Session 1:
Don’t Become a Data Horror Story:
Best Practices, Tips, and Tricks
Jessica Schad: Assistant Professor, Sociology

Tyson Barrett: Research Assistant Professor, Managing Director of the Data Science & Discovery Unit, EEJCEHS
BEST PRACTICES ON HOW TO LAUNCH EFFECTIVE SURVEYS

DR. JESSICA D. ULRICH-SCHAD
ASSISTANT PROFESSOR OF SOCIOLOGY
WHAT ARE SURVEYS?

- **Survey**: collects information from a sample of individuals through their responses to *standardized* questions on a questionnaire
  - Mostly **closed-ended** questions, some **open-ended**

- **3 main advantages:**
  - 1) Versatile
  - 2) Efficient
  - 3) Generalizable

| 24. To what degree do you attribute the following problems to the recent oil and gas industry development in your area? |
|---|---|---|---|---|
| Violent crime (e.g., homicide, aggravated assault). | Not a problem in my area. | A problem here, but not caused by oil and gas development. | A problem here caused partially by oil and gas development. | A problem here caused solely by oil and gas development. |
| Property crime (e.g., burglary; theft, vandalism). | | | | |
| Manufacturing or sales of illegal drugs. | | | | |
| Public drinking or public drug use. | | | | |
| Prostitution. | | | | |
| Human/sex trafficking. | | | | |
| Homelessness. | | | | |
| Panhandling or begging. | | | | |
| Vandalism or graffiti. | | | | |
| Gangs. | | | | |
KEY ISSUE IN SURVEY RESEARCH TODAY

• Response rates down
  • Stedman et al. (2019) found a decline on average of .76%/year

• Key problem: can hurt representativeness

• Why are response rates declining?
COLLECTING SURVEY DATA TODAY

• Some tips on gathering valid, reliable, and generalizable data
  • Planning
  • Collecting
  • Presenting
1) PLAN HOW YOU WILL USE YOUR SURVEY DATA BEFORE YOU COLLECT IT.

- “I wish I would have added a question about........” OR “I really wish I would have asked the question like this instead of.......”

- Ask yourself:
  - What kind of analysis do I plan to do?
  - What are my I.V.? D.V.?
  - What level of measurement for each variable do I need?

20. Since the most recent oil boom, do you feel violent crime (e.g., homicide, aggravated assault) in your community has increased or decreased?
   - Increased
   - Stayed the same
   - Decreased

21. Since the most recent oil boom, do you feel property crime (e.g., burglary, theft, vandalism) in your community has increased or decreased?
   - Increased
   - Stayed the same
   - Decreased

22. Please rate how much of a problem, if at all, you feel the level of violent crime (e.g., homicide, aggravated assault) is in your community.
   - Very much a problem
   - Somewhat of a problem
   - A slight problem
   - Not a problem at all

23. Please rate how much of a problem, if at all, you feel the level of property crime (e.g., burglary, theft, vandalism) is in your community.
   - Very much a problem
   - Somewhat of a problem
   - A slight problem
   - Not a problem at all

30. What year were you born?
2) TO CREATE A GOOD SURVEY, PUT YOURSELF IN THE SHOES OF POTENTIAL PARTICIPANTS.

- Questions need to “make sense” and be answerable
- Survey can’t be too long
- Timing is important
- Topic saliency=more responses
- Pilot study, expert review
WRITING GOOD SURVEY QUESTIONS

• Use short/non-academic words and sentences

• Do you believe in anthropogenic climate change?

• Do you believe in human-caused climate change?
WRITING GOOD SURVEY QUESTIONS

• Avoid **double-barreled** questions

• *In your opinion, how would you rate the speed and accuracy of your work?*
  • Excellent
  • Good
  • Fair
  • Poor

• *In your opinion, how would you rate the accuracy of your work?*
WRITING GOOD SURVEY QUESTIONS

• **Minimize bias/leading**

• More people have attended the movie, Gone with the Wind, than any other motion picture produced this century. Have you seen this movie?
  • Yes
  • No
  • Not sure

• Have you seen the movie Gone with the Wind?
WRITING GOOD SURVEY QUESTIONS

• **Allow for disagreement**

• *Do you agree that this community is a good place to live?*
  • *Strongly Agree*
  • *Agree*
  • *Neither Agree nor Disagree*
  • *Disagree*
  • *Strongly Disagree*

• *Do you agree or disagree that this community a good place to live?*
WRITING GOOD SURVEY QUESTIONS

• Don’t ask questions they can’t answer

• How much money did you spend on groceries in the last year?

• Approximately how much money did you spend on groceries in the past week?
WRITING GOOD SURVEY QUESTIONS

• Make response categories **exhaustive**

• What type of family did you primarily grow up in?
  • Mother only
  • Father only
  • Both mother and father

• What type of family did you primarily grow up in?
  • Mother only
  • Father only
  • Both mother and father
  • Other: _____________________
WRITING GOOD SURVEY QUESTIONS

• Make response categories **mutually exclusive** (except for “check all that apply” questions)
  • How many hours do you typically spend on social media in a week?
    • None
    • 1-3
    • 3-10
    • 10+
  • How many hours do you typically spend on social media in a week?
    • None
    • 1-3
    • 4-10
    • 11+
INCENTIVES

- Social exchange theory: pre-incentives work better than promised incentives

- “Thanks for the $2 bill, I will give it to my granddaughter for her piggy bank.” – 2018 SD Sustainability Survey
3) KNOW THE LIMITATIONS OF YOUR SURVEY DESIGN.

• ALL research has weaknesses – including surveys
  • Flaws in design and/or lack of resources

• Awareness = address or temper findings
  • Non-probability sampling or coverage error (e.g., incomplete sampling frame) = might not be generalizable
  • Non-response error = weighting
EMERGING METHODS IN SURVEY RESEARCH

• Online panels

• Multiple modes of contact and/or response
KEY TAKEAWAYS REGARDING SURVEY RESEARCH

• 1) Plan how you will use your survey data before you collect it.

• 2) To create a good survey, put yourself in the shoes of potential participants.

• 3) Know the limitations of your survey design.
INTERESTED IN LEARNING MORE ABOUT SURVEY DESIGN?

• SOCIOLOGY 7100
  • Advanced Survey Research Techniques
  • Spring 2020
  • Day/time TBD

• Email me if interested and I can send you more information
  • Jessica.schad@usu.edu
Finding Data Balance

Tyson S. Barrett, PhD
Microsoft Excel or Google Sheets

SPSS, Jamovi, or JASP

R or SAS
Data in Spreadsheets

A million ways to make your life harder

<table>
<thead>
<tr>
<th>File</th>
<th>Data Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataFINAL_v5_comments.xlsx</td>
<td>02/09/2019</td>
</tr>
<tr>
<td>dataFINAL_FINAL_v2.xlsx</td>
<td>02/28/2019</td>
</tr>
<tr>
<td>dataFINAL_FINAL_FINAL.xlsx</td>
<td>02/20/2019</td>
</tr>
<tr>
<td>dataForAnalysis.xlsx</td>
<td>12/01/2018</td>
</tr>
<tr>
<td>data_Jim_FINAL.xlsx</td>
<td>12/15/2018</td>
</tr>
</tbody>
</table>
Data in Spreadsheets

A million ways to make your life harder

<table>
<thead>
<tr>
<th>File</th>
<th>Data Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataFINAL_FINAL_FINAL.xlsx</td>
<td>02/09/2019</td>
</tr>
<tr>
<td>dataFINAL_FINAL_v2.xlsx</td>
<td>02/28/2019</td>
</tr>
<tr>
<td>dataFINAL_FINAL.xlsx</td>
<td>02/20/2019</td>
</tr>
<tr>
<td>dataForAnalysis.xlsx</td>
<td>12/01/2018</td>
</tr>
<tr>
<td>data_Jim_FINAL.xlsx</td>
<td>12/15/2018</td>
</tr>
</tbody>
</table>

Copy-and-Pasting data

Color coding without indicators

Dates and things that look like dates

Hiding columns
# Data in Spreadsheets

A million ways to make your life harder

<table>
<thead>
<tr>
<th>File</th>
<th>Data Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataFINAL_v5_comments.xlsx</td>
<td>02/09/2019</td>
</tr>
<tr>
<td>dataFINAL_FINAL_v2.xlsx</td>
<td>02/28/2019</td>
</tr>
<tr>
<td>dataFINAL_FINAL_FINAL.xlsx</td>
<td>02/20/2019</td>
</tr>
<tr>
<td>dataForAnalysis.xlsx</td>
<td>12/01/2018</td>
</tr>
<tr>
<td>data_Jim_FINAL.xlsx</td>
<td>12/15/2018</td>
</tr>
</tbody>
</table>

**Copy-and-Pasting data**

**Color coding without indicators**

**Data in Spreadsheets**

Datapalooza 2020
Data in Spreadsheets

Some Basic Principles

- Be consistent
- Write dates like YYYY-MM-DD
- Don’t leave any cells empty
- Put just one thing in a cell
- Organize data as a single rectangle

Create a data dictionary

- Don’t use color as data
- Don’t include calculations in raw data

Broman et al., 2018
Data in Spreadsheets

Some Basic Principles

- Don't leave any cells empty
- Be consistent
- Put just one thing in a cell
- Write dates like YYYY-MM-DD

Organize data as a single rectangle

Create a data dictionary

Don't use color as data

Don't include calculations in raw data

These are ways to be kind to your future self and obtain data balance!
Avoid Headaches!

Have a consistent folder structure for where Spreadsheet analyses are vs Spreadsheet data are

- **Project Name**
  - **Data**
    - raw_data.csv
    - cleaned_data.csv
  - **Analyses**
    - descriptives.xlsx
    - regressions.Rmd
  - **Manuscripts**
    - to_submit.docx
    - writeup.Rmd
## A Grammar of Data

Wickham, SQL, dplyr

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>data wrangling</td>
<td>obtaining and cleaning data to get ready for analyses</td>
<td>data cleaning, data manipulating</td>
</tr>
<tr>
<td>selecting</td>
<td>grabbing columns (&quot;variables&quot;) and leaving others out</td>
<td></td>
</tr>
<tr>
<td>filtering</td>
<td>grabbing rows (&quot;observations&quot;) and leaving others out</td>
<td></td>
</tr>
<tr>
<td>mutating</td>
<td>creating a new variable and adjusting an existing one</td>
<td>transform, recoding</td>
</tr>
<tr>
<td>summarizing</td>
<td>reducing the amount of data using summary information (e.g. means, medians, etc.)</td>
<td>aggregating</td>
</tr>
<tr>
<td>pivoting</td>
<td>Reformatting the data to make it longer or wider (not changing any values in the data)</td>
<td>reshaping, elongating, widening</td>
</tr>
<tr>
<td>tidying</td>
<td>Creating a data set that has: 1. All columns are variables 2. All rows are observations</td>
<td>rectangulating, cleaning, pivoting</td>
</tr>
</tbody>
</table>
## A Grammar of Data

**Wickham, SQL, dplyr**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>data wrangling</td>
<td>obtaining and cleaning data to get ready for analyses</td>
<td>data cleaning, data manipulating</td>
</tr>
<tr>
<td>selecting</td>
<td>grabbing columns (&quot;variables&quot;) and leaving others out</td>
<td></td>
</tr>
<tr>
<td>filtering</td>
<td>grabbing rows (&quot;observations&quot;) and leaving others out</td>
<td></td>
</tr>
<tr>
<td>mutating</td>
<td>creating a new variable and adjusting an existing one</td>
<td>transform, recoding</td>
</tr>
<tr>
<td>summarizing</td>
<td>reducing the amount of data using summary information (e.g. means, medians, etc.)</td>
<td>aggregating</td>
</tr>
<tr>
<td>pivoting</td>
<td>Reformatting the data to make it longer or wider (not changing any values in the data)</td>
<td>reshaping, elongating, widening</td>
</tr>
</tbody>
</table>
| tidying    | Creating a data set that has:  
1. All columns are variables  
2. All rows are observations | rectangulating, cleaning, pivoting            |