Introduction to Social Science Research and Survey Design & Analysis

ENG 7360 Topics in Rhetoric

Prof. Mya Poe

Fall 2022

Taught By: Juniper Johnson & Ana Abraham

Northeastern University

NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Workshop Objectives

- Understand key fundamentals of social science research
- Understand the basic concepts of survey design and issues to avoid
- Understand key interpretative methods for analyzing survey data
- Explore and think critically about bias in survey design, analysis, and interpretation, and data display

Slides, handouts, and sample survey available at:
https://bit.ly/fa22-poe-surveydesign

Feel free to ask questions at any point during the presentation!

Fundamentals of Social Research

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Brief Introduction to Social Science Research

- The object of social study is the social world, including **human behavior, relations, and systems**—political, economic, religious, and social.
- In large part, social research aims to find **patterns of regularity** in social life through theory and empirical observation.
- The focus is often on **aggregate phenomena** (of groups, organizations, collectives, etc.) rather than individuals

Feel free to ask questions at any point during the presentation!

Examples of Research Questions

- Do women earn less money than men and, if so, why?
- What causes support for or opposition to same-sex marriage?
- Which academic departments offer the broadest degree of liberal arts training?
- Is it true, as some have argued, that the United States was established as a “Christian nation”?
- Do writing assessments lead to learning improvements?

Feel free to ask questions at any point during the presentation!

The Use of Variables in Social Research

Variables are sets of attributes or characteristics. For example, proficiency in reading may be a variable used by education researchers. This variable might take on different values (or attributes), depending on how it is designed and measured. This might be a scale from 1–10, a qualitative measurement of "satisfactory" or "unsatisfactory", and so on.

Another example of a variable is occupation. This variable's attributes might be *teacher*, *doctor*, *farmer*, *PhD student*, etc.

Feel free to ask questions at any point during the presentation!

Common Errors in Inquiry

- Inaccurate Observations
  - Poor variable construction, measurement error
- Overgeneralization
  - Assuming that what is true for some case studies is true for all case studies
- Selective Observation
  - Focusing only on cases that conform to a predefined pattern
- Illogical Reasoning
  - Making erroneous conclusions based on observation

Feel free to ask questions at any point during the presentation!

Surveys: Introduction to Design

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Discuss!

- Have you encountered any surveys recently? Either through your research, studies, or in your daily life.

- Think back to one recent survey you have encountered. What questions stood out? Were there any misleading, overly technical, or biased questions?

Feel free to ask questions at any point during the presentation!

Research Design

- Social science research can be conducted with the purpose of exploration, description, or explanation (sometimes, more than one of these).
- The purpose of the research should influence the research design. There are many ways to collect data on social phenomena, including surveys, experiments, and unobtrusive research like content analysis.

Feel free to ask questions at any point during the presentation!

Surveys Have a Long History in Research

Surveys are a very old research technique. Censuses, a type of survey, were used as far back as Ancient Egypt and Rome.

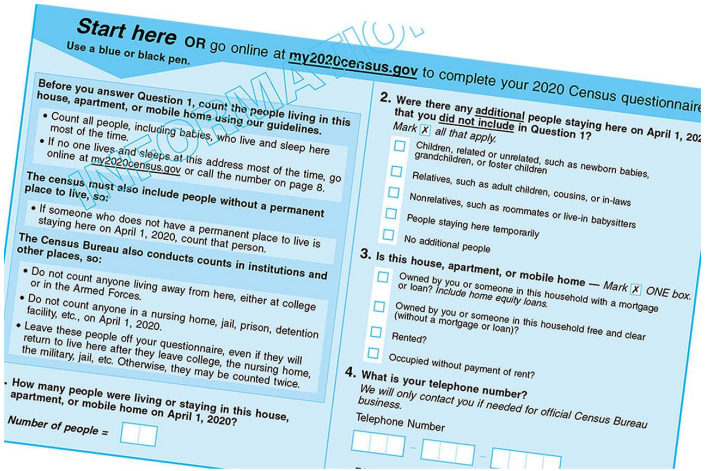
Today, surveys are commonly used in social research for descriptive, exploratory, and explanatory purposes.

"Every fourteen years the inhabitants of Egypt were required to submit a declaration to local authorities containing the names, ages, and other identifying information of all co-residents."

W. Graham Claytor and Roger S. Bagnall in *Greek, Roman, and Byzantine Studies* (2015)

Feel free to ask questions at any point during the presentation!

Some Well-Known Surveys



Start here OR go online at my2020census.gov to complete your 2020 Census questionnaire

Use a blue or black pen.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.
- If no one lives and sleeps at this address most of the time, go online at my2020census.gov or call the number on page 8.

The census must also include people without a permanent place to live, so:

- If someone who does not have a permanent place to live is staying here on April 1, 2020, count that person.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away from here, either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2020.
- Leave these people off your questionnaire, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2020?

Number of people =

2. Were there any additional people staying here on April 1, 202 that you did not include in Question 1?

Mark ☒ all that apply.

- ☐ Children, related or unrelated, such as newborn babies, grandchildren, or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in babysitters
- ☐ People staying here temporarily
- ☐ No additional people

3. Is this house, apartment, or mobile home — *Mark ☒ ONE box.*

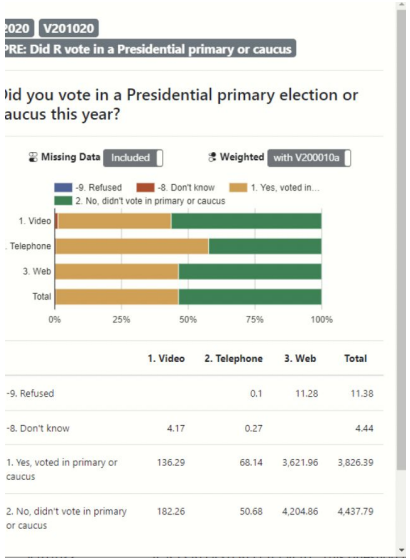
- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

4. What is your telephone number?

We will only contact you if needed for official Census Bureau business.

Telephone Number

US 2020 Census



	1. Video	2. Telephone	3. Web	Total
-9. Refused		0.1	11.28	11.38
-8. Don't know	4.17	0.27		4.44
1. Yes, voted in primary or caucus	136.29	68.14	3,621.96	3,826.39
2. No, didn't vote in primary or caucus	182.26	50.68	4,204.86	4,437.79

America National Elections Studies survey

A-17. **[HAND CARD.]** How difficult is it for you to **(item)** written in English?

HAND CARD 1

Would you say you have...

	No difficulty	Some difficulty	Moderate difficulty	or a	Great deal of difficulty	NEVER TRIED
A. Understand a utility bill, such as telephone or electric?	1	2	3		4	5
B. Understand the dosage information on over-the-counter medicines?	1	2	3		4	5
C. Look up information in dictionaries, encyclopedias, phone books, or other reference books?.....	1	2	3		4	5

A-18. **[HAND CARD.]** How difficult is it for you to fill out forms in English, such as at the doctor's office or at school? Would you say you have...

HAND CARD 1

A. No difficulty, ...	1
B. Some difficulty, ...	2
C. Moderate difficulty, or a.....	3
D. Great deal of difficulty?	4
E. NEVER TRIED	5

The NCES 2003 National Assessment of Adult Literacy

Feel free to ask questions at any point during the presentation!

Survey Design: Research Ethics

While doing research involving human subjects, always:

1. Get IRB (Institutional Review Board) approval
2. Find the right, secure survey tool to use; Northeastern uses Qualtrics, which is secure for keeping your results private
3. Understand that your specific questions will gather specific answers. Try not to steer your participants in a particular direction, and always be wary of the vulnerabilities of the demographic(s) you're studying.

Feel free to ask questions at any point during the presentation!

Survey Design: Types of Questions

Multiple choice: questions that only have specific answers and the user can only click one. For example, “What is your favorite physical exercise?”

Likert-scale: a type of response scale in which respondents can specify their level of agreement, importance or satisfaction typically in 5 points, (1) strongly agree to (5) strongly disagree.

Checkboxes: questions that only have specific answers and the user can click multiple. For example, “select all the physical exercises you did last week”.

Linear Scale: questions that invite users to choose from a lower to higher number that match their experience. Typically on a scale from 1 to 10; allows a more granular measure of affect and participants are able to express their degrees of response.

Paragraph/Short Answer: open-ended questions that usually invite longer prose. For example, “Describe your favorite gym exercise and how it makes you feel”.

Feel free to ask questions at any point during the presentation!

Questions to Avoid

- “On a scale from 1 to 10 please rate the discounted semester pass program.”
    - Technical or industry jargon. Your respondents may or may not be aware of the program or, perhaps, it’s the first time they've heard this title.
- “Do you think that many programs that are currently reviewed by the university administration, like the expansion of bike lanes on campus, would be environmentally friendly and affordable?”
    - Vaguely worded and double-barreled (contains different issues or topics within one question- “environmentally friendly and affordable”)

Feel free to ask questions at any point during the presentation!

Questions to Avoid (cont'd)

- "Do you think the inept university administration should for once finally do something about improving our campus commuting experience? "
  - Try to avoid biased judgmental language and anything that your respondents might find off-putting!

- How long does it take you to commute to campus on a typical day?

- Less than 10 minutes
- 11-20 minutes
- 15-35 minutes
- 25-40 minutes
- 41-50 minutes
- 51-60 minutes
- More than one hour

Overlapping Answers are a Common Mistake

Feel free to ask questions at any point during the presentation!

Questions to Avoid (cont'd)

- Do you always take the MBTA when you commute to campus?
  Choices: Always, Sometimes, Rarely, Never
  - "**Always**" is confusing. The feedback from this question will not be very useful.
- Does the MBTA not provide adequate resources for commuters? Choices: Yes/No
  - Double-negatives are incredibly confusing for survey takers
- Mismatched answer choices, random questions that do not fit thematically and many other problems that can be fixed by reading the questions out loud or to peers!

Feel free to ask questions at any point during the presentation!

Surveys: Analysis and Interpretation

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Strengths and Weaknesses of Surveys

- Surveys are particularly useful in **describing the characteristics of a large population.**
- Survey administration is **flexible** (in person, by phone, online), which allows researchers to collect data from large samples.
- Standardization leads to **high reliability.**
- But, the standardized nature of many surveys can lead to questions set for the **"lowest common denominator"**
- Similarly, surveys can **miss important context** and provide superficial cover of complex topics.
- The artificiality of surveys can lead to **issues in validity.**

Feel free to ask questions at any point during the presentation!

Population and Sample

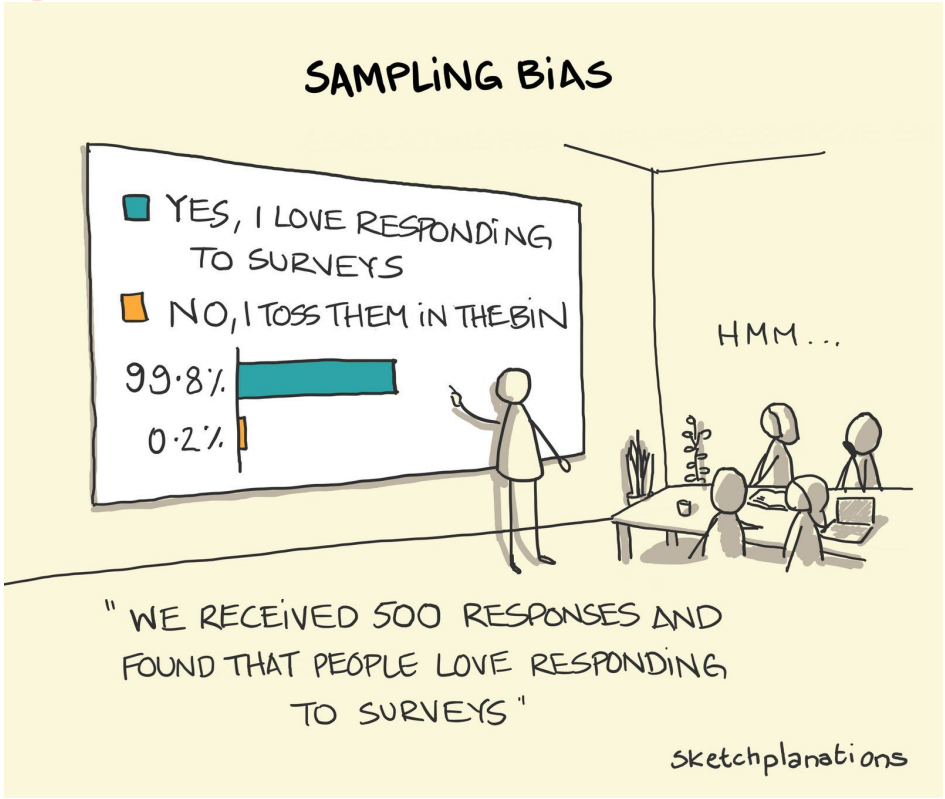
The **population** is the group we are interested in studying. This can be *Americans*, *middle school children*, *Republican voters*, or any other theoretically specified group.

From there, researchers define the **study population,** from which they will draw a **sample**.

A sample is a sub-group within the study population, which allows researchers to make inferences about the population.

Feel free to ask questions at any point during the presentation!

Sampling: Beware of Bias



Source: sketchplanations

Secondary Analysis of Surveys

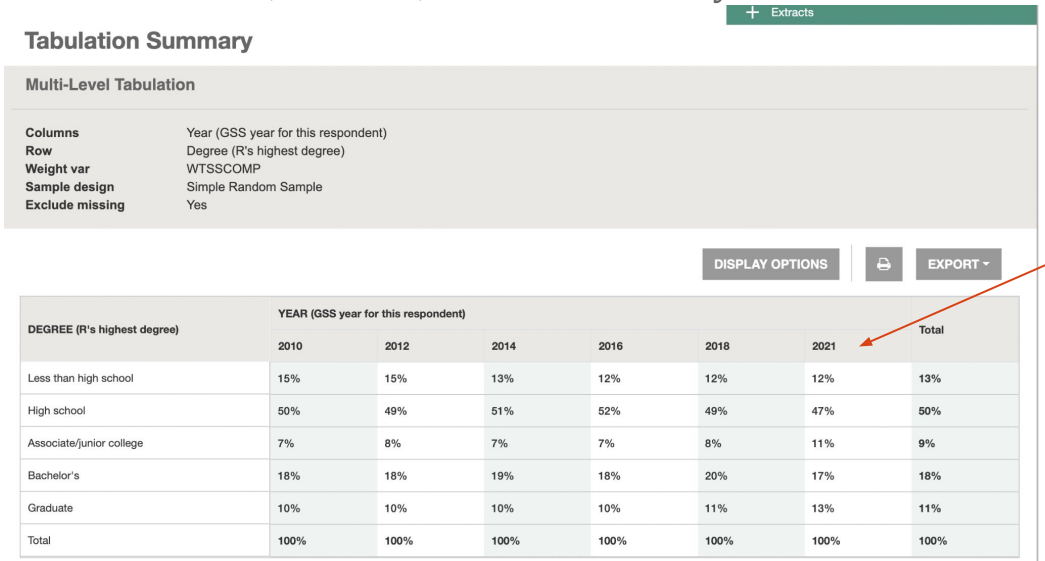
Oftentimes, researchers use data collected by others for their research. This is called **secondary analysis**, which occurs when data collected and processed by one researcher are reanalyzed—often for a different purpose—by another.

One major source of survey questions for secondary analysis are the General Social Surveys.

Feel free to ask questions at any point during the presentation!

Example: Data From General Social Surveys

1. Go to https://gss.norc.org/ and create a free account.
2. Under my GSS, select variables of interest (organized under tags, such as “Education”)
3. Create a tabulation, or table, and choose a way to show results



+ Extracts

Tabulation Summary

Multi-Level Tabulation

Columns	Year (GSS year for this respondent)
Row	Degree (R's highest degree)
Weight var	WTSSCOMP
Sample design	Simple Random Sample
Exclude missing	Yes

DISPLAY OPTIONS | EXPORT

DEGREE (R's highest degree)	YEAR (GSS year for this respondent)						Total
	2010	2012	2014	2016	2018	2021	
Less than high school	15%	15%	13%	12%	12%	12%	13%
High school	50%	49%	51%	52%	49%	47%	50%
Associate/junior college	7%	8%	7%	7%	8%	11%	9%
Bachelor's	18%	18%	19%	18%	20%	17%	18%
Graduate	10%	10%	10%	10%	11%	13%	11%
Total	100%	100%	100%	100%	100%	100%	100%

Column percentages of the highest educational degree attained by respondents.

In 2021, 17% of respondents had attained a bachelor’s degree as their final degree.

Feel free to ask questions at any point during the presentation!

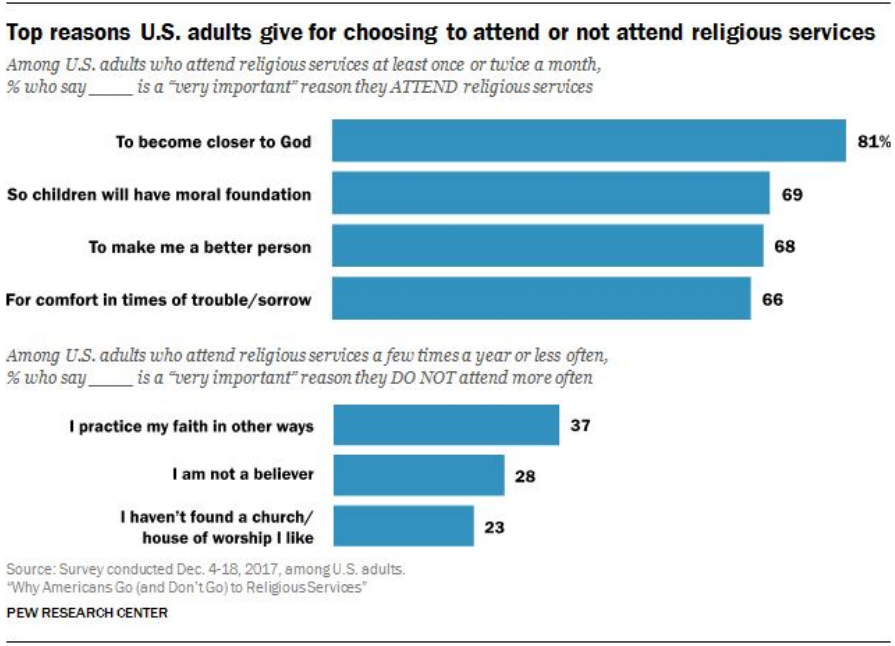
Making Sense of Data

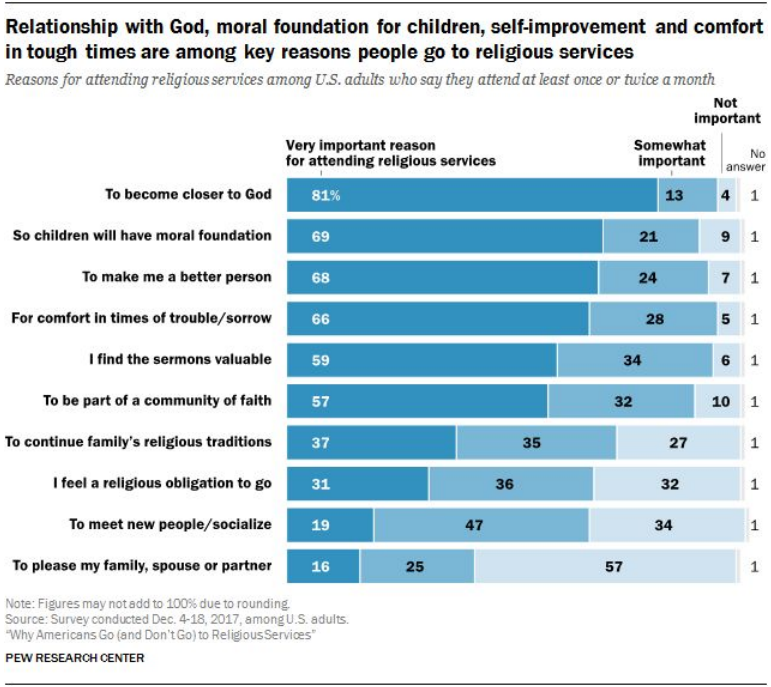
Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Displaying Survey Data

Findings from surveys are often turned into numerical data for analysis and presented in graphs, charts, and tables.



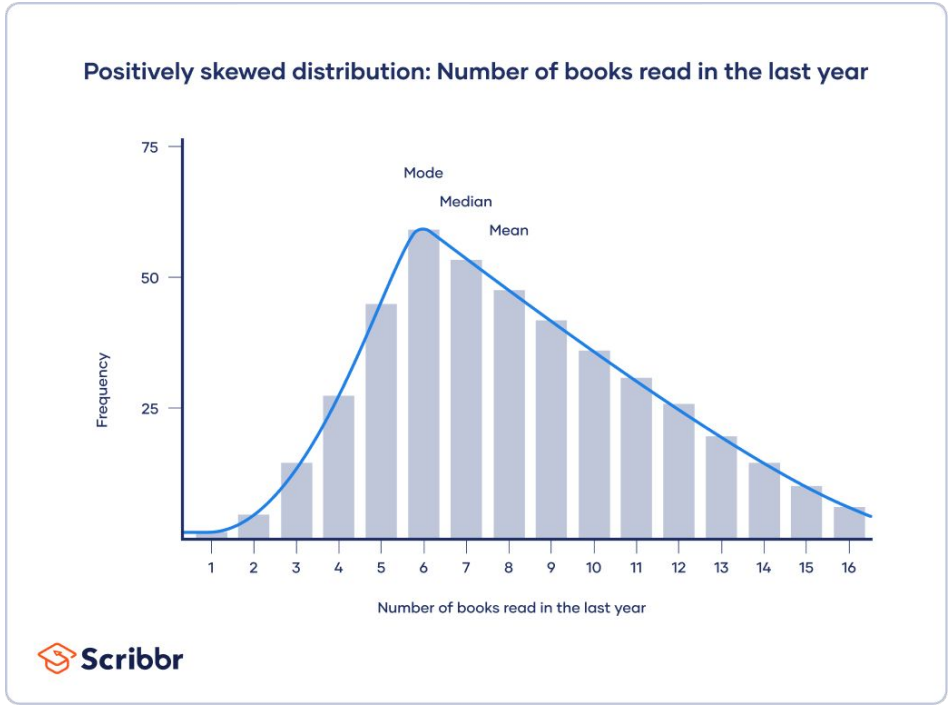


Descriptive Statistics: Central Tendency

Measures of **central tendency** (mean, median, mode) indicate **typical values** in the data.

But be cautious of skewness in the data: in this example, the **mean** (sum of values/number of observations) is higher than the **mode** (value with the highest number of observations) and **median** (value of the "middle" case in a rank-ordered set of observations.)

Why? A few people who read many books per year are dragging the mean up. Hence, the distribution is positively skewed. But which measure here most closely reflect the "typical" value?



Inferential Statistics: Association

A **measure of association** is a numerical value that tells us how strongly related two variables are. For nominal variables, the Chi Square is often used. Here's what it looks like:

Expected frequencies reflect a scenario where there is no association between gender and church attendance.

These are the actual survey responses

Computations of Chi Square

A Hypothetical Illustration of Chi Square

I. Expected Cell Frequencies	*Men*	*Women*	*Total*
Attended church	28	42	70
Did not attend church	12	18	30
Total	40	60	100
II. Observed Cell Frequencies	*Men*	*Women*	*Total*
Attended church	20	50	70
Did not attend church	20	10	30
Total	40	60	100
III. (Observed – Expected)² ÷ Expected	*Men*	*Women*	
Attended church	2.29	1.52	$\chi^2 = 12.70$
Did not attend church	5.33	3.56	$p < .001$

The Chi Square measure is statistically significant. **There is an association between gender and church attendance**

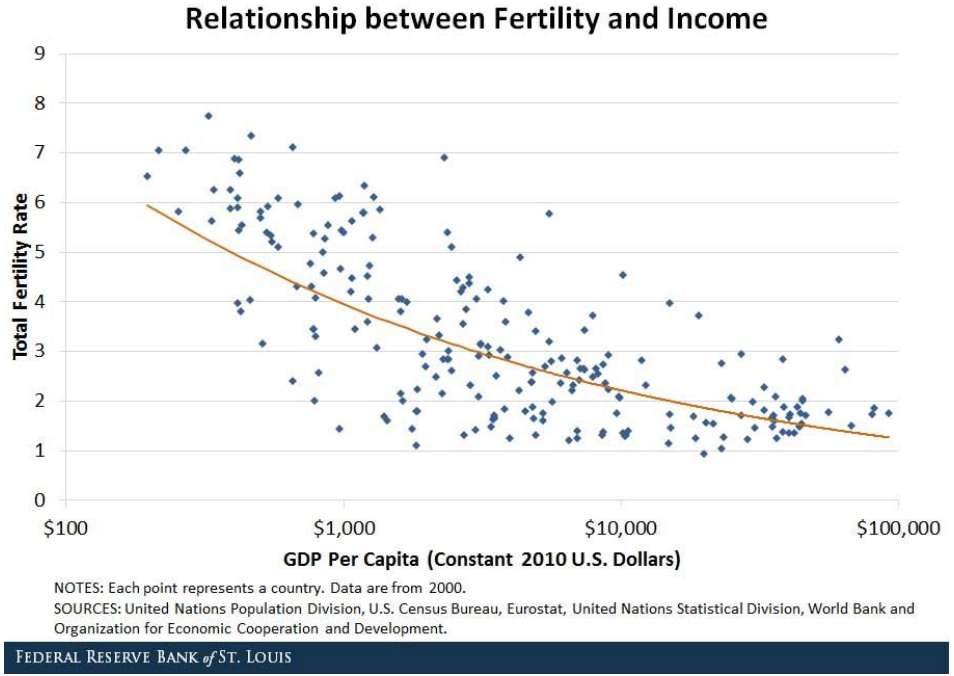
Source: Babbie, 2014

Feel free to ask questions at any point during the presentation!

Inferential Statistics: Correlation

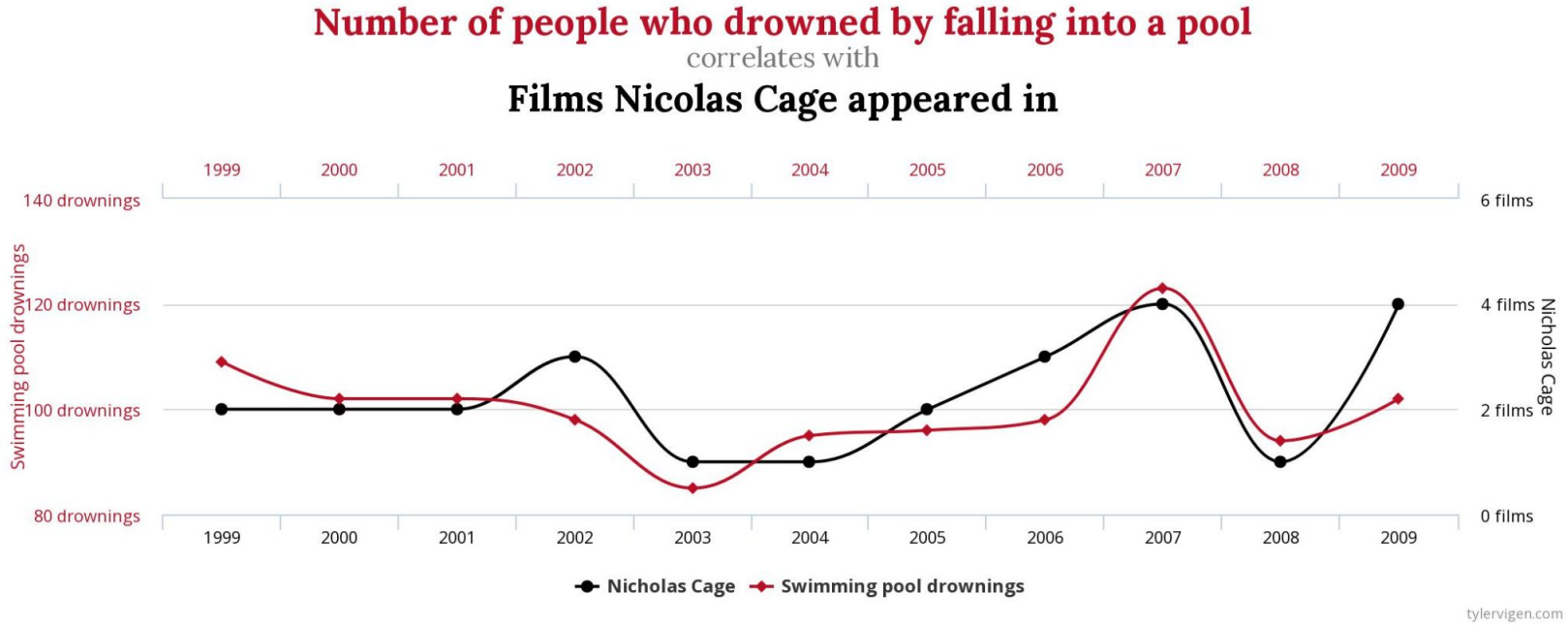
Correlation pertains to an empirical relationship between two variables such that (1) changes in one are associated with changes in the other, or (2) particular attributes of one variable are associated with particular attributes of the other.

This graph shows a well-documented relationship between a country's income and fertility rate.



Feel free to ask questions at any point during the presentation!

Correlation Does Not Mean Causation!



Source: Data from Centers for Disease Control & Prevention and Internet Movie Database, calculations by Tyler Vigen.

Feel free to ask questions at any point during the presentation!

Representing Data in Visuals and Infographics

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Misrepresentation of Data

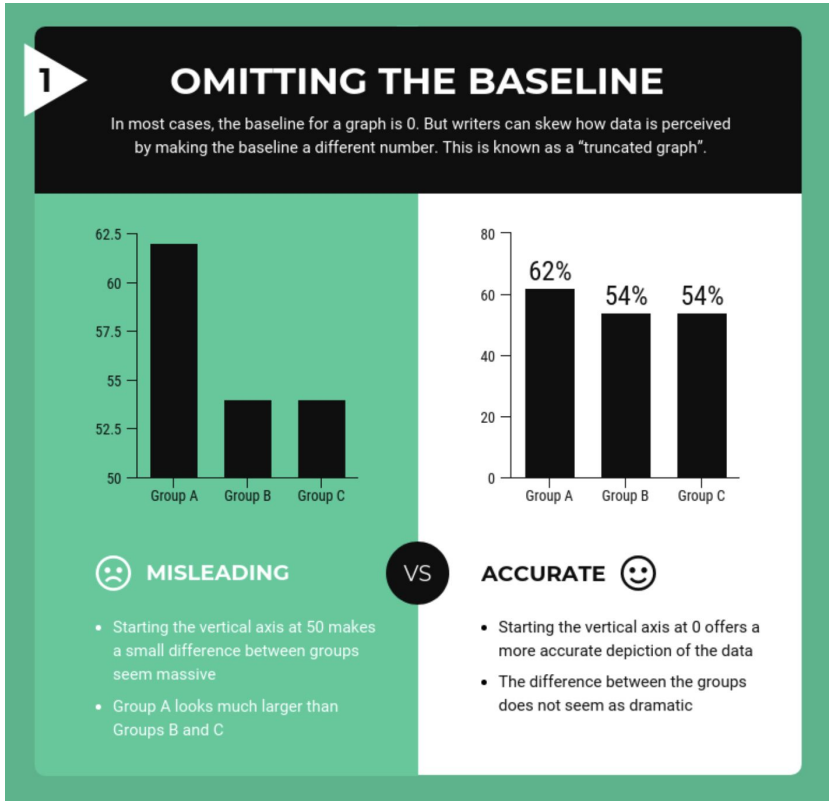
From D.B. Resnik, in International Encyclopedia of the Social & Behavioral Sciences, 2001:

"*The concept of 'misrepresentation,' unlike 'fabrication' and 'falsification,' is neither clear nor uncontroversial. Most scientists will agree that fabrication is making up data and falsification is changing data.* ***But what does it mean to misrepresent data? As a minimal answer to this question, one can define 'misrepresentation of data' as 'communicating honestly reported data in a deceptive manner.***"

- This **online book from The Data School** covers some common ways data could be misrepresented at multiple points in the process of gathering, analyzing, and presenting findings on data-based research.

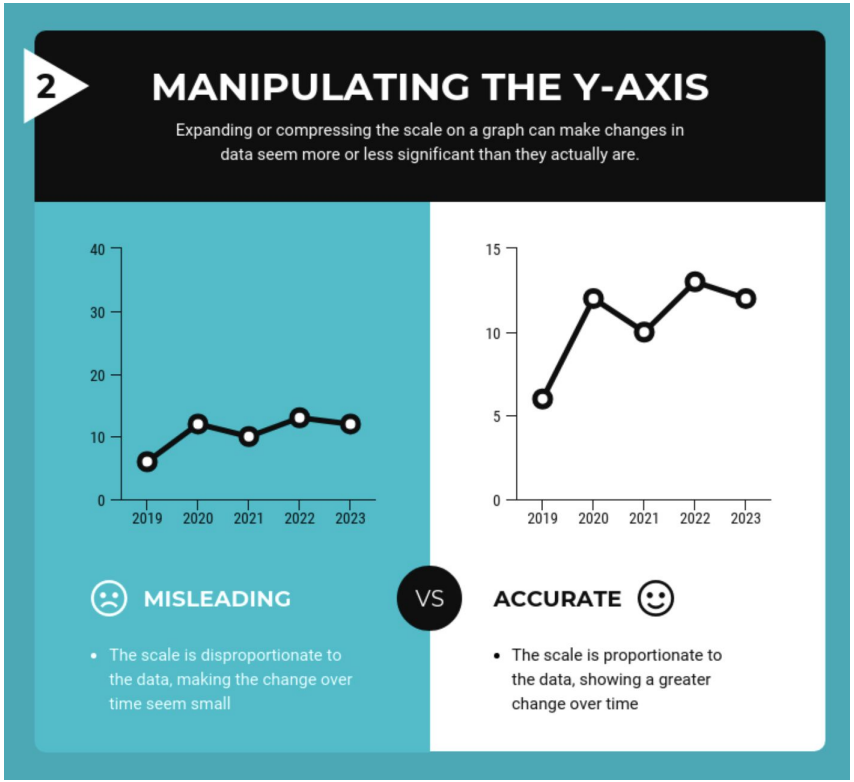
Feel free to ask questions at any point during the presentation!

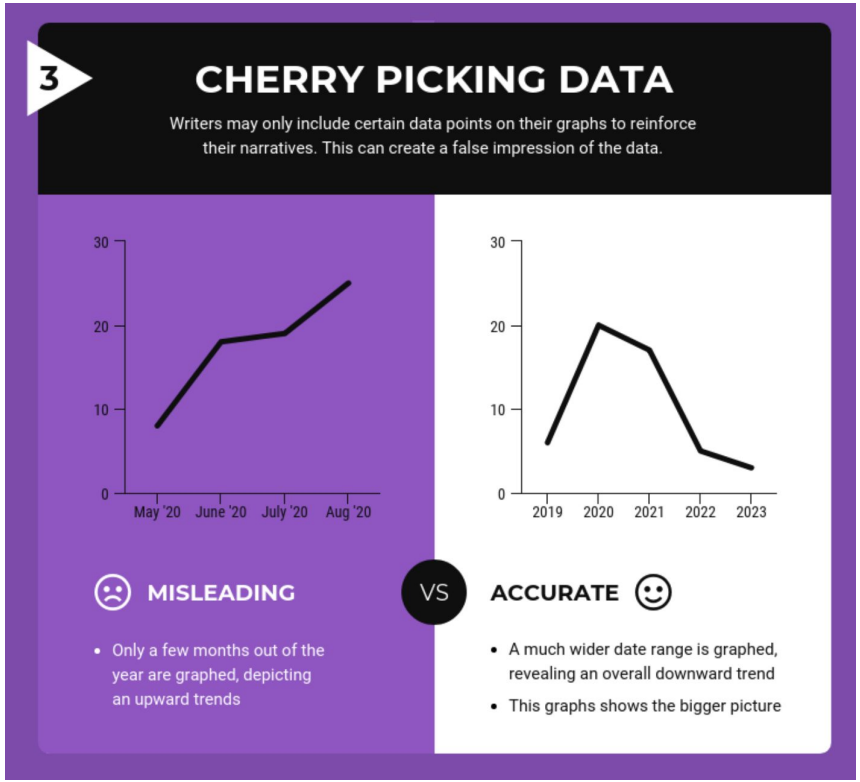
Limitations of Charts, Diagrams, Graphs, & Maps



Feel free to ask questions at any point during the presentation!

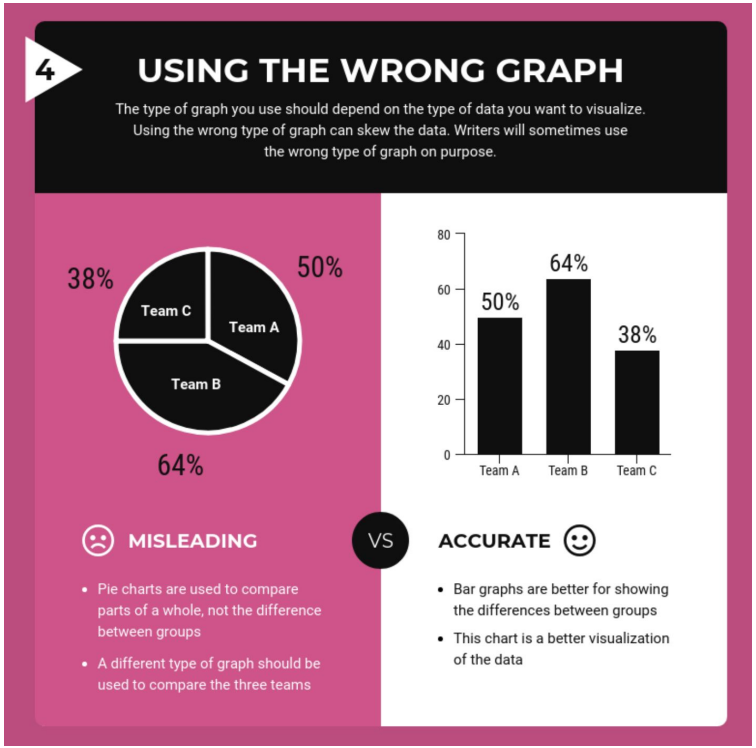
More Limitations

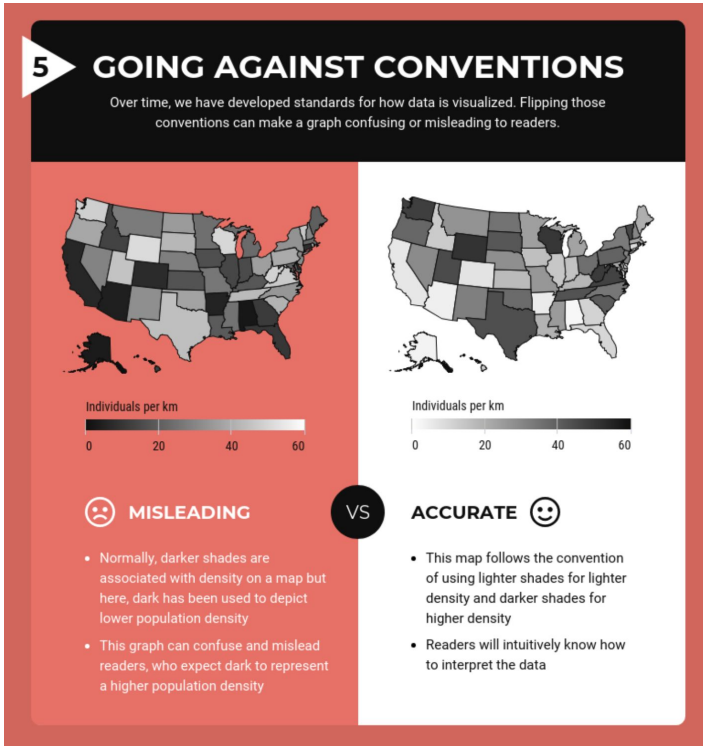




Feel free to ask questions at any point during the presentation!

More Limitations





Feel free to ask questions at any point during the presentation!

Discussion

- What **commonalities** do you notice among the more misleading and more accurate versions of graphs and charts in these examples?
- How would you define "**accuracy**" in the context of data presentation? Why is that question essential to ask?
- In what **contexts** does it make the most sense to use these kinds of visuals to present data? Are there other times where they're inappropriate? How so?

Feel free to ask questions at any point during the presentation!

Final Thoughts on Data Interpretation

- Always keep in mind the original research purpose when analyzing secondary data, including data from surveys.
- Do the variables used properly reflect the concepts you are trying to explore?
- When looking at graphs and other visual representations, pay attention to what they are trying to show: *is the baseline omitted? Is the scale deceiving? What are the units of measure? What timeline is captured?*

Feel free to ask questions at any point during the presentation!

Thank you!

If you have any questions, contact us at: nulab.info@gmail.com
Sign up for office hours at: http://calendly.com/diti-nu/

We'd love your feedback! Please fill out a short survey here: https://bit.ly/diti-feedback

Developed by DITI Research Fellows: Margarida Rodrigues and Yunus Emre Tapan

Slides & handouts available at: https://bit.ly/fa22-poe-surveydesign

Feel free to ask questions at any point during the presentation!