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)
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

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

SUMMARY: APPROVAL OF CONGRESS HANDLING ITS JOB

Built from B2 and B2a:
1. Approve strongly
2. Approve not strongly
3. Disapprove not strongly
4. Disapprove strongly
5. Don’t know (B2 or B2a)
6. Refused (B2 or B2a)
7. NA (B2 or B2a)

V023009K: Frequency

<table>
<thead>
<tr>
<th></th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>264</td>
</tr>
<tr>
<td>2</td>
<td>567</td>
</tr>
<tr>
<td>3</td>
<td>264</td>
</tr>
<tr>
<td>4</td>
<td>250</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>

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 I 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 Thermomote

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing eq 887, ge 888</td>
<td></td>
</tr>
</tbody>
</table>

C1a.
Correlation View
Raw Data View (1)
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

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
2. Smaller
3. About the same
4. Don't know
5. Refused
6. NA

V023060 Frequency

<table>
<thead>
<tr>
<th>Value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1141</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
</tr>
<tr>
<td>3</td>
<td>237</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>
Correlation View
Questionnaire View
Raw Data View
Interesting correlation with:

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 K1a (highways fixed as 1st):

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
Questionnaire View
Interesting correlation with:

 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, C1h, C1j, 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
Raw Data View
Another scenario
Filter
Questions

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
2. No
3. Don't know
4. Refused
5. NA

INAP:
5,8,0,0 in N4

VO23089 Frequency

1. 1026
2. 2
3. 215
4. 266
5. 1
6. 1

VO23090a

P1a2. Blacks Worse Jobs: less able to Ir
Numeric
Missing eq 0, go 8
P1a2.

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 worse), 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 P1a,c,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
2. Somewhat important
3. Not important at all
4. Statement isn't true (VOL)
5. Don't know
6. Refused
Find a few correlations of interest
Select correlations
Cross-correlations
Questions?