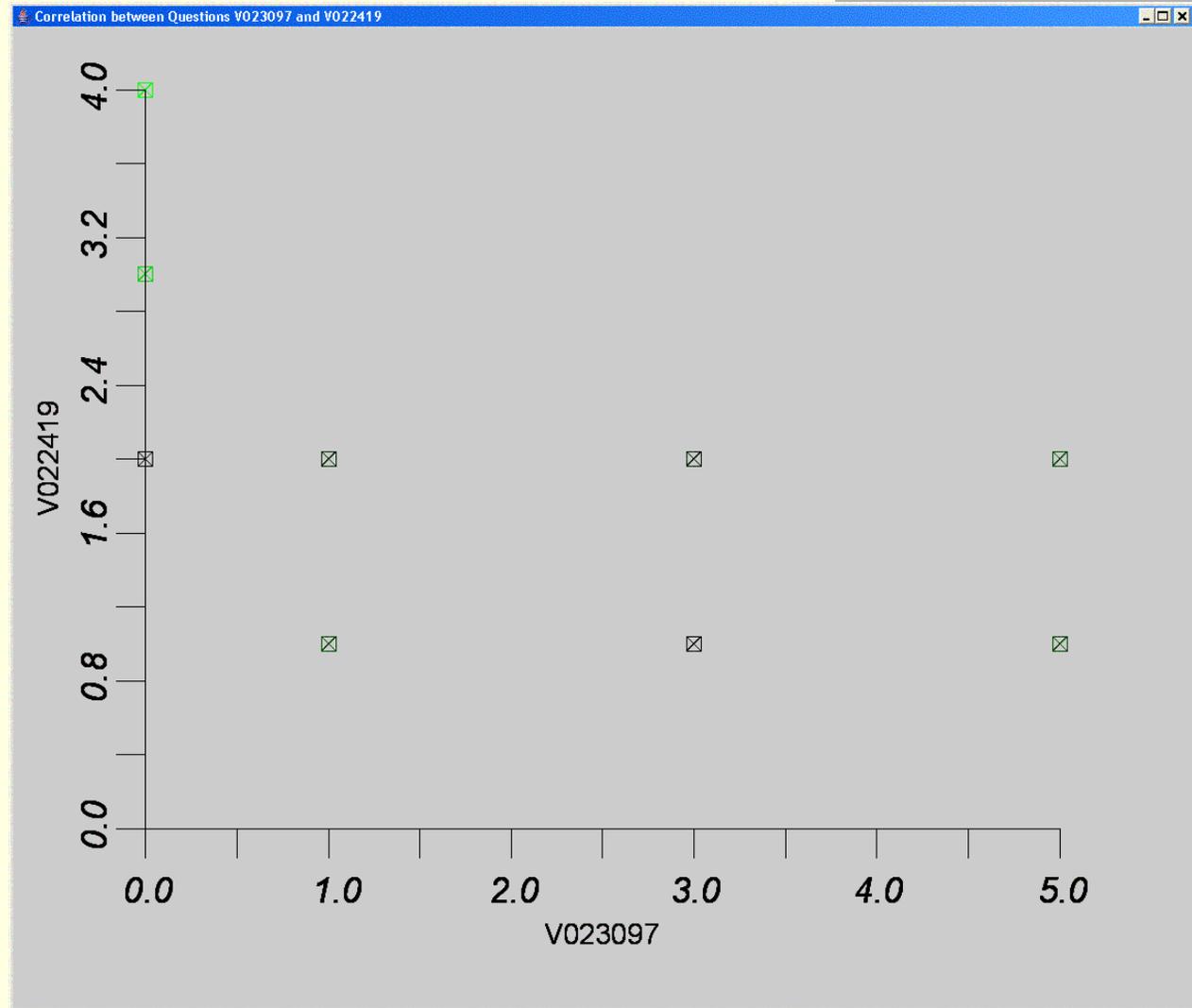
Questionnaire View



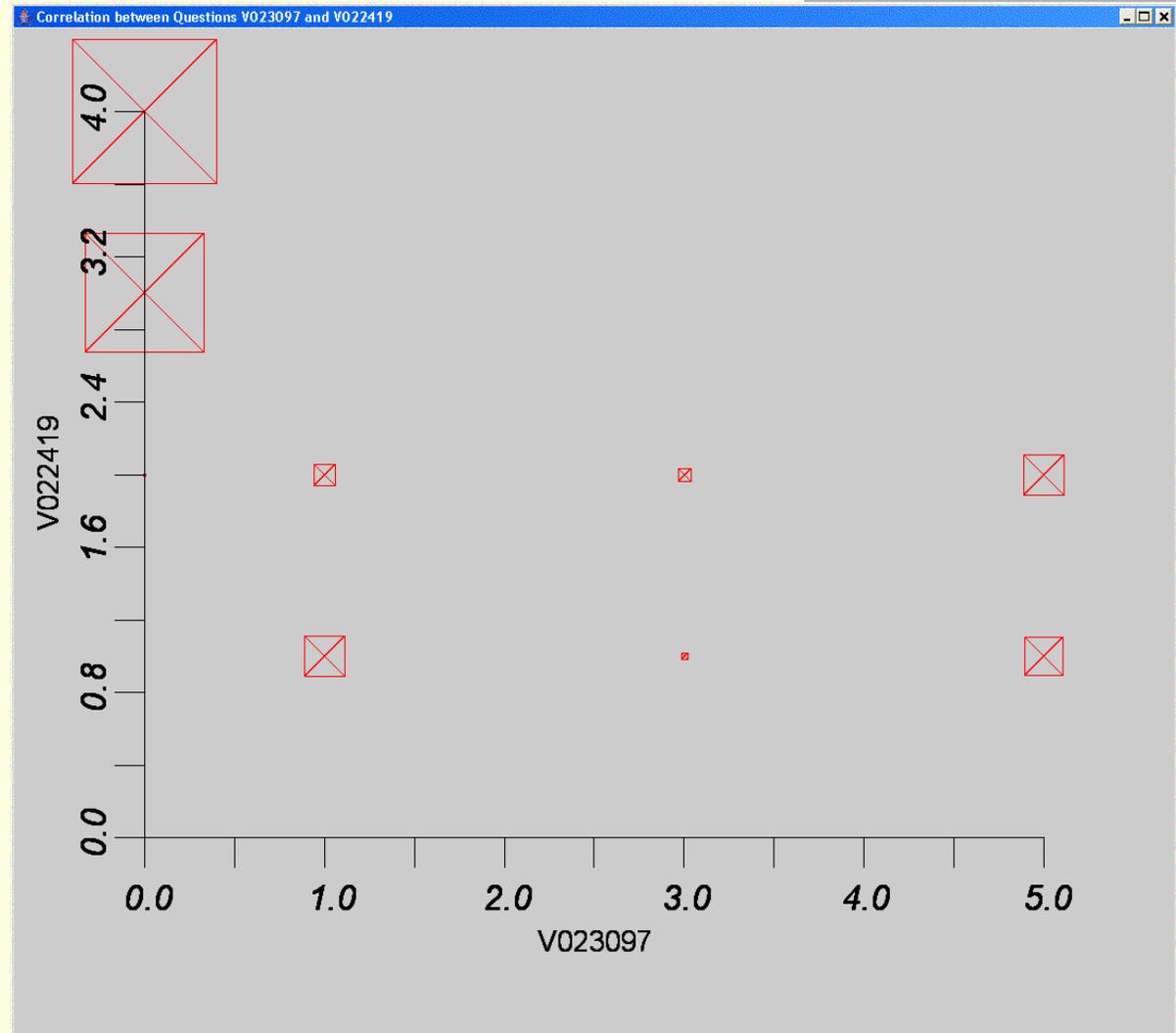
Correlation View



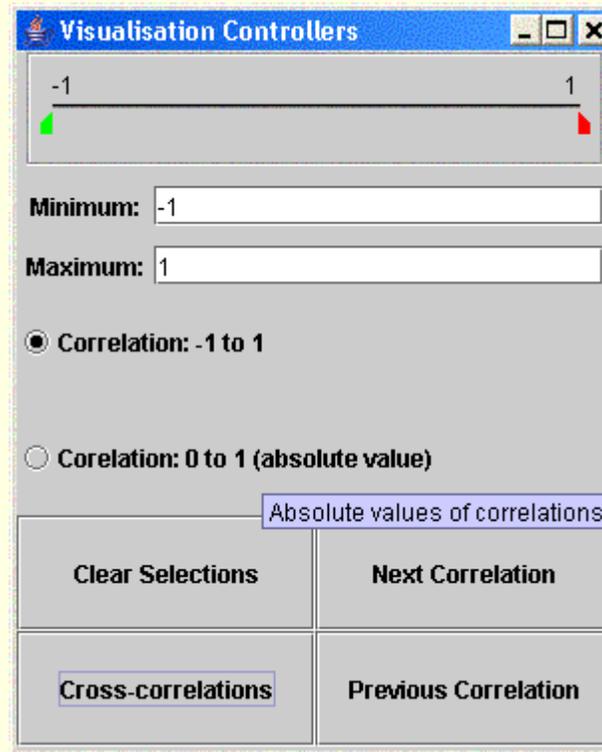
Raw Data View (1)



Raw Data View (2)



Control Panel



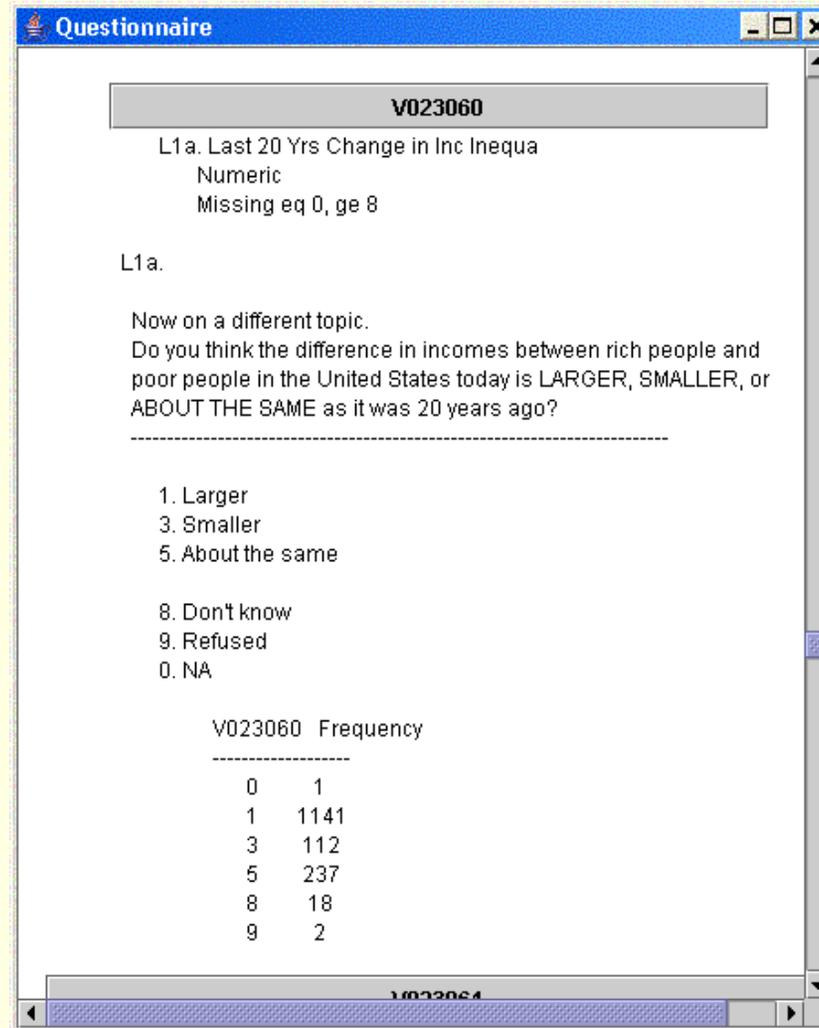
Implementation

- InfoVis Toolkit
- SGT Toolkit
- Java 2D and Swing
- Data formats
 - CSV, converted to TQD
 - TXT



Let's explore one possible scenario

Find a question



Questionnaire

V023060

L1a. Last 20 Yrs Change in Inc Inequa
Numeric
Missing eq 0, ge 8

L1a.

Now on a different topic.
Do you think the difference in incomes between rich people and poor people in the United States today is LARGER, SMALLER, or ABOUT THE SAME as it was 20 years ago?

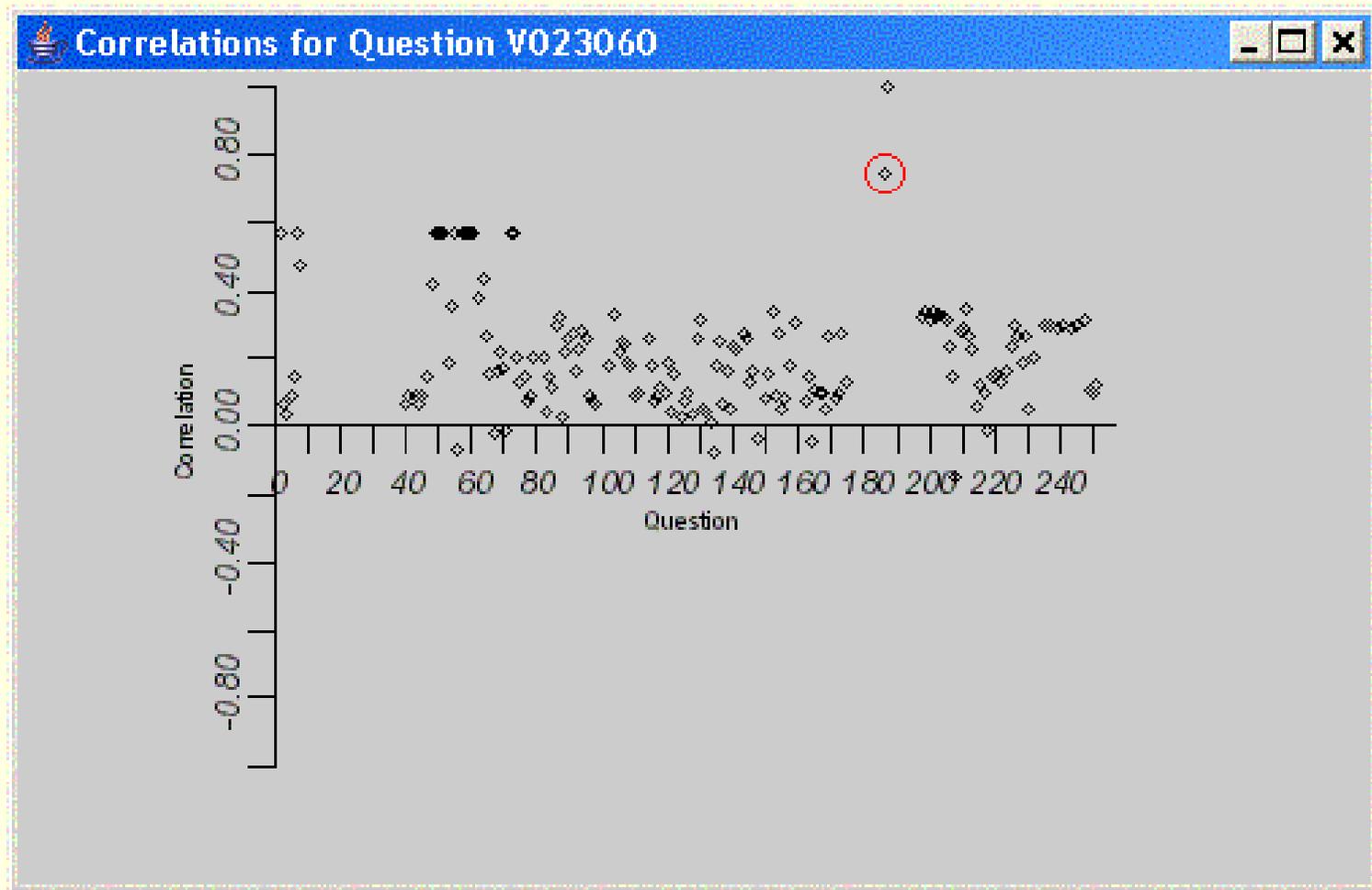
1. Larger
3. Smaller
5. About the same

8. Don't know
9. Refused
0. NA

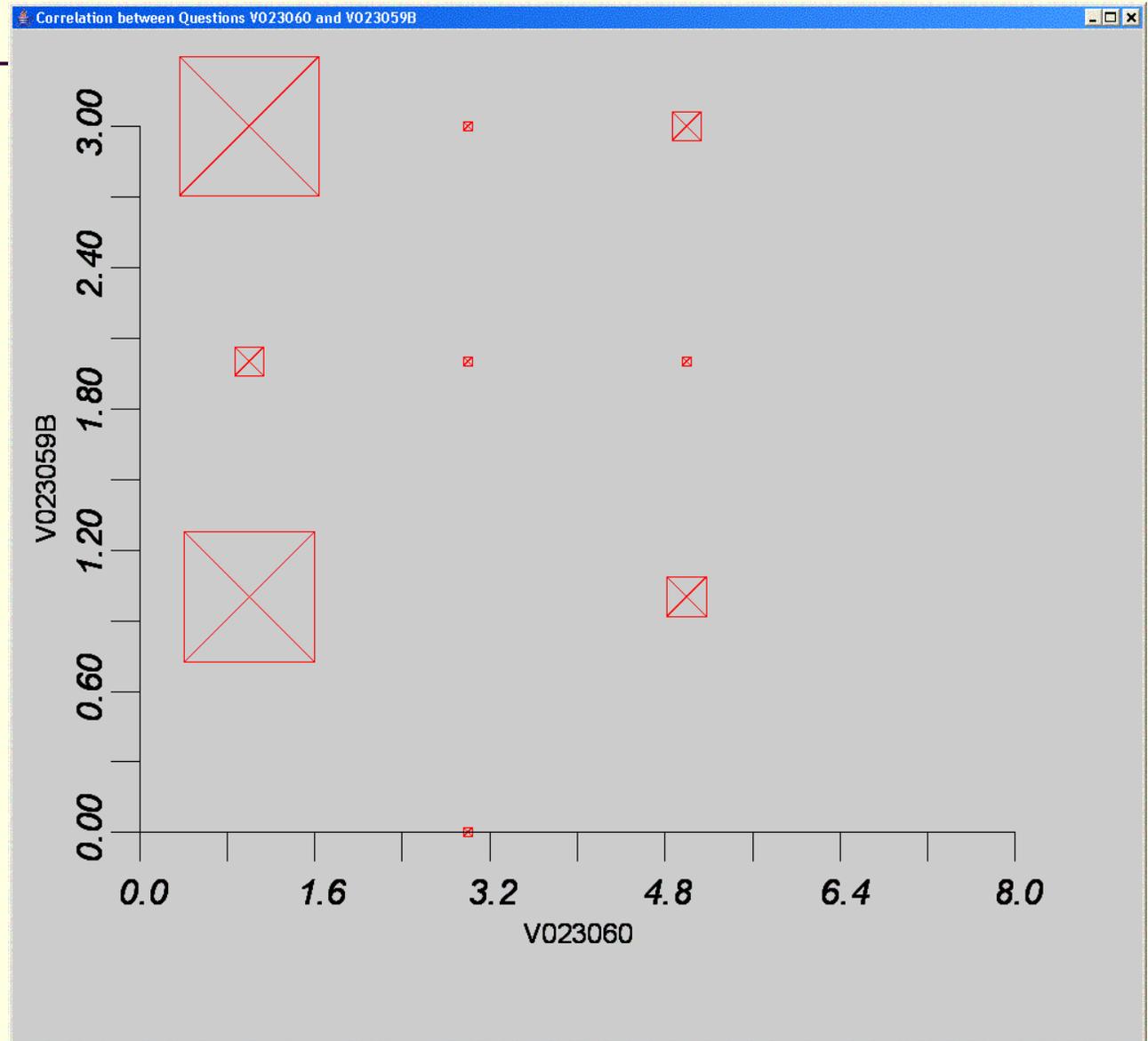
V023060 Frequency

0	1
1	1141
3	112
5	237
8	18
9	2

Correlation View

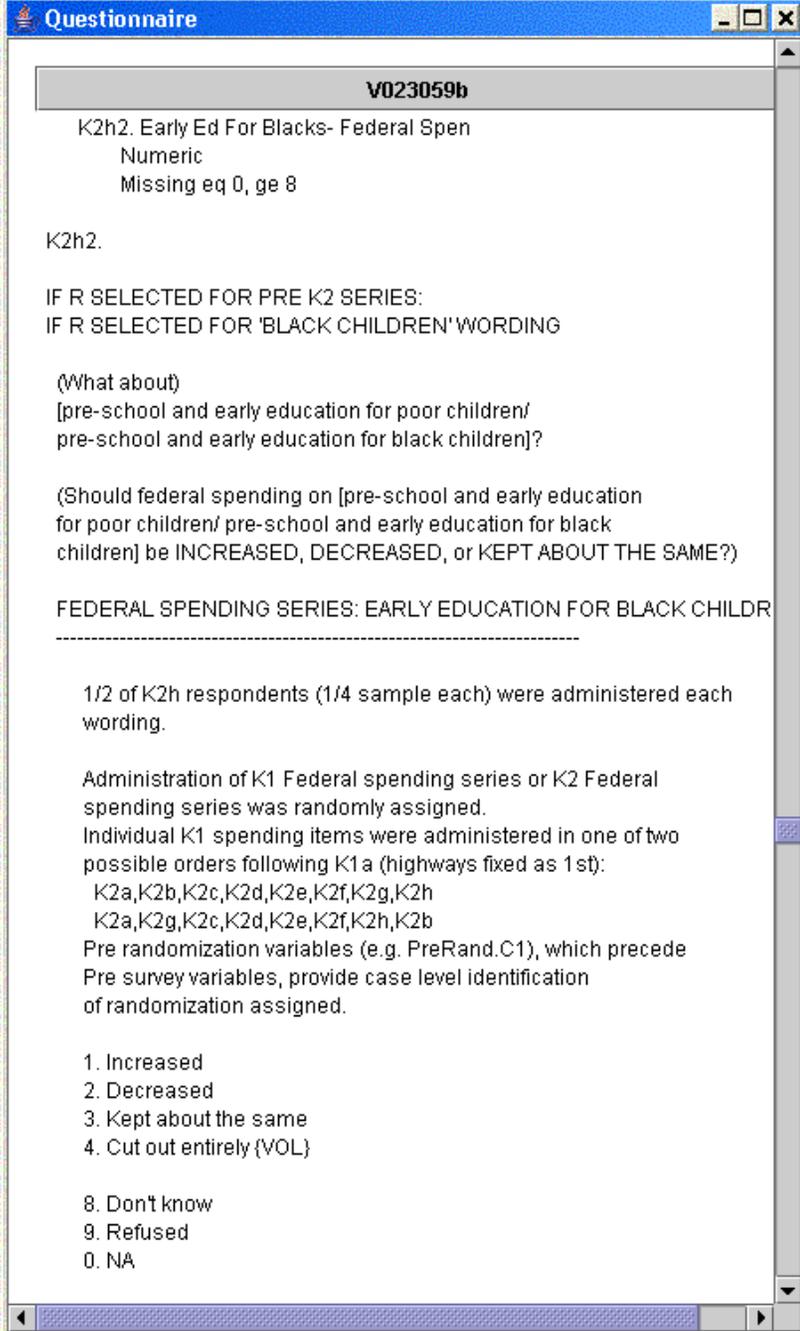


Raw Data View



Interesting correlation

with:



Questionnaire V023059b

K2h2. Early Ed For Blacks- Federal Spen
Numeric
Missing eq 0, ge 8

K2h2.

IF R SELECTED FOR PRE K2 SERIES:
IF R SELECTED FOR 'BLACK CHILDREN' WORDING

(What about)
[pre-school and early education for poor children/
pre-school and early education for black children]?

(Should federal spending on [pre-school and early education
for poor children/ pre-school and early education for black
children] be INCREASED, DECREASED, or KEPT ABOUT THE SAME?)

FEDERAL SPENDING SERIES: EARLY EDUCATION FOR BLACK CHILDREN

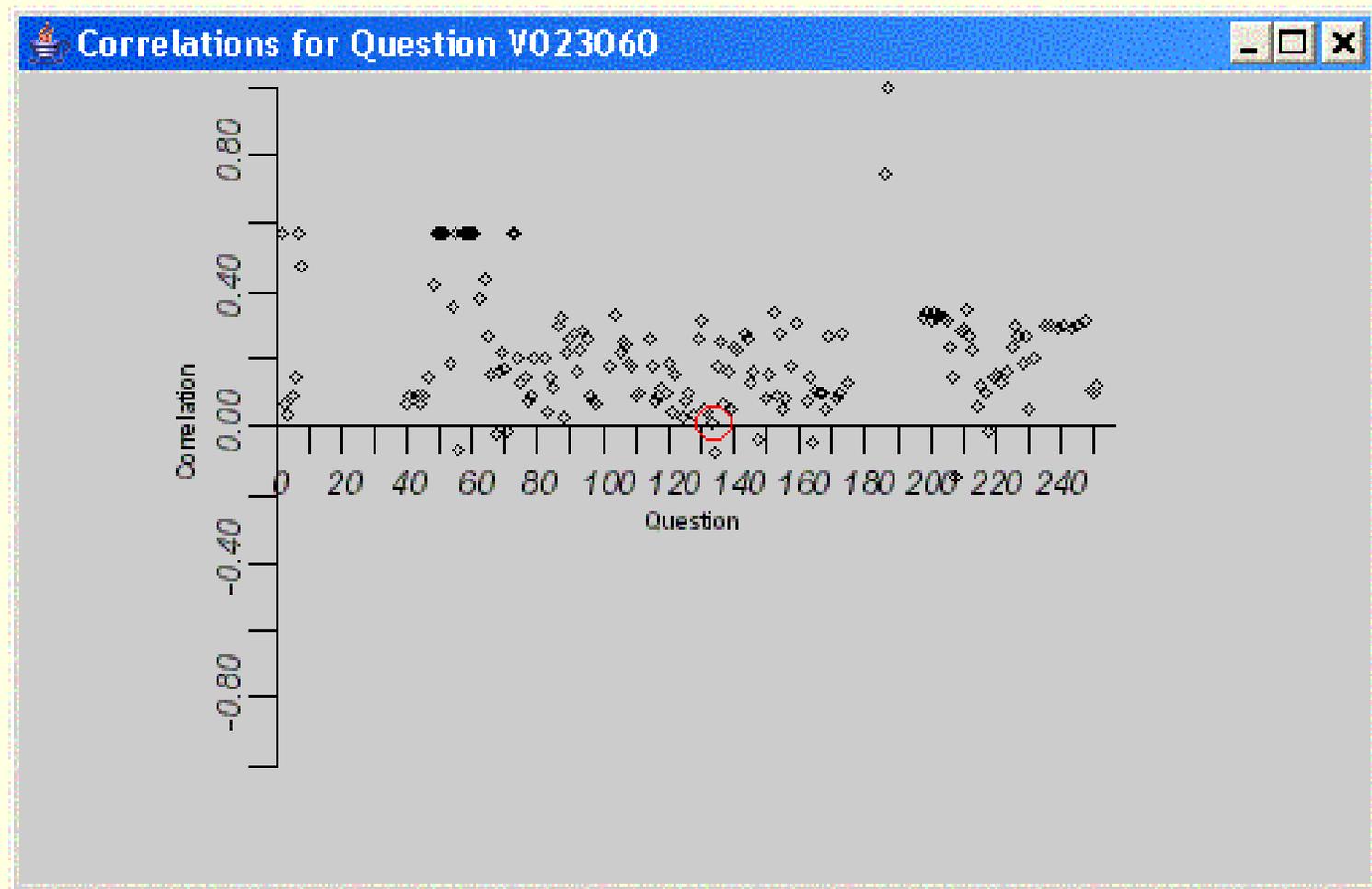
1/2 of K2h respondents (1/4 sample each) were administered each
wording.

Administration of K1 Federal spending series or K2 Federal
spending series was randomly assigned.
Individual K1 spending items were administered in one of two
possible orders following K1 a (highways fixed as 1 st):
K2a,K2b,K2c,K2d,K2e,K2f,K2g,K2h
K2a,K2g,K2c,K2d,K2e,K2f,K2h,K2b
Pre randomization variables (e.g. PreRand.C1), which precede
Pre survey variables, provide case level identification
of randomization assigned.

1. Increased
2. Decreased
3. Kept about the same
4. Cut out entirely (VOL)

8. Don't know
9. Refused
0. NA

Correlation View



Interesting correlation with:

Questionnaire
999. NA

V023013

C1d. Lieberman Thermomete
Numeric
Missing eq 887, ge 888

C1d.

(The next person is:
Joseph Lieberman
(Where on that thermometer would you rate him?)

{PROBE FOR DON'T KNOW RESPONSE: WHEN YOU SAY "DON'T KNOW" YOU MEAN THAT YOU DON'T KNOW WHO THE PERSON IS OR DO YOU HAVE SOMETHING ELSE IN MIND?}

THERMOMETER RATING: JOSEPH LIEBERMAN

Individual thermometers (C1) were administered in one of two possible orders following George W. Bush (fixed as 1st thermometer):
C1a,C1b,C1c,C1d,C1e,C1f,C1j,C1h,C1m,C1p,C1r
C1a,C1b,C1c,C1d,C1e,C1f,C1r,C1p,C1m,C1h,C1j
Pre randomization variables (e.g. PreRand.C1), which precede Pre survey variables, provide case level identification of randomization assigned.

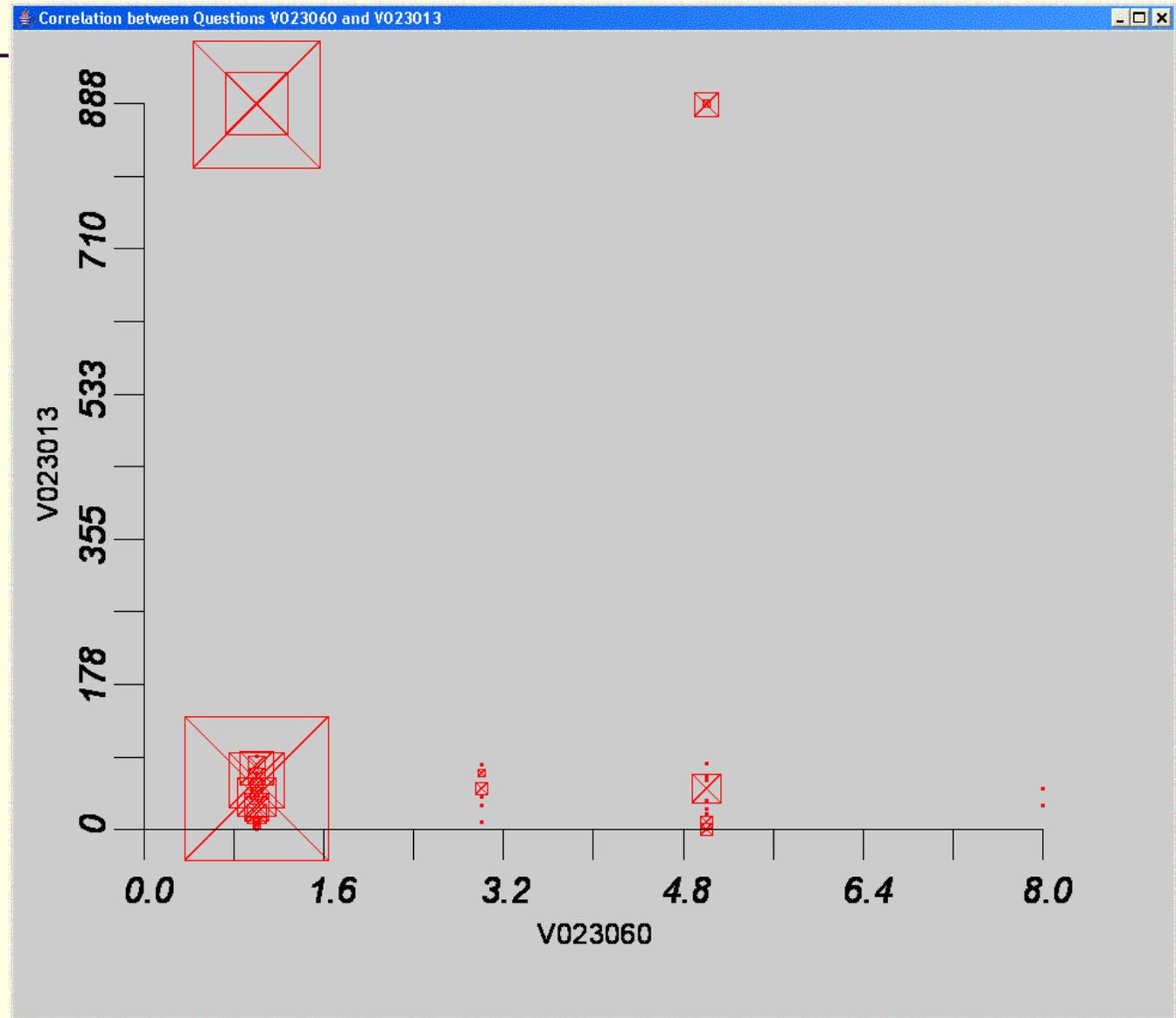
0-100.

887. Don't recognize
888. Don't know where to rate
889. Refused
999. NA

V023014

C1e. Nader Thermomete

Raw Data View





Another scenario

Questions

Questionnaire

8 1

V023089

N4b. Presentation at Place Worshi
 Numeric
 Missing eq 0, ge 8

N4b.

IF R REPORTS BEING AN ACTIVE MEMBER AT R'S PLACE OF WORSHIP

(As part of these activities,) have you given a presentation or speech (in the last six months)?

1. Yes
 5. No

8. Don't know
 9. Refused
 0. NA

INAP.
 5,8,9,0 in N4

V023089	Frequency
.	1026
0	2
1	215
5	266
8	1
9	1

Questionnaire

V023090a

P1 a2. Blacks Wrse Jobs: less able to lr
 Numeric
 Missing eq 0, ge 8

P1 a2.

IF R SELECTED FOR INCOME INEQUALITY BY RACE SERIES:
 IF R SELECTED FOR 'BLACKS WORSE JOBS-INCOME' WORDING:

[The first is:/(Next:)]
 'Because whites have more in-born ability to learn'.

(Would you say that this is VERY important, SOMEWHAT important, or NOT IMPORTANT AT ALL as an explanation for why [blacks/ whites] tend to have [worse/better] jobs and [lower/higher] incomes than [whites/blacks] do ?)

BLACKS WORSE JOBS: MORE IN-BORN ABILITY TO LEARN

Respondents were randomly assigned to 1 of 4 race/gender income inequality series.

Series 1: P1a1,b1,c1,d1,e1,f1,g1 ('whites better'), P1j,k,m
 Series 2: P1a2,b2,c2,d2,e2,f2,g2 ('blacks worse'), P1j,k,m
 Series 3: P2a1,b1,c1,d1,e1,f1,g1. ('women worse'), P2j,k,m
 Series 4: P2a2,b2,c2,d2,e2,f2,g2 ('men better'), P2j,k,m

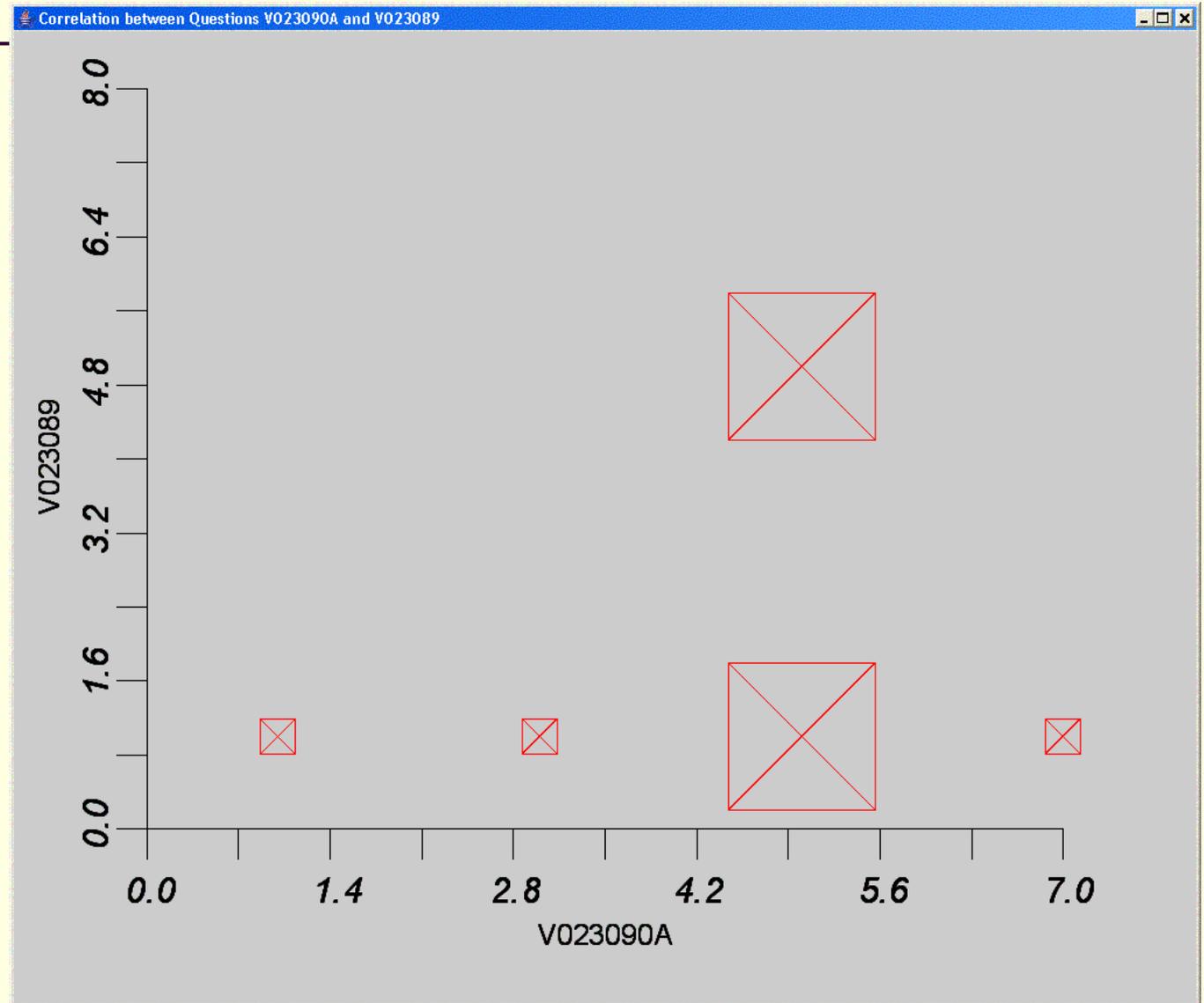
Reasons for inequality by race (either P1 series) were administered in one of two possible orders:
 P1a,b,c,d,e,f,g OR P1b,d,e,f,c,a,g

Pre randomization variables (e.g. PreRand.C1), which precede Pre survey variables, provide case level identification of randomization assigned.

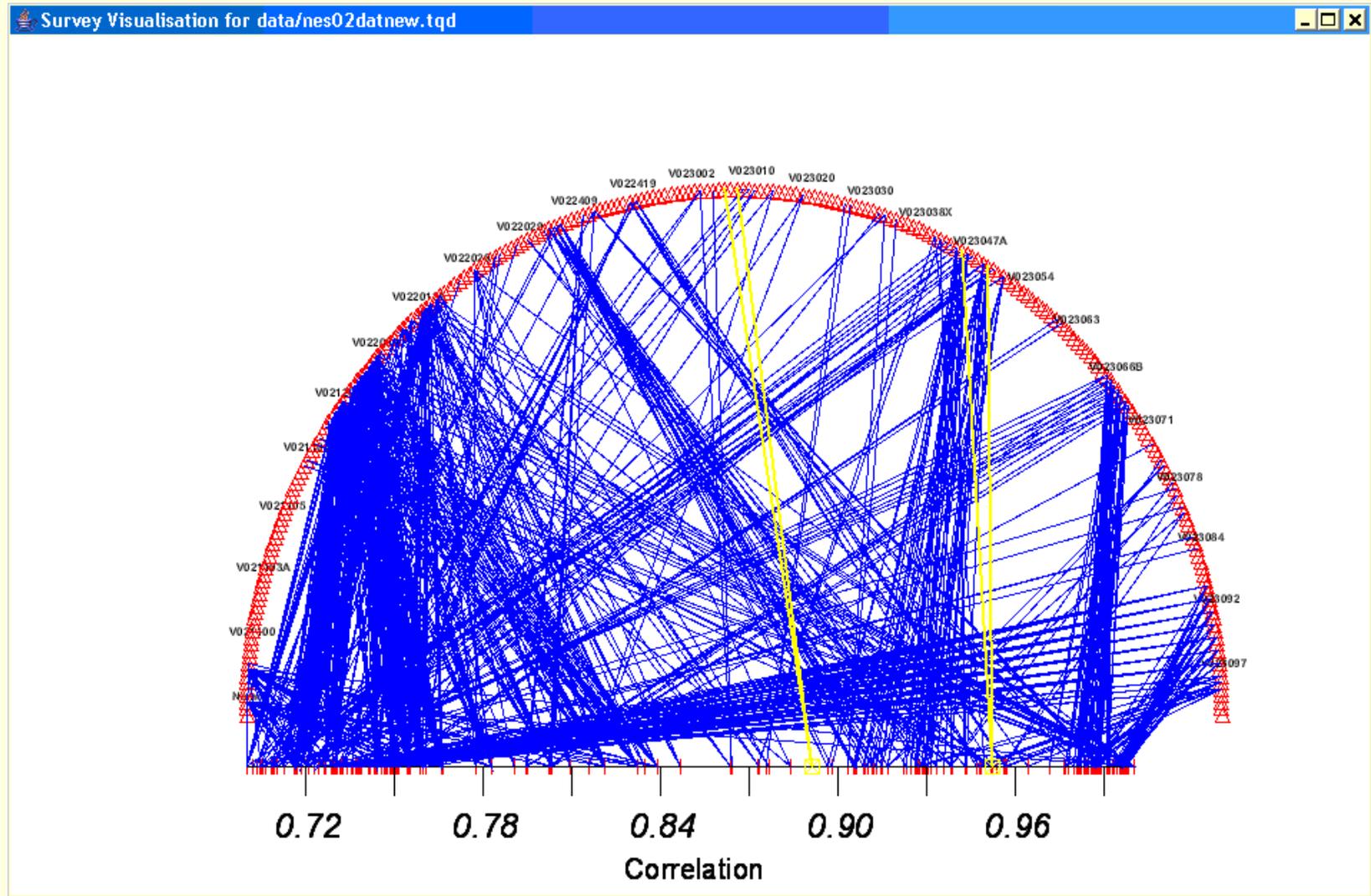
1. Very important
 3. Somewhat important
 5. Not important at all
 7. Statement isn't true (VOL)

8. Don't know
 9. Refused

Find a few correlations of interest



Select correlations





Questions?