# Formalizing your Curiosity

Wisconsin Public Library Consortium (WPLC)
Final Workshop

ANALYSIS

Joshua H. Morrill, PhD
Morrill Solutions Research

joshua@morrillsolutions.com
608-772-0924

# Joshua H. Morrill, PhD
# Morrill Solutions Research

Josh has been providing research and evaluation consulting with corporate, non-profit, and academic groups for the last 13 years. He worked as an emerging trends analyst for Gartner where he helped uncover some early usage of cell phones for game play.

As a consultant, Josh has worked extensively with the National Science Foundation in the assessment of various STEM (Science Technology Engineering & Math) technology initiatives. He has worked with National Science Digital Library project to assess the effectiveness of this initiative for faculty and librarians across the nation. And, more recently, Josh has worked with NSF and partners at the University of Illinois and University of Central Florida to conduct a large national study to develop a learner segmentation based on what people want out of learning technologies.

Josh has also partnered extensively with libraries. In particular, Josh has partnered with the Wisconsin Public Library Consortium (WPLC) in the development and implementation of a longitudinal study of Wisconsin public library users (and non-users) in 2003, 2007 and 2012. Josh also partnered with the University of Wisconsin and University of Illinois systems to assess student use and reactions to eTextbooks.

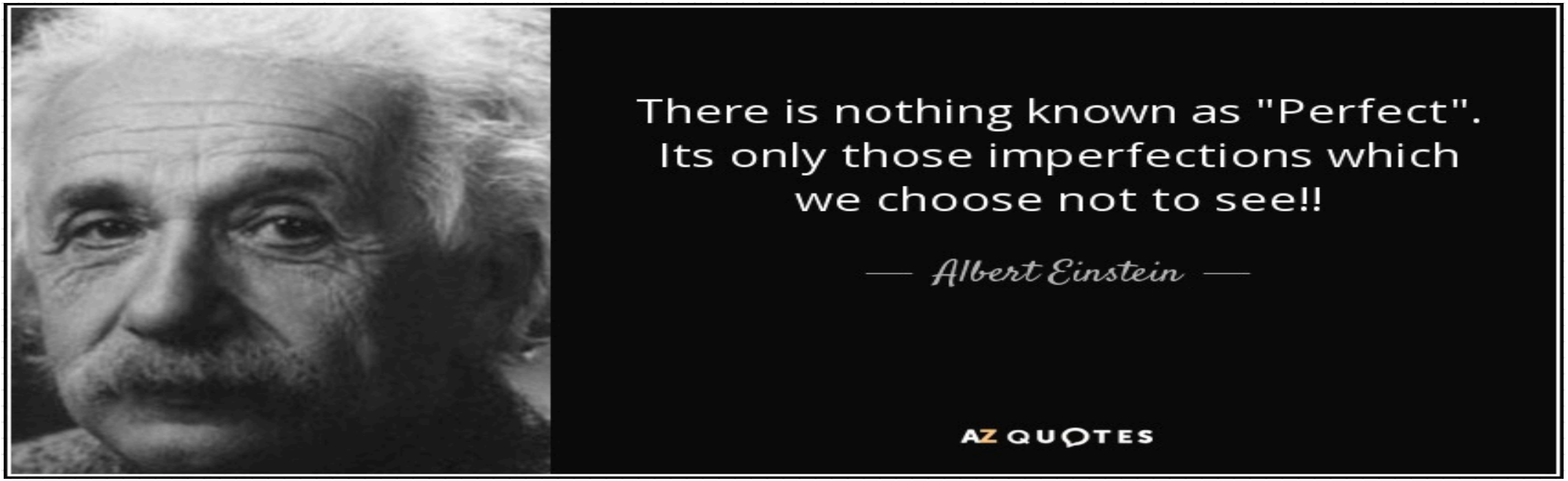
In addition to research, Joshua has also conducted several strategic planning and future-visioning workshops for corporate, library, education and non-profit groups that use scenario-based planning to help these organizations plan around future uncertainties. Joshua is a frequent speaker nationally. Josh holds a Ph.D. in communication science from the University of Wisconsin-Madison with a concentration in small group decision-making and quantitative research design and analysis

***Research is formalized curiosity. It is poking and prying with purpose***

-Zora Neale Hurston

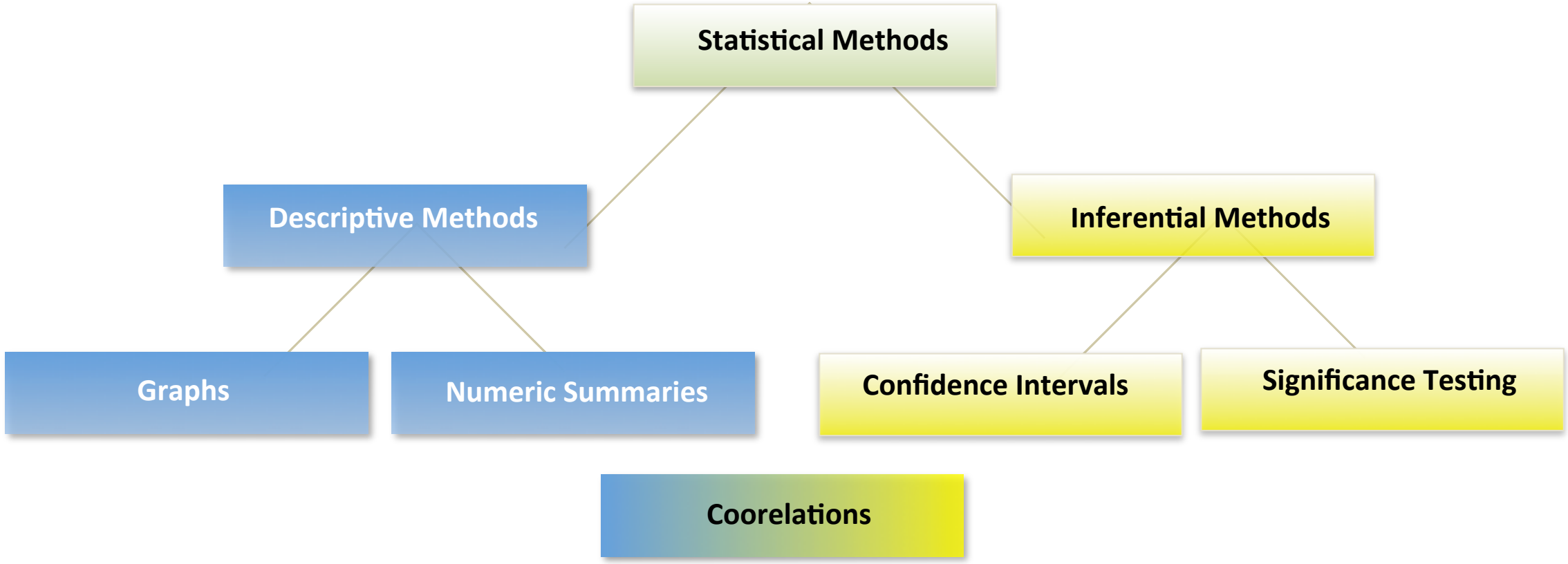
# ALL Measurement Has Error!

*Some measures have less/ some more but it is <u>always</u> there*




**The goal of research then is to minimize and parse (explain) the components that contribute to this error in order to make the best informed decision possible. Analysis is one way to do that.**

# What exactly do we mean by "Analysis"?



# You have data…Now what?
# Steps for approaching your data

1

**Ask yourself**

*How generalizable is this information? Is the sample sufficiently diverse/ large?*

2

**Do some basic frequencies of your data?**

*For every question frequency, mean, standard deviation*

3

**Ask yourself if any significance testing is needed**

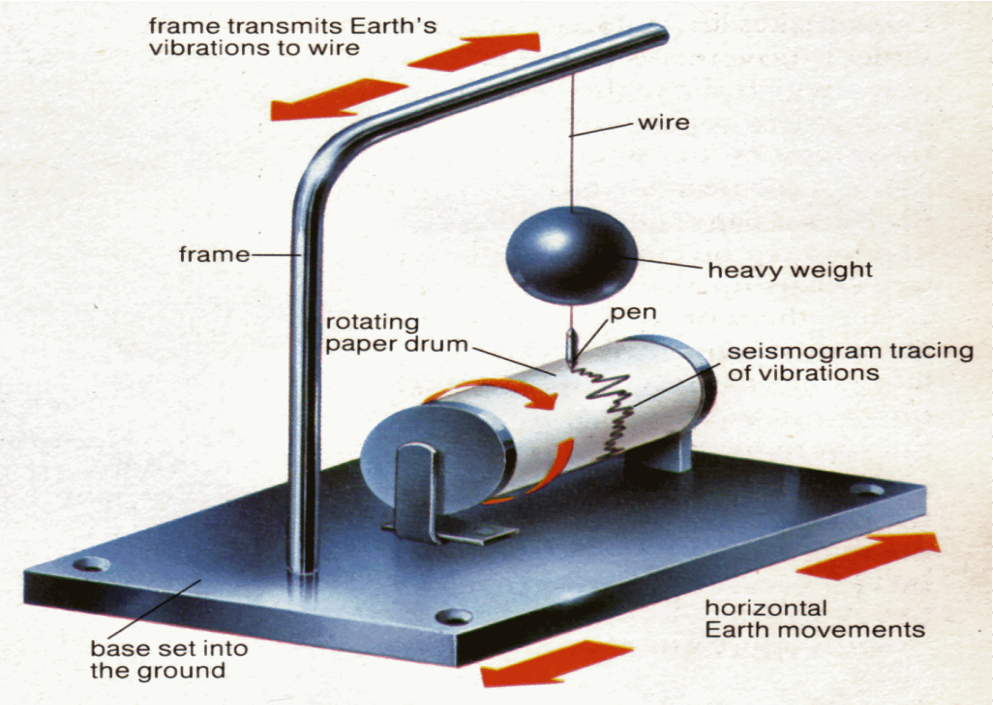
*Confidence Intervals? Mean tests?*

4

**How** do I represent what I have found?

*What is the story you want others to take away.*

# Sample Size:
# Do I have enough / Too much?



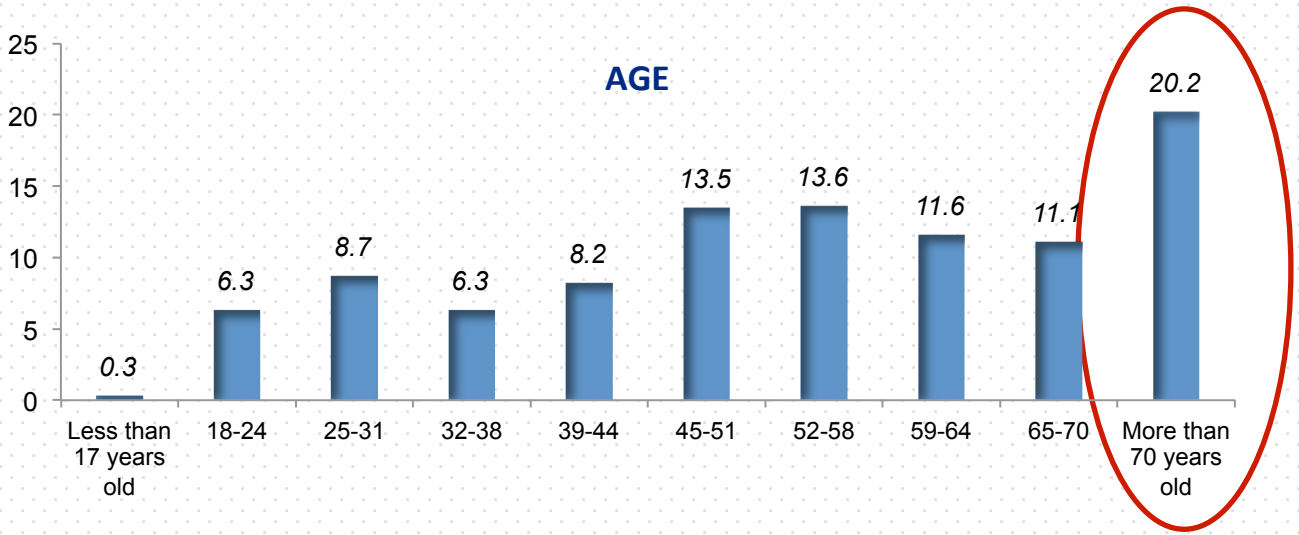

The phenomena we are trying to measure is the earthquake

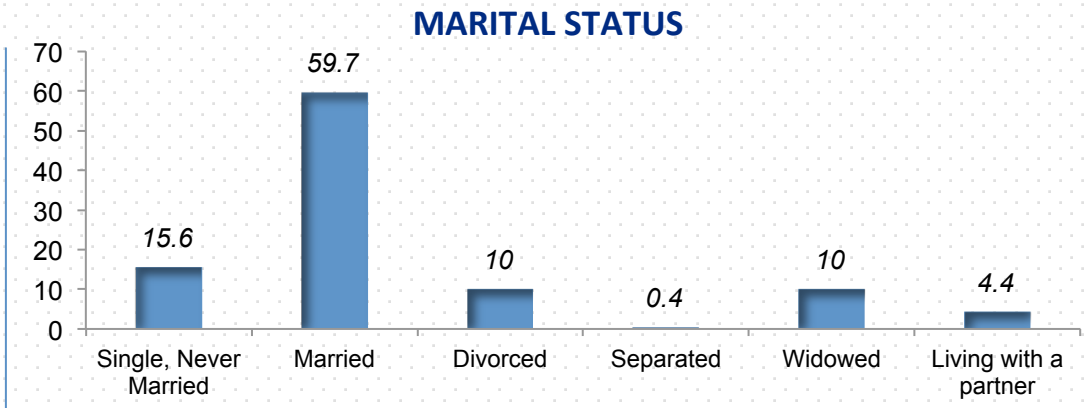
The amount of people/ cases we have determines is our seismograph is more of less sensitive.

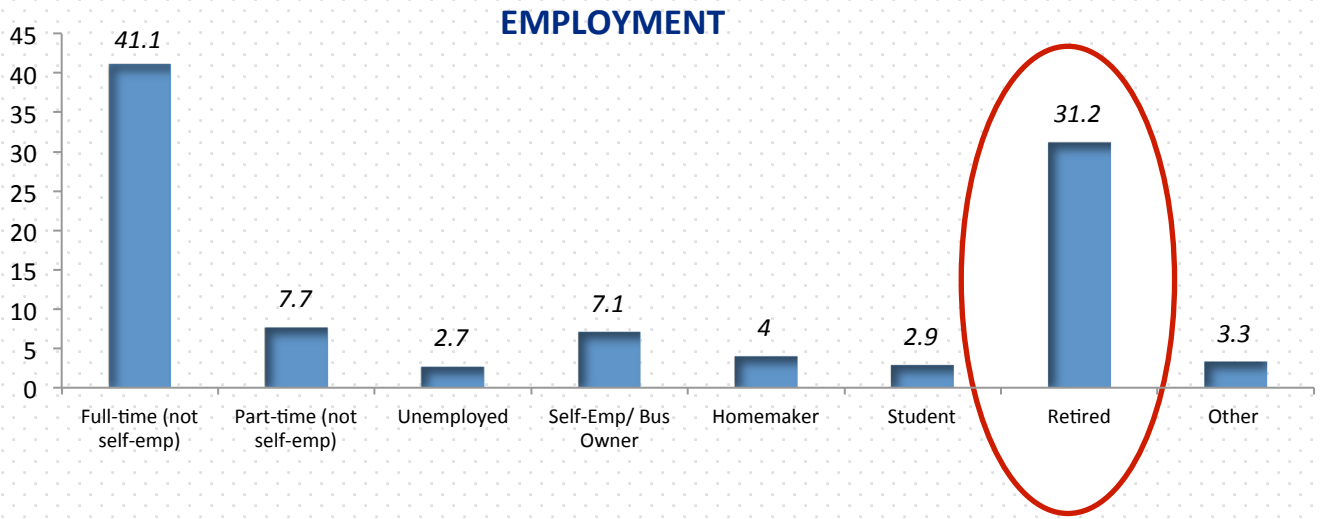
Too FEW cases and we can only pick up on the BIG earthquakes.

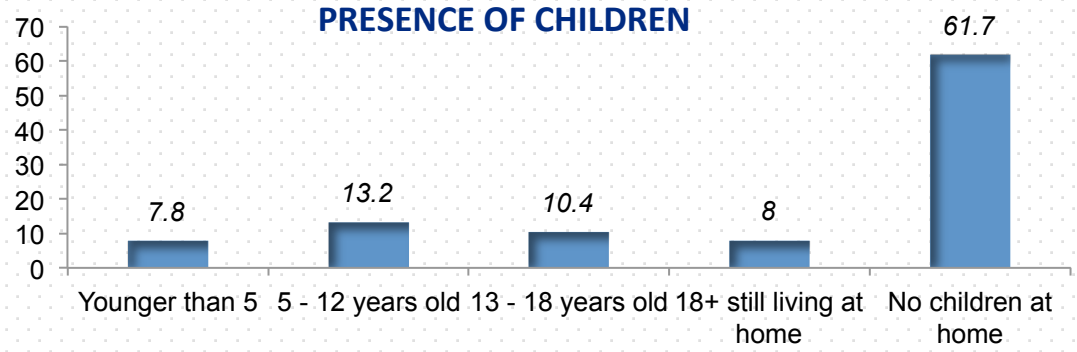
Too MANY cases and every minor bump registers as an earthquake.

# Profile of Respondents




## POLITICAL AFFILIATION

- 30% Democrat
- 20% Republican
- 29% Not affiliated
- 15% Declined to answer
- 6% Other

## VOTING BEHAVIOR

- 81% consider themselves a regular voter
- 17% consider themselves a non-voter
- 2% unsure

# You have data…Now what?
# Steps for approaching your data

1

**Ask yourself**

*How generalizable is this information? Is the sample sufficiently diverse/ large?*

2

**Do some basic frequencies of your data?**

*For every question frequency, mean, standard deviation*

3

**Ask yourself if any significance testing is needed**

*Confidence Intervals? Mean tests?*

4

**How** do I represent what I have found?

*What is the story you want others to take away.*

# Measures of Central Tendency

- Mean … the average score
- Median … the value that lies in the middle after ranking all the scores
- Mode … the most frequently occurring score

# Measures of Central Tendency

Mean … the most frequently used but is sensitive to extreme scores

e.g. 1 2 3 4 5 6 7 8 9 10

Mean = 5.5 (median = 5.5)

e.g. 1 2 3 4 5 6 7 8 9 20

Mean = 6.5 (median = 5.5)

e.g. 1 2 3 4 5 6 7 8 9 100

Mean = 14.5 (median = 5.5)

# Variation or Spread of Distributions

Measures that indicate the spread of scores:

- Range
- Standard Deviation

# Variation or Spread of Distributions

Standard Deviation

- It tells us how much the scores in the data set vary around the mean

- Low standard deviations indicate data clustered close to the mean, high standard deviations indicate the data are spread over a range of values.

*MEANS ARE MEANINGLESS WITHOUT A STANDAD DEVIATION.*

# Standard Deviation

*Scores:*

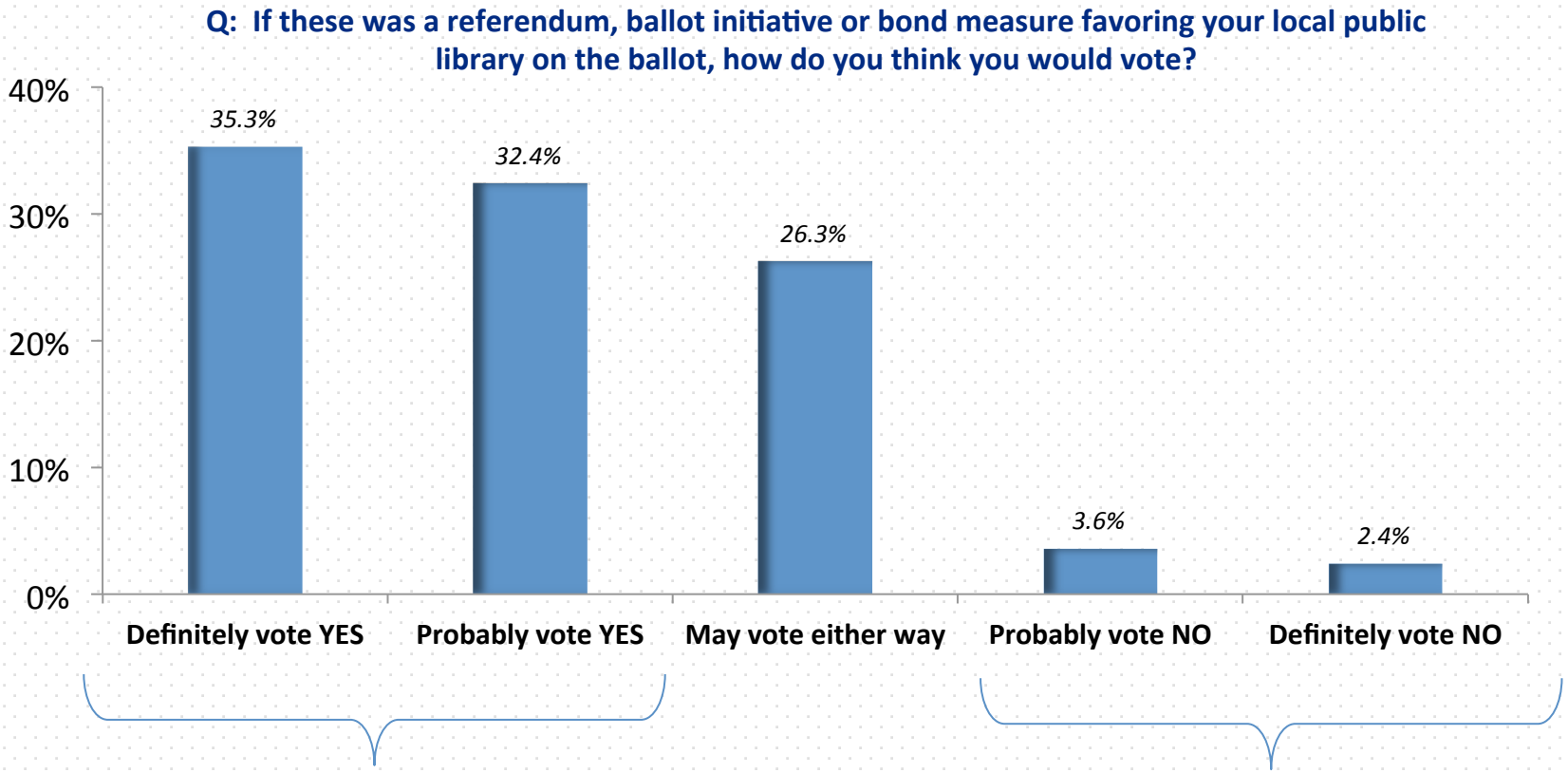
I. e.g. 9 , 10, 11, 7, 13

Mean = 10 (SD = 2.2)

II. e.g. 1, 10, 13, 13, 13

Mean = 10 (SD = 5.2)

# ...and would support their public library in a vote if the opportunity arises.




*Over 2/3 of respondents would likely vote YES*

*Less than 10% of respondents would likely vote NO*

# What would students pay?

*In pilot one price was the most significant concern among students, and it remains a chief concern. However, in looking at comparative pricing most students would pay only about half of what they would pay for a used paper textbook for the Courseload eTextbook. The one exception was the Engineering course. The Engineering students would be willing to pay (on average) the same for an eTextbook and a used textbook,*

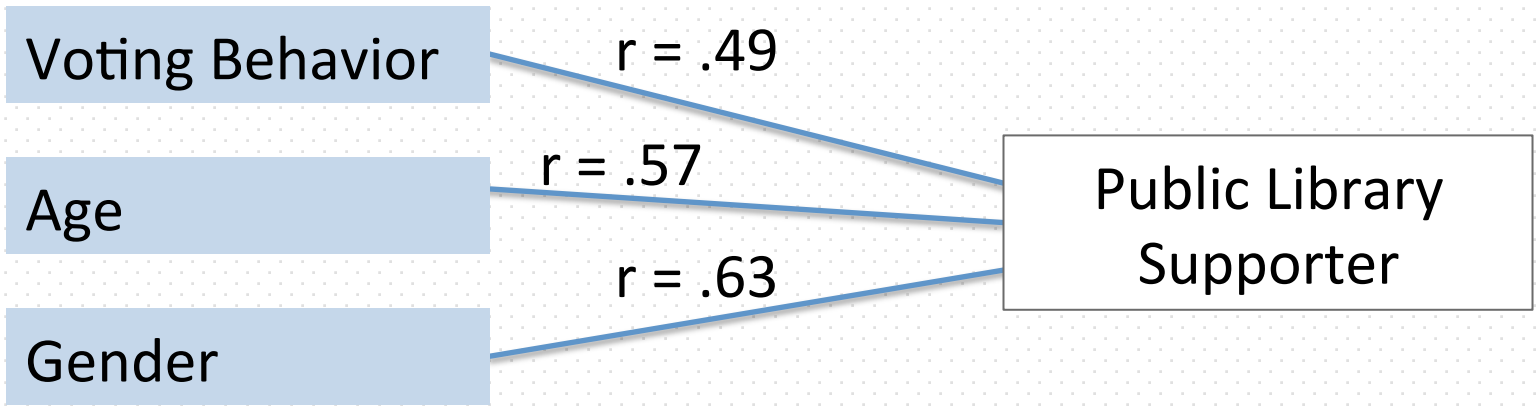
***What would you pay for…?***

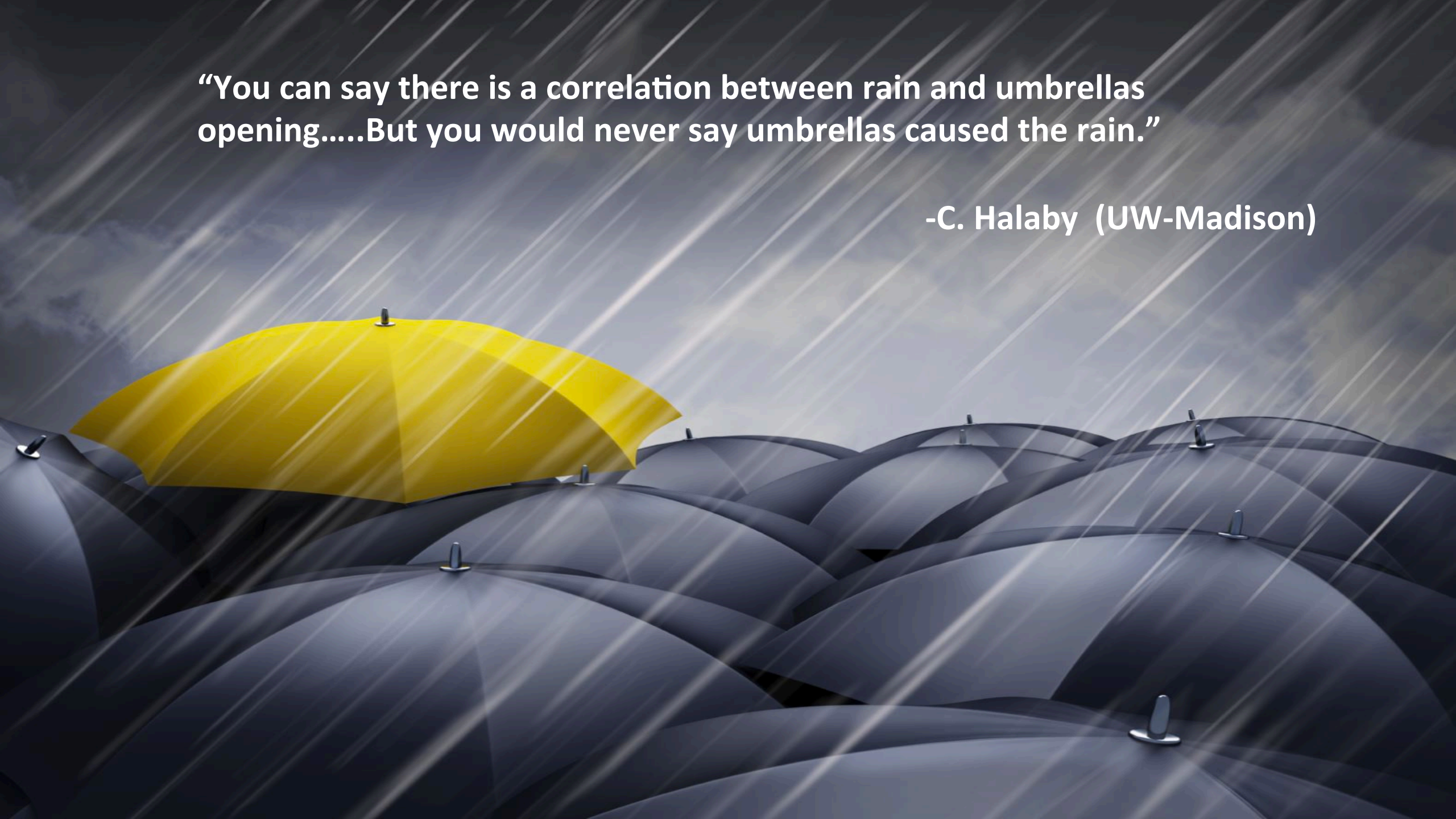
| ***Pilot courses*** | A NEW hard copy paper textbook you used in this class. | A USED hard copy paper textbook you used in this class. | A copy of the eText you used in this class. |
|---|---|---|---|
| ***Business*** | $118.52 | $71.22 | $43.37 |
| ***Educational Psychology*** | $98.38 | $63.00 | $29.00 |
| ***Philosophy*** | $69.77 | $44.84 | $22.16 |
| ***Engineering*** | $93.47 | $35.33* | $35.32* |

***! Note in all mean dollar amounts listed above, Standard deviations were nearly identical. The standard deviation was approximately $10.***

# Correlation

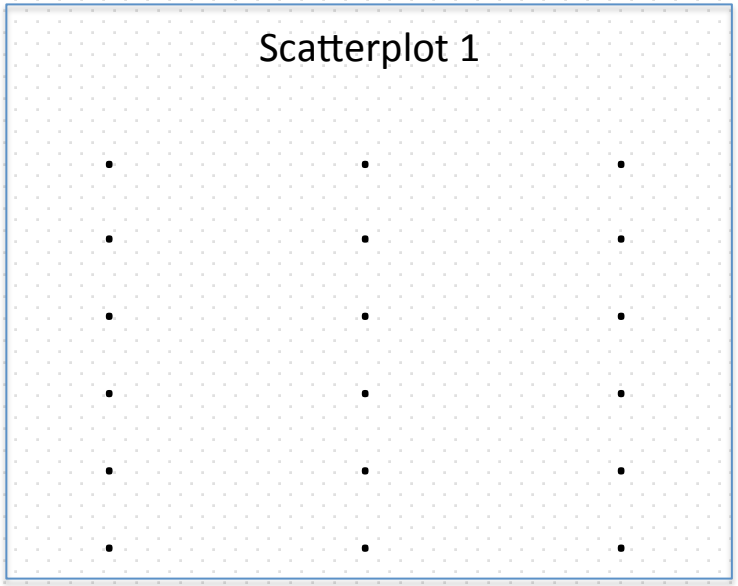
*Correlation is a strength of association (-1 to +1), but it is not predictive. We do not know the cause of the association. Still it can be very helpful.*



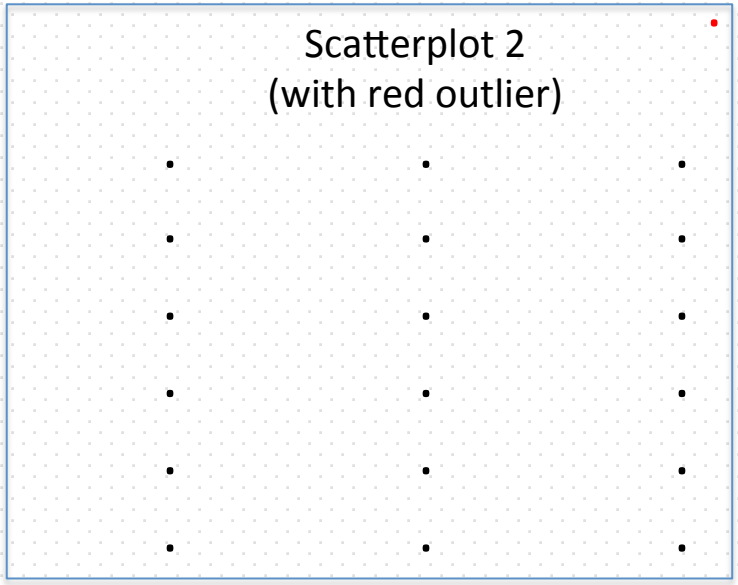

"You can say there is a correlation between rain and umbrellas opening…..But you would never say umbrellas caused the rain."
-C. Halaby (UW-Madison)

# Correlation

*Correlations are useful but they are very susceptible to outlier data points.*




r = 0




r = .80

# You have data…Now what?
# Steps for approaching your data

1

**Ask yourself**

*How generalizable is this information? Is the sample sufficiently diverse/ large?*

2

**Do some basic frequencies of your data?**

*For every question frequency, mean, standard deviation*

3

**Ask yourself if any significance testing is needed**

*Confidence Intervals? Mean tests?*

4

**How** do I represent what I have found?

*What is the story you want others to take away.*

# Confidence Intervals Handout

**Formula for C.I. around a mean:**

**X= mean estimate**

**N= sample size**

**S= Standard Deviation of the mean**

**Z= Desired confidence interval (95% 90%...)**

$$\overline{x} \pm z \frac{s}{\sqrt{n}}$$

Useful info to have when computing a confidence interval:
99% confidence Interval = z score of 2.58
95% confidence Interval = z score of 1.96
90% confidence Interval = z score of 1.65

**This gives an upward bound and a lower bound to the mean estimate. So, if we have a sample with a mean of 8.5, a standard deviation of 1.25, a sample of 1,000 and we want to be 95% confident in the interval we derive, we get the following:**

**1.96 (1.25/ sqrt-1000) = .08**

**We add and subtract this to the mean: 8.5 + .08 = 8.58 & 8.5 -.08 = 8.42**

**So we can say that we are 90% confident that the "true" estimate falls between 8.42 and 8.58**

# Confidence Intervals to Compare Means

**Using the previous example interval of 8.42 to 8.58, let's look at two other datasets that measured the same construct (i.e. satisfaction)**

Comparison Set 1:

**X= 6.5**

**N= 300**

**S= 2.5**

**Z= 90%**

Upper CI = 6.78

Lower CI = 6.22

Comparison Set 2:

**X= 8.25**

**N= 100**

**S= 1.5**

**Z= 90%**

Upper CI = 8.54

Lower CI = 7.96

In the first example at the right, we see that 6.78 -6.22 is an interval that does not overlap with 8.42 – 8.58. So we can say with 95% certainty that these intervals are significantly different.

In the second example, we see that 7.96 – 8.54 DOES overlap with 8.42 – 8.58. So we can say with 95% confidence that there is no difference between these two findings

# Confidence Intervals Around Percentages

**$P_{\%}$ +/- Z [$P_{\%}$ (1- $P_{\%}$) / Sqrt of N]**

*So, if we want to put a 95% CI around a finding that 40% of 275 people rated the taste of a new cola as "Very Good":*

**1.96 [.40 (1-.40) / sqrt of 275 = .40(.60)/ 16.58 = .24/16.58 = .015] = 1.96 * 0.15 = .029**
**.029 +/- .40**

We add .029 to.40 and get upper value = **.429** / We subtract from .40 and get a lower value = **.371.**

So we can now say that we are 95% confident that the "true" percentage is between 37% and 43%

# You have data…Now what?
# Steps for approaching your data

1

**Ask yourself**

*How generalizable is this information? Is the sample sufficiently diverse/ large?*

2

**Do some basic frequencies of your data?**

*For every question frequency, mean, standard deviation*

3

**Ask yourself if any significance testing is needed**

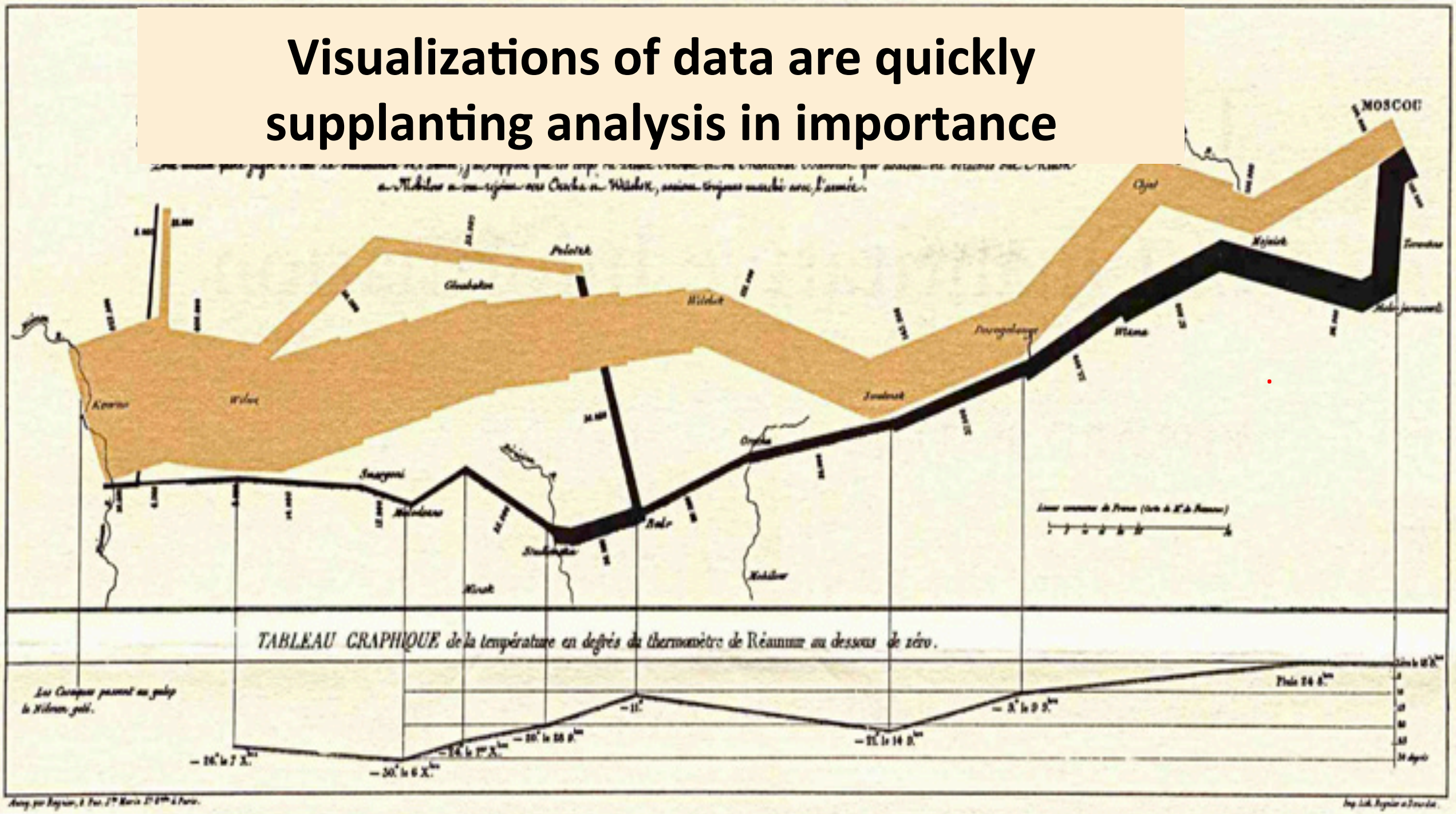
*Confidence Intervals? Mean tests?*

4

**How** do I represent what I have found?

*What is the story you want others to take away.*

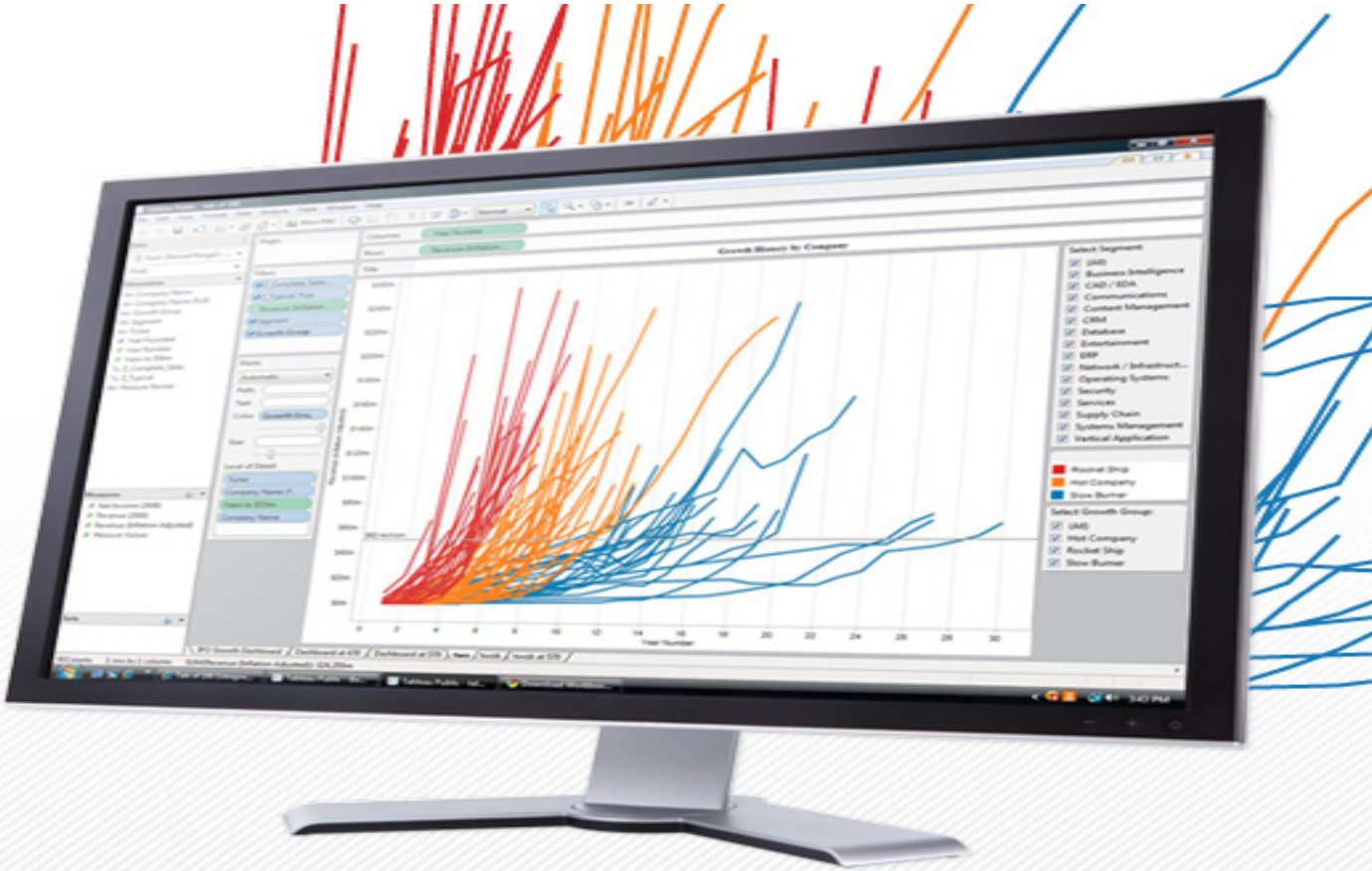
# Visualizations of data are quickly supplanting analysis in importance



# New Software Data Visualization Tools

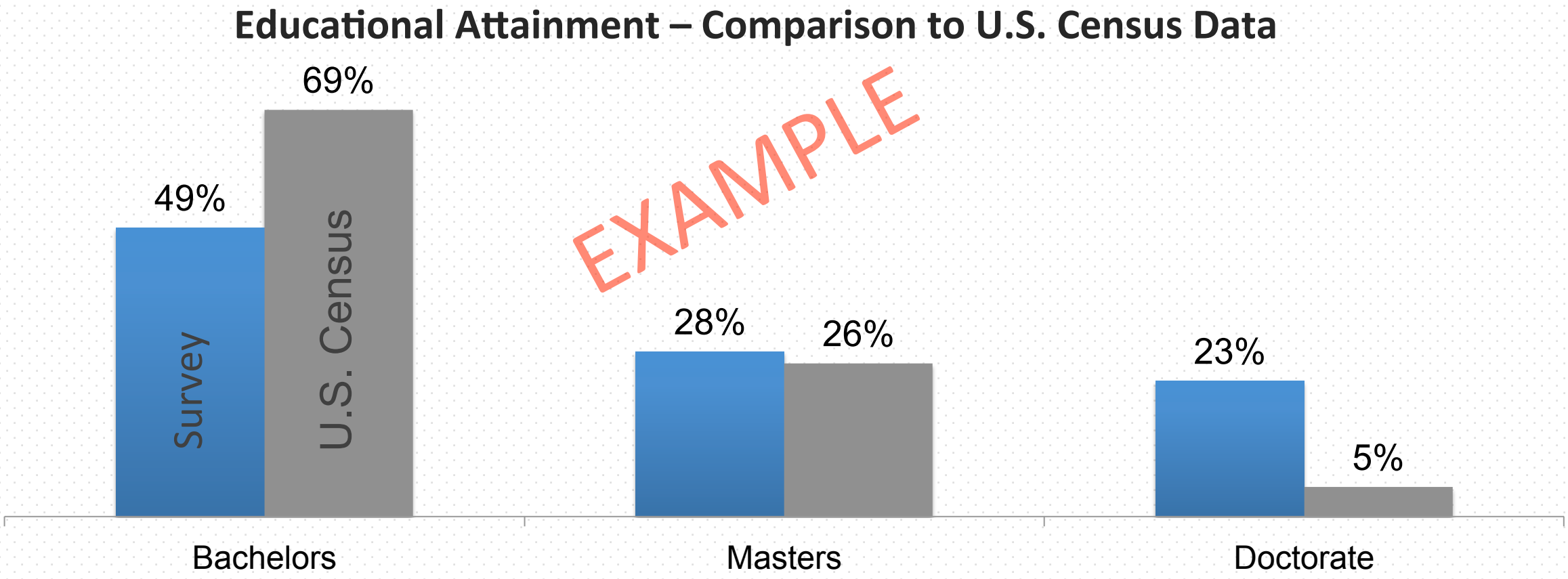
Tableau Desktop

Business analytics anyone can learn and use.

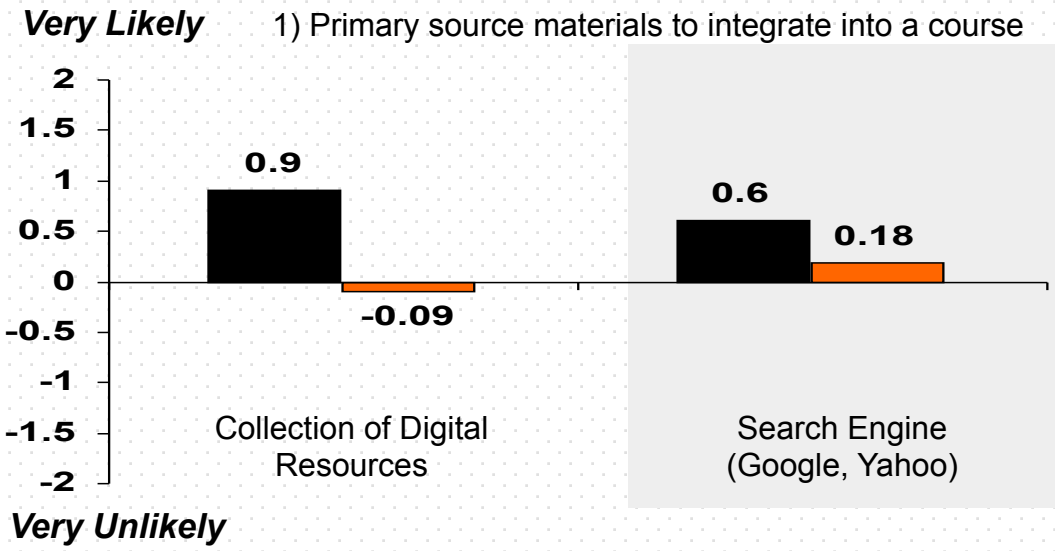


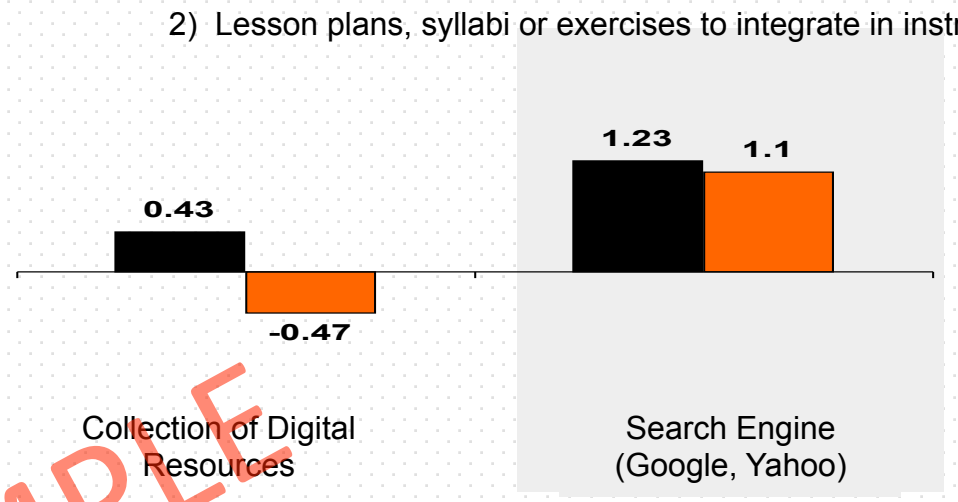
Increasingly people want to be shown the conclusions—not told the conclusions.

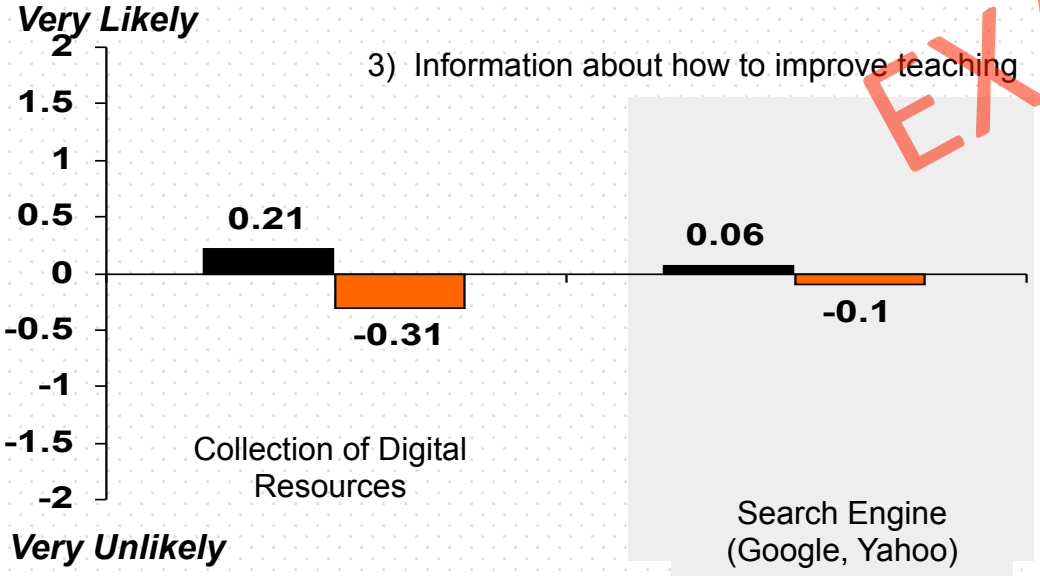
# Think about Contrast or Comparison Points

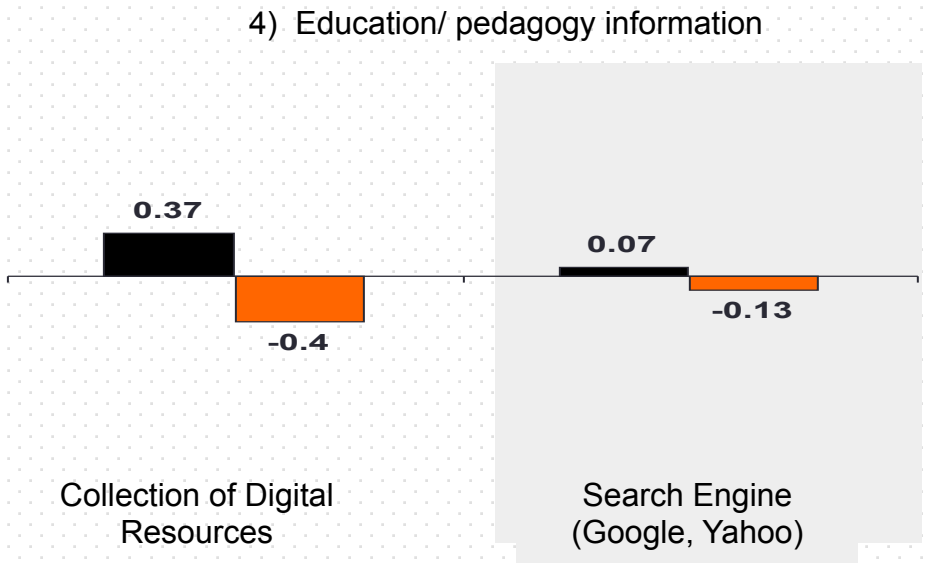


# Think about Information Density (handouts vs slides)



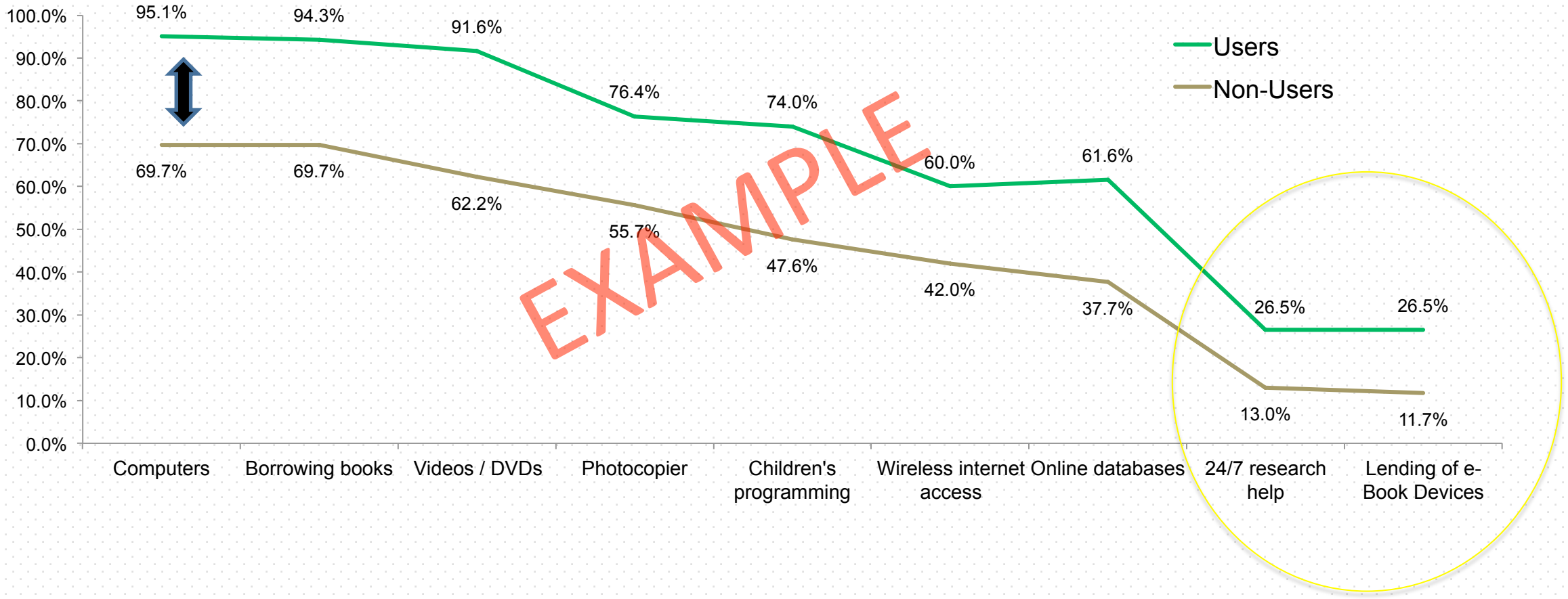




Faculty Overall

Math Faculty

# Think about the Story you want to tell with a Visualization

**Across non-users, there appears to be an opportunity to increase awareness of what public libraries offer.**



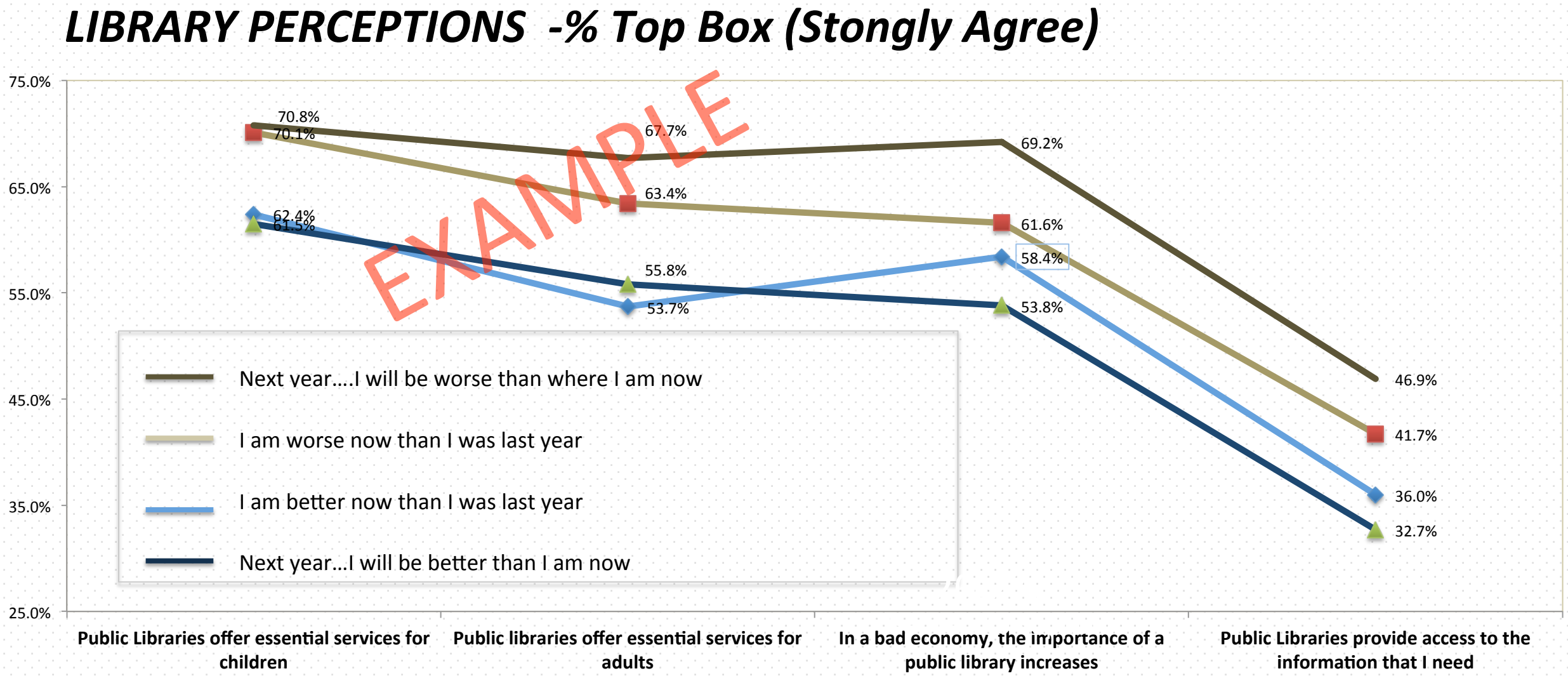
# Visualizations work best to show complex (sometimes subtle) relationships

## *LIBRARY PERCEPTIONS -% Top Box (Stongly Agree)*

| | Essential services for children | Essential services for adults | In a bad economy the importance of libraries increases | Provide access to the information that I need |
|---|---|---|---|---|
| **I will am worse now than I was last year** | 70% | 63% | 62% | 42% |
| **I am better now than I was last year** | 62% | 54% | 58% | 36% |

| | Essential services for children | Essential services for adults | In a bad economy the importance of libraries increases | Provide access to the information that I need |
|---|---|---|---|---|
| **Next year, I will be worse than I am now** | 71% | 68% | 69% | 47% |
| **Next year, I will be better than I am now** | 62% | 56% | 54% | 33% |

# Visualizations work best to show complex (sometimes subtle) relationships



# Questions

*Thank You*

**Joshua H. Morrill, PhD**
joshua@morrillsolutions.com
608-588-2874