Evaluating Data

Margarita Corral
Data Analysis Specialist, Brandeis University
she/her/hers mcorral@brandeis.edu

Aimee Slater
Academic Outreach Librarian for Government Information and the Social Sciences, Brandeis University
she/her/hers aslater@brandeis.edu
Today’s Agenda

Data Overview: 10 mins
Activity: Evaluate a Dataset in Small Groups: 20 mins
Brief report from each group: 15 mins
Data Literacy

“The ability to collect, manage, evaluate, and apply data, in a critical manner”

(Ridsdale et al. 2015)

Data Information Literacy. 12 competencies (Carlson et al. 2013)

1. Data Processing and Analysis
2. Data Management and Organization
3. Data Preservation
4. Databases and Data Formats
5. Ethics and Attribution
6. **Data Quality** and Documentation
Data Options

Research Question

Do I Collect my Own Data?

Do I Look for Data?

Experiment?
Survey?
Tools?
DMP

Locating Data
Who collected the data?
How was the data collected?
When was the data collected?
Population/Sample?
Metadata/Codebook/Data User Dictionary
Format?
Clean?

IRB?
Where to Find Data

Governments/NGOs
- USGS
- UNdata
- DATA.GOV
- WORLD BANK

Research Orgs
- ICPSR
- SCIENTIFIC DATA
- Socrata
- OSDC

Parking Lots
- GitHub
- Amazon Web Services
- Data.world
Data Quality Checklist

• Origin
  – Who collected the data?
  – Why did they start collecting the data?
  – Who paid to collect the data?
  – Date of creation
  – Is it being updated/maintained? Are there different versions?
  – Quality Control? Audits?
  – Is there supporting documentation? Does there need to be supporting documentation?
Data Quality Checklist

• Format
  – Portable?
  – Do you need software to open/analyze it?
  – Which version?
Data Quality Checklist

• Fields/Values
  – Codebook or Data Dictionary
  – Metadata. Is it easy to find?
  – Is it “clean?”
  – Are there missing values?

Question: “Does the Earth go around the Sun, or does the Sun go around the Earth?”

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>Earth goes around Sun</td>
</tr>
<tr>
<td>2.00</td>
<td>Sun goes around Earth</td>
</tr>
<tr>
<td>7.00</td>
<td>Won’t say</td>
</tr>
<tr>
<td>8.00</td>
<td>Don’t know</td>
</tr>
<tr>
<td>9.00</td>
<td>Not Asked</td>
</tr>
</tbody>
</table>

National Science Foundation Surveys of Public Attitudes Toward and Understanding of Science and Technology, 1979-2006: [United States]
Tools for Cleaning Data
Tools for Data Analysis

- STATA
- R
- SPSS
- Excel
- ArcGIS
Some Resources


Brandeis Library
Let’s Evaluate a Few Datasets!

1. Break into small groups. We will give you a notecard with your group number.
2. Head to https://tinyurl.com/slbcData and open the file with your Group Number named.
3. Check out the dataset and supporting documentation. If available, we’ve included a URL to the source, so check the link for additional information.
4. Fill out the sheet.