

Making sense of your data

Evaluation Workshop Series: Session 2
November 12, 2010

Presenters:

Kristin Dillon and Jennifer Maxfield



WILDER
FOUNDATION

Outline

- Preliminary steps
- Organizing your data
- Analyzing your data
- Interpreting your results and drawing conclusions
- Excel demonstration

Preliminary steps

Preliminary steps

- Develop your evaluation plan
 - What are your key evaluation questions?
 - What information is needed to answer the evaluation questions?
 - What/who are your information sources?
 - How will you collect data?
 - How will you analyze the data?
- Collect data

Organizing your data

Organizing your data

- Name variables using a consistent format
 - Short
 - Intuitive
 - Single word is preferable

Don't

VAR001

Date of referral

Do

Q1_location

ReferralDate

Organizing your data

- Assign a unique identifier to each individual
 - To prevent duplicates
 - To prevent entering data on the wrong person
 - To link information across datasets

Organizing your data

■ Using name as an identifier

Name
MyLinh Nguyen
My Linh Nguyen
Kenneth Roberts, Jr.
Ken Roberts
emily ann meyers
EMELY MEYER
Juan Hernandez Romero
Juan Hernandez
Gloria Jones
Gloria Rogers

■ Pros:

- How you refer to participants

■ Cons:

- Typos
- Prefixes and suffixes
- Middle name or initial
- Multiple last names
- Upper and lower casing
- Name changes

Organizing your data

■ Using SSN as an identifier

■ Pros:

- May be required for federal applications

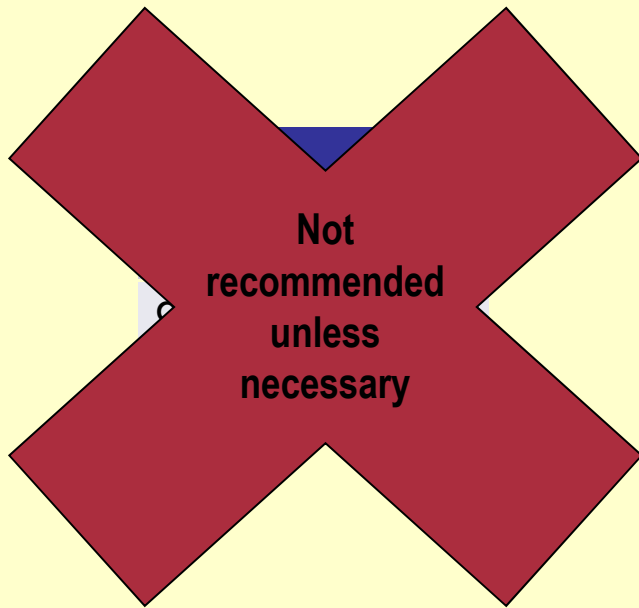
■ Cons:

- Hyphens, spaces, or none
- Privacy concerns

SSN
999-99-9999
999 99 9999
999999999

Organizing your data

■ Using SSN as an identifier



■ Pros:

- May be required for federal applications

■ Cons:

- Hyphens, spaces, or none
- Privacy concerns

Organizing your data

■ Using telephone number as an identifier

Phone
(999)999-9999
999-999-9999
999 999 9999
9999999999
999-9999
99999999

■ Pros:

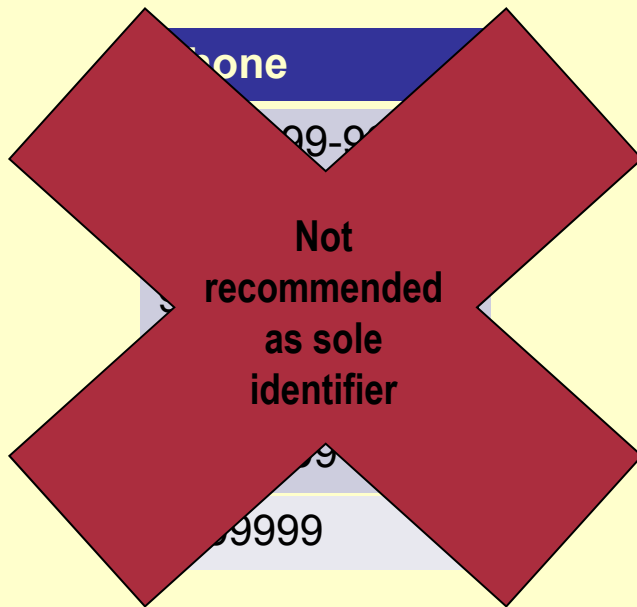
- This may be something you already collect for program purposes

■ Cons:

- Area code
- Parentheses, hyphens, or none
- Changes
- Not unique

Organizing your data

■ Using telephone number as an identifier



■ Pros:

- This may be something you already collect for program purposes

■ Cons:

- Area code
- Parentheses, hyphens, or none
- Changes
- Not unique

Organizing your data

■ Using student ID as an identifier

StudentID
162345
345628
466585
100326
799866

■ Pros:


- Pre-existing ID
- Allows you to link your data to other data

■ Cons:

- Might be hard to obtain
- Privacy concerns

Organizing your data

■ Using student ID as an identifier

StudentID	
162345	
345628	
466585	Recommended
100326	with privacy
799866	controls

■ Pros:

- Pre-existing ID
- Allows you to link your data to other data

■ Cons:

- Might be hard to obtain
- Privacy concerns

Organizing your data


- Assigning a unique identifier

IntakeNumber
100
101
102
103
104

- Assign a unique ID number at intake and use in conjunction with other identifying information

Organizing your data

- Assigning a unique identifier

IntakeNumber	
100	
101	
102	
103	Recommended
104	

- Assign a unique ID number at intake and use in conjunction with other identifying information

Organizing your data

- Multi-record

- Multiple rows of data per individual

- Single record

- One row of data per individual
- Usually preferable for analysis

- Identifying duplicate cases can be a challenge

- The CDC's Link Plus software can help.

Free download online:

www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm

Organizing your data

- Do not use color coding
 - Colors cannot be sorted or analyzed

Don't

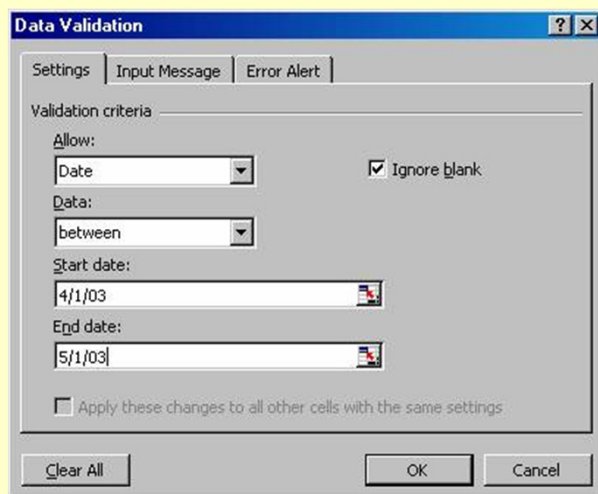
StudentID
162345
345628
466585
100326
799866
162345

Do

StudentID	Status (0=exited, 1=current)
162345	1
345628	1
466585	0
100326	0
799866	0
162345	1

Organizing your data

- Enter data in a consistent format
- Benefits of using numeric codes
 - E.g., 0 = no, 1 = yes
- Limit permissible responses
 - Data validations in Excel



Organizing your data

- Avoid leaving anything blank
- Instead, use a code to explain why there are no data
 - 6 = Missing
 - 7 = Don't know
 - 8 = Refusal
 - 9 = Not applicable

Organizing your data

- Usually it is best to create new variables rather than override previous information
 - E.g., Status changes

OriginalStatus	StatusChange1	StatusChange1 _Date	StatusChange2	StatusChange2 _Date	CurrentStatus
Enrolled	-9	-9	-9	-9	Enrolled
Enrolled	Exited	10/11/2009	Enrolled	12/1/2009	Enrolled
Waitlist	Enrolled	08/05/2010	-9	-9	Enrolled
Enrolled	Exited	03/15/2008	-9	-9	Exited
Ineligible	-9	-9	-9	-9	Ineligible

Organizing your data

- Keep documentation, such as a codebook
 - Variable name
 - Variable description
 - Response options or categories
 - Assigned values
 - Data source
 - Timing of data collection
 - Explanation of any changes

Analyzing your data

Analyzing your data

- Continuum of complexity
- Descriptive analysis
 - Frequency distribution
 - Central tendency
 - Variability
- Inferential analysis

Analyzing your data

- Types of data

- Categorical

- Nominal

- Ordinal

- Continuous



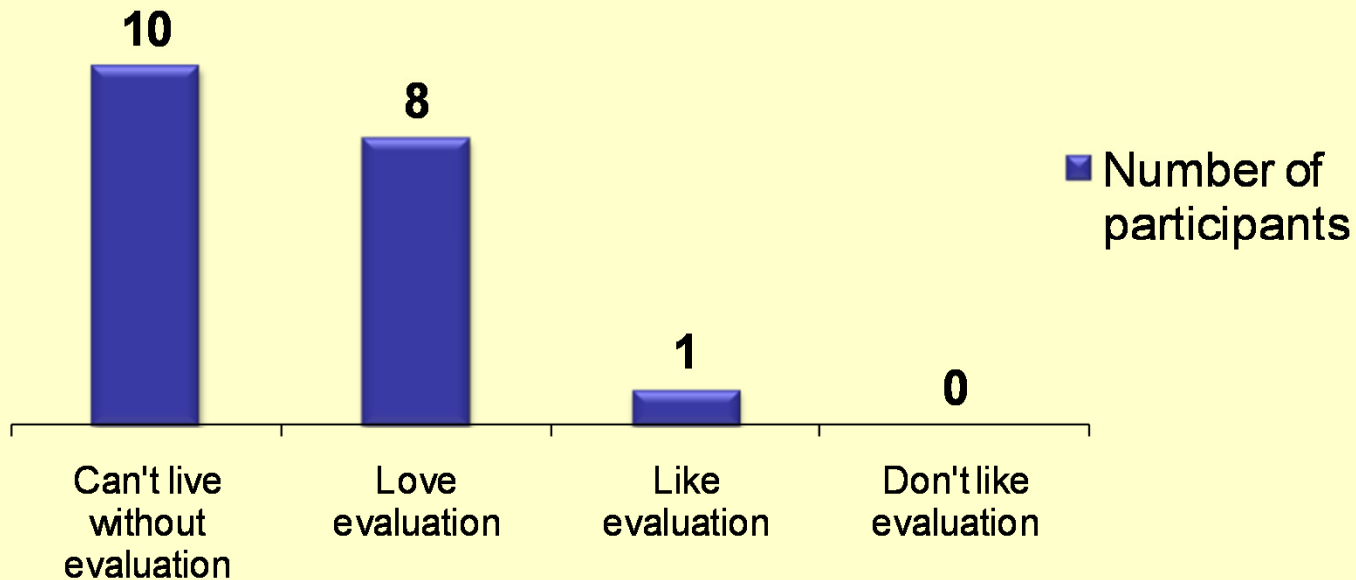
When I hear “data analysis,” I mostly feel...

- 7% 1. Scared or anxious
- 33% 2. Overwhelmed
- 4% 3. Happy
- 44% 4. Excited
- 11% 5. Neutral
- 0% 6. None of the above

Analyzing your data – Descriptive

■ Frequency distributions

Participant thoughts about evaluation



Analyzing your data – Descriptive

- Central tendency
 - Average or Mean

Number of siblings

$$1 + 1 + 1 + 2 + 2 + 3 + 5 + 9 = 24$$

$$24 \div 8 = 3 \text{ siblings}$$

Analyzing your data – Descriptive

- Central tendency

- Median

Number of siblings

$$1 + 1 + 1 + 2 + 2 + 3 + 5 + 9 = 24$$

2 siblings

Analyzing your data – Descriptive

- Central tendency

 - Mode

Number of siblings

$$1 + 1 + 1 + 2 + 2 + 3 + 5 + 9 = 24$$

1 sibling

Analyzing your data – Descriptive

■ Variability

- Minimum and maximum

Number of siblings

1 1 1 2 2 3 5 9

1 to 9

Analyzing your data – Descriptive

■ Variability

— Range

Number of siblings

1 1 1 2 2 3 5 9

$$9 - 1 = 8$$

Analyzing your data – Descriptive

■ Variability

- Standard deviation



Number of siblings

1 1 1 2 2 3 5 9

= 2.777

Analyzing your data – Inferential

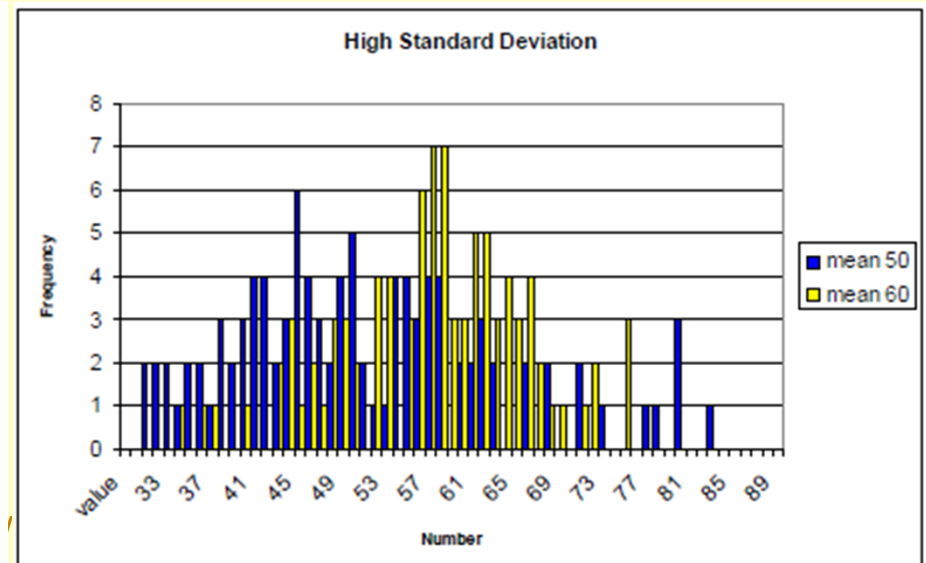
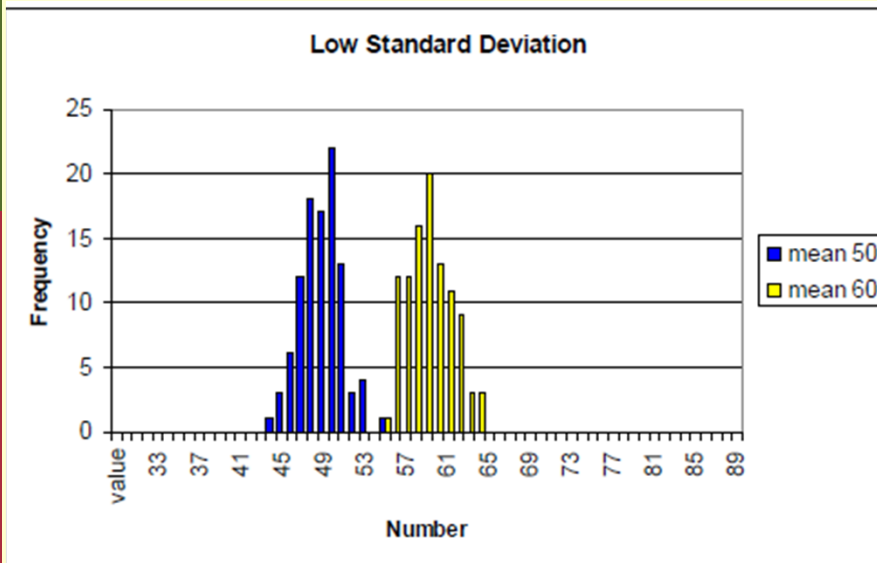
- Common types of tests
 - Chi squares
 - Correlations
 - T-tests
 - Analysis of variance

Analyzing your data – Inferential

- Statistical significance
 - Strength of the relationship
- Substantive or clinical significance
 - Based on agreed upon criteria

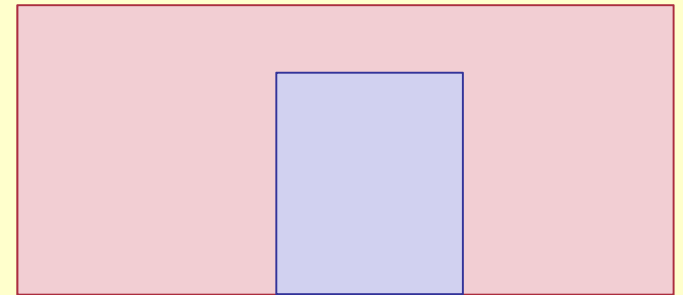
Analyzing your data – Inferential

- Factors impacting statistical significance
 - Amount of variability



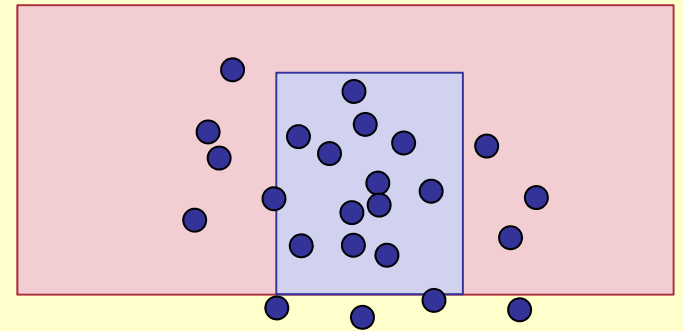
Analyzing your data – Inferential

- Factors impacting statistical significance
 - Effect size



Analyzing your data – Inferential

- Factors impacting statistical significance
 - Size of the sample



Interpreting your data



WILDER
FOUNDATION

Interpreting your results

Involves stepping back to consider what the results mean

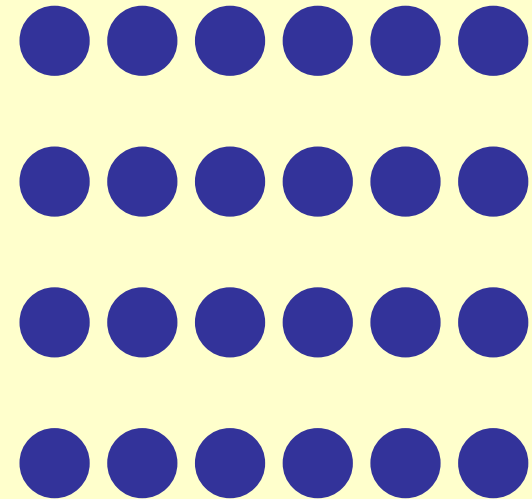
Don't forget to:

- Involve stakeholders
- Consider practical value
- Acknowledge limitations
- Seek consultation as needed

Interpreting your results

Look for what stands out:

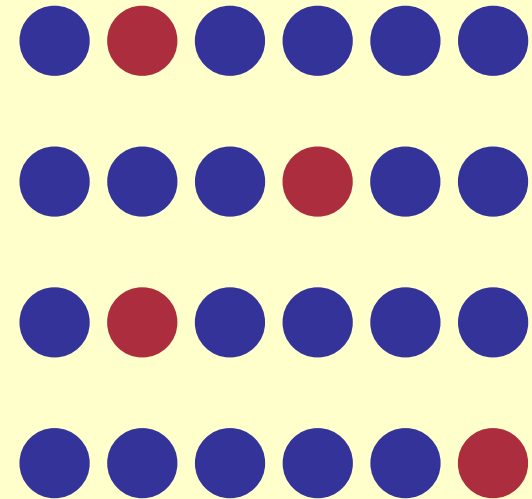
- Patterns and themes



Interpreting your results

Look for what stands out:

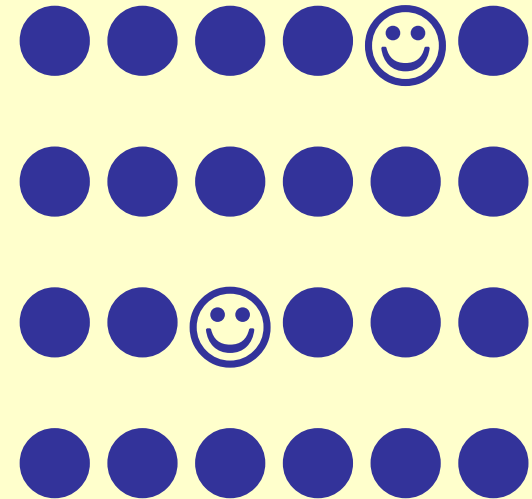
- Surprising findings



Interpreting your results

Look for what stands out:

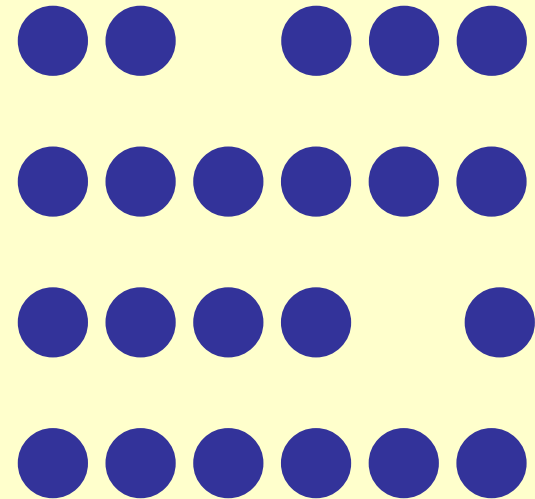
- Interesting stories



Interpreting your results

Look for what stands out:

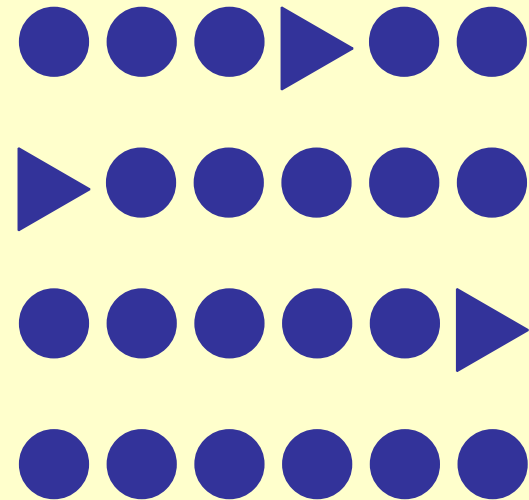
- Additional data needs



Interpreting your results

Look for what stands out:

- Recommendations or suggestions for the future



Interpreting your results

Think about the context:

- Are there exceptions to the patterns or themes?
- Do the results make sense?
- Are the results statistically or clinically significant?
- Are there inconsistencies in the results?
- What is the overall picture?

Interpreting your results

Common pitfalls:

- Cherry picking data
- Not looking at the overall picture
- Misrepresenting findings
- Straying from the results

Interpreting your results activity

Select the best interpretation of these data.

Thirty-seven percent of 17- to 20-year-olds are comfortable discussing their alcohol consumption habits with their family and friends.

3%

1. Thirty-seven percent of 17- to 20-year-olds consume alcohol regularly.

0%

2. Teenagers who drink often do so while talking with their families.

0%

3. Teenage drinking improves family communication.

97%


4. Not all teenagers are comfortable discussing whether they drink or not.





Select the best interpretation of these data.

Seventy-seven percent of the 3rd graders who stayed in the same school in 2010 read at grade level, but only 59 percent of those who transferred schools during the year did.

- 96% 1. Students who transferred schools were less likely to read at grade level than those who stayed in the same school. 
- 0% 2. Low reading proficiency in third grade makes students more likely to change schools.
- 4% 3. Kids who move in third grade are less likely to graduate from high school.
- 0% 4. One-quarter of third grade students in Minnesota can't read.

Select the best interpretation of these data.

Forty percent of homeless adults reported a job loss or reduced hours as a reason they lost their housing.

0% 1. Job loss or reduction is the most commonly reported cause of homelessness.

0% 2. Creating jobs is the best way to prevent homelessness.

18% 3. Forty percent of homeless adults have lost their job or had decreased work hours.

82% 4. Changes in employment contributed to a loss of housing for many homeless adults.



Select the best interpretation of these data.

In Minnesota, 53% of 6th graders, 38% of 9th graders and 26% of 12th graders (all males) reported that they were bullied at least once by other students during the past 30 days.

3%

1. As males students get older, they get bullied less.

3%

2. A targeted intervention focused on reducing bullying should be provided to half of 6th grade males.

0%

3. The study shows that only 53% of 6th grade boys have ever experienced bullying.

93%

4. As male students get older, a smaller proportion report experiencing bullying.



Select the best interpretation of these data.

When coalition members were asked how much their coalition had increased community awareness of the coalition's efforts, 71% of respondents said "a lot," 29% said "a little," and 0% said "not at all."

0% 1. 71% of people in the community have a lot of awareness about the coalition's work.

52% 2. The majority of coalition members surveyed feel that the coalition has increased community awareness of their work a lot. 

0% 3. 29% don't think that the coalition has increased awareness of their efforts.

48% 4. Every coalition member believes the coalition has increased community awareness at least a little.

Select the best interpretation of these data.

In a study of tobacco usage, 23% of adults with an income of \$35,000 or less are current smokers, compared to 11% of those with an income of more than \$75,000. Also, 26% of adults with an education of less than high school are smokers, compared to 6% of those with college degrees or higher education.

8%

1. The study shows that 77% of adults with a household income of \$35,000 or less have never smoked.

12%

2. Level of education plays a greater role in the prevention of smoking than household income.

80%

3. Education appears to be a greater protective factor in tobacco usage than household income.



Thank you!

**For more information please visit
www.wilderresearch.org**