Data Sharing with ICPSR: Fueling the Cycle of Science through Discovery, Access Tools, and Long-Term Availability

Johanna Bleckman and Kaye Marz, ICPSR

IASSIST 2015
Minneapolis, MN
June 2, 2015
Learning objectives

You will become more familiar with:

• Federal data sharing requirements and other good reasons to share data
• Options for sharing data
• Protection of confidentiality when sharing data
• Data discovery tools
• Online data exploration tools from ICPSR
Objective I:
Federal data sharing requirements and other good reasons to share data
Why share data?

- Federal data sharing requirements allow effective use of federal agency resources
  - Office of Science and Technology Policy memorandum (2013)
  - National Science Foundation (2008)
    - Requirement for Data Management Plans in 2010
    - Requirement for Data Sharing Plans included
  - National Institute of Justice (1978)
Why share data?

• Benefits to the social science community (NIH, 2003)
  – Reinforces open scientific inquiry
  – Encourages diversity of analysis and opinion
  – Promotes new research, testing of new hypotheses and methods of analysis
  – Supports studies on data collection methods and measurement
  – Facilitates education of new researchers
  – Enables exploration of topics not envisioned by initial investigators
  – Permits creation of datasets combined from multiple sources
Why share data?

• Benefits to the researcher
  – Data in the public domain generate new research which cites the original research
  – Data are preserved and can be obtained by the original researchers if their copies are lost or destroyed
  – Archiving data helps researchers meet requirements of NIH and NSF data management plans and NIJ data archiving special conditions requirements
  – Information on research community’s interest in a dataset can assist with the success of future grant funding proposals
Where are the data now?

- Lost in a technical malfunction
- Destroyed in a flood in the department
- Files were on the university server but are long gone
- Data are kept for 10 years beyond the last time used
- Purged after retirement
- Institutional review boards required the data be destroyed after the project

Sharing data...more publications!

Table. Median Number of Publications by Data Sharing Status

<table>
<thead>
<tr>
<th></th>
<th>Data Archived (n=111)</th>
<th>Data Shared Informal (n=415)</th>
<th>Data Not Shared (n=409)</th>
<th>Total (n=935)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary PI Publications</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Secondary Publications</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Publications with Students</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Cycle of Science

New data collected

Data deposited in archive

Discovered and accessed from archive

New datasets or research ideas

Analyzed and results published
Data deposited in archive

Discovered and accessed from archive

Analyzied and results published

New datasets or research ideas

New data collected

Cycle of Science
Cycle of Science

Data deposited in archive

Discovered and accessed from archive

Analyzed and results published

New datasets or research ideas

New data collected
Data deposited in archive

Discovered and accessed from archive

New data collected

New datasets or research ideas

Cycle of Science

Analyzed and results published

Published
Objective II: Options for sharing data
Options for sharing data through ICPSR
How do ICPSR and openICPSR compare to other data service providers?

- Have professional data curators to review deposited materials who are experts in developing metadata (tags) for the social and behavioral sciences
- Provide an immediate distribution network of over 760 institutions looking for research data, that has powerful search tools, and a data catalog indexed by major search engines
- Sustained by a respected organization with over 50 years of experience in reliably protecting research data
- Prepared to accept and disseminate sensitive and/or restricted-use data

Discoverable

Accessible

Sustainable

Trusted
How is openICPSR different from ICPSR?

**ICPSR**

- Sustained by institutional member fees; depositors are not charged to deposit data
- Data are freely available to individuals affiliated with member institutions; non-members pay an access fee; both may pay fee to access restricted-use data in VDE
- Data are fully curated including professional processing, value-added documentation, and renderings in popular statistical programs and online analysis

**openICPSR**

- Sustained by deposit fee to individuals from non-members institutions ($600/project)
- Data freely available to the public (or for a nominal charge for restricted-use data)
- Most data available only in the raw (bit-level) form as deposited and described by the depositor; may be fully curated if the Professional Curation option was chosen and the quoted fee paid for ICPSR’s curation and/or disclosure review services
Who might use openICPSR?

- Researchers required to share data freely with the public to comply with grant/contract requirements
- Authors required to share data for replication purposes to comply with journal requirements
- Researchers required to share sensitive data or data with disclosure risk with the public from a secure digital environment
- Researchers, including students, who want to share data publicly as good practice or for the purposes of replication
Deposit agreements

- Has implicit or explicit copyright and have the right to make it publicly available through ICPSR
- Permits ICPSR to redisseminate, promote, catalog, reformat, store, and preserve the data collection
- Permits ICPSR to transform or enhance the collection to protect confidentiality and for usability
- Has removed all direct identifiers and done due diligence to prevent disclosure of subject identities
- Holds ICPSR and University of Michigan harmless from liability for breaches of subject confidentiality or invasions of privacy
openICPSR deposit terms includes more

- Deposited work will be distributed under an Attribution
  4.0 Creative Commons License
- Depositor has institutional approval to share the data
  collection
- Data collection will be preserved as-is and available at no
  cost to data users
- Research subjects have consented to sharing the data
  and/or the depositor has institutional approval to share
  the data
openICPSR: Public-access sharing solution for Institutions and Journals

• Branded repository to represent and showcase their research data using a fully-hosted (cloud) service
• Able to fulfill grant requirements that will pass an audit
• No need to have technical staff or equipment to manage the repository
• Easy and clean interface for deposit and access
• Administrative (usage) reporting available 24/7
• Able to assure the Research Administration Office that data are safe and secure for the long term
openICPSR: Examples of branded sites and services

Your logo. Your colors.
A unique URL.

On-demand deposit.
On-demand reports.
Should the data be restricted?

- **Re-identification**
  - Can individuals be identified from information in this material?
  - If the data were made public, could someone use a combination of variables (e.g. age, sex, race, occupation, geography) to find individuals in a publicly available database?

- **Harm**
  - Does this material include sensitive information?
  - Would the release of individually identifiable information create a risk of harm (e.g. psychological distress, social embarrassment, financial loss) greater than the risks that people experience in everyday life?
Identifiers

- **Direct**
  - Names
  - Addresses, including ZIP codes and other postal codes
  - Telephone and fax numbers, including area codes
  - Social security numbers
  - Other linkable numbers such as driver’s license numbers, certification numbers, medical device numbers, etc.

- **Indirect**
  - Detailed geographic information
  - Organizations belonged to, offices or posts held by respondent
  - Educational institutions attended, year of graduation
  - Detailed occupational titles
  - Place where respondent was born or grew up
  - Exact dates of significant life events (birth, death, marriage, etc)
  - Detailed income

*Direct identifiers must be removed from data before sharing! Data with indirect identifiers may need to be restricted.*
What sensitive topics could cause harm?

- Psychological well being or mental health
- Sexual attitudes, preferences, or practices
- Use of alcohol or drugs
- Illegal behavior
- Behavior that puts the respondent at risk of criminal or civil liability
- Behavior damaging to an individual’s financial standing, employability, or reputation
- Medical information that could lead to discrimination, stigmatization, etc.
Let’s make a deposit with openICPSR

Share your social and behavioral science research data

Get started now »

Maximize Access
Be recognized and cited

Store Safely
Store your data with confidence

Protect Confidentiality
Ensure confidentiality and privacy

www.openicpsr.org
Objective III: Protection of confidentiality when sharing data
Common Objection/Misperception: “My data are too sensitive to share. . .”

- Many restricted data files can be processed for public release
- ICPSR has been sharing restricted-use data for over a decade via three methods:
  - Secure Download
  - Virtual Data Enclave
  - Physical Enclave
- ICPSR stores & shares over 6,400 restricted-use datasets associated with over 2,000 ‘active’ restricted-use data contracts
Virtual Data Enclave

- Virtual machine launched from your desktop machine
- Functions just like your local machine, but with restrictions on what can enter and exit the environment
- Full suite of statistical packages and other software
Virtual Data enclave (VDE) – user experience
The Visual

University of Michigan Information Technology Services
Virtual Desktop Infrastructure (VDI)

- Virtual Data Enclave
  - Secure Data Analysis
- Secure Data Environment
  - Secure Ingest

- VM Client
  - Virtual Machine client software
- virtual firewall

- researcher
- ICPSR staff member

- ICPSR

- EMC²
  - Secure Storage
  - NS-120
Examples of Disclosure Risk Concerns

- Tabular output including demographics or unique characteristics/activities, resulting in small cell sizes
- Geographic information (zip codes, cities/towns, counties)

<table>
<thead>
<tr>
<th>Gender, Date of birth</th>
<th>and Type of location information</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-digit zipcode</td>
<td>city, town, municipality</td>
</tr>
<tr>
<td>87%</td>
<td>county</td>
</tr>
<tr>
<td>53%</td>
<td>18%</td>
</tr>
</tbody>
</table>

- Geocodes, GIS data, maps
- Longitudinal data – data from research subjects at multiple time points
- Verbatim text – interviews, video transcriptions, short answer responses
- Photos, videos, audio recordings
Common Rules for Tables

Each cell size is sufficiently large to prevent identifiability

- Establish minimum threshold (often 3, 5, or 10), dependent on type and context of the data

- Rules for cells, rows, and columns:
  - All cases in any row or column should not be in a single cell
  - Cell percentage should not correspond to a cell size less than a threshold number
  - A cell should not be a high percent of all cases included in a row or column (more than 60%)

- Combinations of tables should also meet guidelines
Disclosure Risk Protections (DRP)

- Release data from only a sample of the population
- Remove/mask obvious identifiers and high-risk variables
- Limit geographic detail
- Limit the number and detailed breakdown of categories
- Top or bottom coding
- Recode into intervals or round values
- Addition of noise
- Swap records
- Blank and impute
- Aggregate and replace with mean value (aka blurring)
## Disclosure Risk Remediation, Example 1

<table>
<thead>
<tr>
<th>Years of experience</th>
<th>0 to 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>110</td>
<td>166</td>
<td>183</td>
<td>143</td>
<td>133</td>
<td>450</td>
<td>280</td>
<td>174</td>
<td>117</td>
<td>97</td>
<td>58</td>
<td>19</td>
<td>1930</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>24</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>108</td>
</tr>
<tr>
<td>Hispanic</td>
<td>35</td>
<td>55</td>
<td>65</td>
<td>62</td>
<td>51</td>
<td>128</td>
<td>78</td>
<td>50</td>
<td>24</td>
<td>21</td>
<td>12</td>
<td>2</td>
<td>583</td>
</tr>
<tr>
<td>White, other</td>
<td>60</td>
<td>91</td>
<td>95</td>
<td>80</td>
<td>81</td>
<td>298</td>
<td>188</td>
<td>112</td>
<td>84</td>
<td>72</td>
<td>58</td>
<td>20</td>
<td>1239</td>
</tr>
<tr>
<td>Gener</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>22</td>
<td>35</td>
<td>37</td>
<td>25</td>
<td>20</td>
<td>90</td>
<td>56</td>
<td>38</td>
<td>24</td>
<td>18</td>
<td>10</td>
<td>4</td>
<td>379</td>
</tr>
<tr>
<td>Female</td>
<td>88</td>
<td>131</td>
<td>141</td>
<td>123</td>
<td>113</td>
<td>360</td>
<td>224</td>
<td>130</td>
<td>99</td>
<td>78</td>
<td>49</td>
<td>15</td>
<td>1551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years of experience</th>
<th>0 to 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>110</td>
<td>166</td>
<td>183</td>
<td>143</td>
<td>133</td>
<td>450</td>
<td>280</td>
<td>174</td>
<td>117</td>
<td>97</td>
<td>77</td>
<td>2007</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or Hispanic</td>
<td>40</td>
<td>63</td>
<td>73</td>
<td>73</td>
<td>60</td>
<td>152</td>
<td>92</td>
<td>62</td>
<td>33</td>
<td>25</td>
<td>18</td>
<td>691</td>
</tr>
<tr>
<td>Hispanic</td>
<td>40</td>
<td>63</td>
<td>73</td>
<td>73</td>
<td>60</td>
<td>152</td>
<td>92</td>
<td>62</td>
<td>33</td>
<td>25</td>
<td>18</td>
<td>691</td>
</tr>
<tr>
<td>White, other</td>
<td>60</td>
<td>91</td>
<td>95</td>
<td>80</td>
<td>81</td>
<td>298</td>
<td>188</td>
<td>112</td>
<td>84</td>
<td>72</td>
<td>78</td>
<td>1317</td>
</tr>
<tr>
<td>Gener</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>22</td>
<td>35</td>
<td>37</td>
<td>25</td>
<td>20</td>
<td>90</td>
<td>56</td>
<td>38</td>
<td>24</td>
<td>18</td>
<td>14</td>
<td>393</td>
</tr>
<tr>
<td>Female</td>
<td>88</td>
<td>131</td>
<td>141</td>
<td>123</td>
<td>113</td>
<td>360</td>
<td>224</td>
<td>130</td>
<td>99</td>
<td>78</td>
<td>64</td>
<td>1615</td>
</tr>
</tbody>
</table>
### Disclosure Risk Remediation, Example 2

<table>
<thead>
<tr>
<th></th>
<th>Before Mean</th>
<th>After Mean</th>
<th>Change</th>
<th>Before Std. Dev.</th>
<th>After Std. Dev.</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>10 Uniques Removed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REST_ELA_2009</td>
<td>0.0023</td>
<td>0.0033</td>
<td>-0.001</td>
<td>0.19137</td>
<td>0.19239</td>
<td>-0.00102</td>
</tr>
<tr>
<td>REST_MATH_2009</td>
<td>-0.0046</td>
<td>-0.005</td>
<td>0.0004</td>
<td>0.23729</td>
<td>0.23737</td>
<td>-0.00080</td>
</tr>
<tr>
<td><strong>166 Uniques Removed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REST_ELA_2009</td>
<td>-0.0054</td>
<td>-0.0055</td>
<td>0.00010</td>
<td>0.20921</td>
<td>0.20924</td>
<td>-0.00030</td>
</tr>
<tr>
<td>REST_MATH_2009</td>
<td>0.02493</td>
<td>0.02507</td>
<td>-0.00014</td>
<td>0.233647</td>
<td>0.233327</td>
<td>0.000320</td>
</tr>
<tr>
<td><strong>12 Uniques Removed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REST_ELA_2009</td>
<td>-0.0086</td>
<td>-0.00089</td>
<td>-0.00771</td>
<td>0.187485</td>
<td>0.187746</td>
<td>-0.000261</td>
</tr>
<tr>
<td>REST_MATH_2009</td>
<td>0.0054</td>
<td>0.0054</td>
<td>0</td>
<td>0.23612</td>
<td>0.23611</td>
<td>0.000010</td>
</tr>
<tr>
<td><strong>633 Uniques Removed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REST_ELA_2009</td>
<td>-0.0041</td>
<td>-0.0039</td>
<td>-0.0002</td>
<td>0.17251</td>
<td>0.173</td>
<td>-0.00049</td>
</tr>
<tr>
<td>REST_MATH_2009</td>
<td>-0.0063</td>
<td>-0.0064</td>
<td>0.0001</td>
<td>0.21793</td>
<td>0.21828</td>
<td>-0.000350</td>
</tr>
</tbody>
</table>
Documenting Disclosure Mitigation

• Maintain and deposit the syntax used to modify the data
• Include either:
  – a narrative description of changes in the study-level documentation, or
  – a series of item-specific narratives to be included at the variable level (codebook/user guide)
Objective IV:
Data discovery tools
What is ICPSR?

- Then and Now -

• One of the world’s oldest and largest social science data archives, est. 1962

• Data distributed on punch cards, then reel-to-reel tape, now:
  – Data available on demand
  – Over 8,200 studies with over 68,700 data sets

• Membership organization among 21 universities, now:
  – Currently over 760 members world-wide
  – Federal funding of public collections
The Concept of “Data Curation”

• Curation, from the Latin "to care," is the process used to add value to data, maximize access, and ensure long-term preservation

• Data curation is akin to work performed by an art or museum curator.
  – Data are organized, described, cleaned, enhanced, and preserved for public use, much like the work done on paintings or rare books to make the works accessible to the public now and in the future

• Curation provides meaningful and enduring access to data

• Data curation is the foundation for effective, long-term data sharing
Assessing the data in the collection

• Searching for and downloading data
• Codebooks
• Full descriptives
• Variable search and compare
• Simple crosstabs and frequencies
• Online analysis functionality
Study Search Behaviors

In practice, we encounter three typical search behaviors from our users:

• A user has a research question in mind.
• A user is looking for a dataset that contains specific variables.
• A user is looking for a specific dataset and has the study title or investigator name.
Natural Language Searching

• Does juvenile drug use lead to delinquency?
• juvenile “drug use” delinquency
• “juvenile drug use”
Searching by Variable/Concept

1. Juvenile Detention and Correctional Facility Census, 1984-1985 (ICPSR 8495)
United States Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention; United States Department of Justice, Office of Justice Programs. Bureau of Justice Statistics

5 variables match delinquency including:

- V216: TOTAL NUMBER OF DELINQUENTS
- V217: CUST AUTH: 1= ACCESSION DELINQUENT
- V55: JUV POP-COMTD DELINQUENT: MALE
- V390: CUST AUTH: 4= ADJUDICATED DELINQ.
- V311: CUST AUTH: CENSUS USE R01

No variable names/labels match drug use, but the concept might exist.

2. Research on Pathways to Desistance (Maricopa County, AZ and Philadelphia County, PA): Collateral Measures, 2000-2004 (ICPSR 32881)
Mulvey, Edward P.

30 variables match delinquency including:

- C1P08H: C1 P08: Antisocial Behavior; Peer Delinquency-Antisocial behavior; the mean rating of the prevalence of friends who engage in it
- C1P09F: C1 P09F: Antisocial Influence; Peer Delinquency-Antisocial influence; the mean rating of the prevalence of friends who encourage
- C1P09G: C1 P09G: Antisocial Behavior; Peer Delinquency-Antisocial behavior; the mean rating of the prevalence of friends who engage in it
- C1P09H: C1 P09H: Antisocial Influence; Peer Delinquency-Antisocial influence; the mean rating of the prevalence of friends who encourage
- C1P09I: C1 P09I: Antisocial Behavior; Peer Delinquency-Antisocial behavior; the mean rating of the prevalence of friends who engage in it
- C1P09J: C1 P09J: Antisocial Influence; Peer Delinquency-Antisocial influence; the mean rating of the prevalence of friends who encourage

281 variables match drug use including:

- C1YBUSE: C1 Subuse: Number of drugs used in the past year; Number of drugs used in the past year; count of endorsed items. MEASURE: Substance Abuse The Subs
- C1YBUSE: C2 Subuse: Number of drugs used in the past year; Number of drugs used in the past year; count of endorsed items. MEASURE: Substance Abuse The Subs
- C1YBUSE: C3 Subuse: Number of drugs used in the past year; Number of drugs used in the past year; count of endorsed items. MEASURE: Substance Abuse The Subs
- C1SUBUSE DRS: C1 Subuse: Did the subject's drug use ever cause him/her to be treated for drug abuse? Did the subjects drinking ever cause him/her to be treated for drug abuse (hospitalization, couns
- C1SUBUSE DRS: C2 Subuse: Did the subject's drug use ever cause him/her to be treated for drug abuse? Did the subjects drinking ever cause him/her to be treated for drug abuse (hospitalization, couns
- C1SUBUSE DRS: C3 Subuse: Did the subject's drug use ever cause him/her to be treated for drug abuse? Did the subjects drinking ever cause him/her to be treated for drug abuse (hospitalization, couns
- C1SUBUSE DRS: C1 Subuse: Has the subject ever had problems because of his/her drug use? Has the subject ever had problems because of his/her drug use (trouble at work/school, legal probl
- C1SUBUSE DRS: C2 Subuse: Has the subject ever had problems because of his/her drug use? Has the subject ever had problems because of his/her drug use (trouble at work/school, legal probl
Variable Search Tips, 1

• Enter words or strings that are likely to appear in a variable name, label, question, and value labels:

• *Presidential election* will return variables dealing with all presidential elections

• *Presidential election Obama* will return only variables dealing with the 2008 and 2012 presidential elections
Variable Search Tips, 2

- Use quotes to search for specific phrases:
  - "life satisfaction" "minority rights" "community programs"

- The minus sign may be used to remove certain types of results:
  - "Presidential debate" - "Bill Clinton" will eliminate the debates from 1992 and 1996 in which Clinton participated
Variable Search Tips, 3

- A Boolean "and" is implied in the search.
- The search automatically does stemming, there's no need to type in an asterisk. It's also case-insensitive.
- A fielded search is also available.
The Bibliography of Data-related Literature

*It’s really a searchable database* . . . containing over 65,000 citations of known published and unpublished works resulting from analyses of data archived at ICPSR 

*. . . that can generate study bibliographies* associating each study with the literature about it

*. . . Included in the integrated search* on the ICPSR Web site
Demonstrating the Impact of Research

Benefits of the Bibliography

The bibliography facilitates literature searches by social scientists, students, journalists, policymakers, and funding agencies. They use it to

- Identify much of the research that has already been undertaken in an area
- Replicate analyses of data
- Avoid duplicating analysis that has already been done
- Identify cross-disciplinary implications and uses of the data

The bibliography is also valuable for studying data as intellectual output. It allows investigators to

- Study how data resources are used
- Conduct citations analyses
- Investigate the life cycle of data
- Learn more about methodological issues, some of which are covered solely in the published literature
Find publications using the ICPSR BIBLIOGRAPHY
OF DATA-RELATED LITERATURE

The ICPSR Bibliography of Data-related Literature is a continuously-updated database of thousands of citations of works using data held in the ICPSR archive. The works include journal articles, books, book chapters, government and agency reports, working papers, dissertations, conference papers, meeting presentations, unpublished manuscripts, magazine and newspaper articles, and audiovisual materials.

- View all citations
- Browse by author
- Browse by journal
- Browse by study

Search Tips
- Please note that our citation search indexes only the literal citations, not the full text of the publications.
- Instead of entering a research question, use only one or two search words, and use the filters to narrow your results.
- Use quotes for phrase searches; use the minus sign to remove items from results:
  - "drug abuse" -adolescent
  - A Boolean “and” is invoked with multiple search terms.
  - Stemming is automatic; do not use an asterisk.
  - Search is not case-sensitive.

Submit Citations
ICPSR encourages its users to submit bibliographic citations to data we disseminate. If you would like to add a new citation to our Bibliography, please submit it via our citation form, or email us at bibliography@icpsr.umich.edu.

Benefits of the Bibliography
The Bibliography facilitates literature searches by social scientists, students, journalists, policymakers, and funding agencies. They use it to
- Identify much of the research that has already been undertaken in an area
- Replicate analyses of data
- Avoid duplicating analysis that has already been done
- Identify cross-disciplinary implications and uses of the data
Objective V: Online data exploration tools
### Simple Crosstab


**Rows:**
- Q144: Gender of Respondent

**Columns:**
- Q2260: COMB: Being a Christian - making someone a true A/K

**Control:**
- select control variable

**Selection Filter(s):**
- Example: age(18-50)

**Weight:**
- POP WT - Population Weight to US Population over Age 18: 209.129.096

**Percentaging:**
- Column

**N of cases to display:**
- Unweighted

**Summary statistics:**
- Show

**Run the Table**

### Help
- Help with SDA
  - General
    - Recoding Variables: Enter recode syntax into box to the right of row/column/control, without parentheses. For example: r:1-20,21-40,41-65
  - Getting Started
- Help with this Dataset
  - Browse Codebook

Variables
row · Q144: Gender of Respondent
column · Q2260: COMB: Being a Christian - making someone a true American

Frequency Distribution

<table>
<thead>
<tr>
<th>Cells contain:</th>
<th>Q2260 1: Very Important</th>
<th>Q2260 2: Somewhat Important</th>
<th>Q2260 3: Somewhat unimportant</th>
<th>Q2260 4: Very unimportant</th>
<th>Q2260 8: Don't Know</th>
<th>Q2260 9: No Answer</th>
<th>ROW TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of cases</td>
<td>14,262,882.0</td>
<td>11,851,922.5</td>
<td>13,816,481.4</td>
<td>47,365,492.8</td>
<td>638,836.6</td>
<td>919,403.2</td>
<td>88,855,018.4</td>
</tr>
<tr>
<td>Weighted N</td>
<td>36.1%</td>
<td>37.0%</td>
<td>42.7%</td>
<td>46.8%</td>
<td>36.4%</td>
<td>40.5%</td>
<td>42.5%</td>
</tr>
<tr>
<td>Q144 1: Male</td>
<td>197</td>
<td>166</td>
<td>191</td>
<td>631</td>
<td>10</td>
<td>12</td>
<td>1,207</td>
</tr>
<tr>
<td>Q144 2: Female</td>
<td>336</td>
<td>270</td>
<td>251</td>
<td>703</td>
<td>16</td>
<td>17</td>
<td>1,593</td>
</tr>
<tr>
<td>Q144 TOTAL</td>
<td>533</td>
<td>436</td>
<td>442</td>
<td>1,334</td>
<td>26</td>
<td>29</td>
<td>2,800</td>
</tr>
</tbody>
</table>

COL TOTAL       | 39,556,391.1             | 32,048,581.7                | 32,342,205.4                 | 101,157,915.2             | 1,757,841.6       | 2,271,959.1       | 205,128,094.0 |
Chart

Gender of Respondent BY COMB: Being a Christian – making someone a true American

Compare Variables

Variables

- List all 202 variables in this study
- Search the variables in this study

Instructional Resources

- ICPSR has created the following instructional guides that utilize data from this study:
  - American Identity and Immigrant Resentment

Additional materials can be found on our Resources for Instructors site.

Utilities
Variables

- List all 202 variables in this study
- Search the variables in this study

q11
GO
### Variable Search Results

Your query returned 3 variables. Use the search box below to revise your search, or start a new search.

**q11**

#### Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Label/Question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q11X</strong></td>
<td><strong>Recode of open-ended race question from Q11 and Q9</strong></td>
</tr>
<tr>
<td></td>
<td>Recode of open-ended race question from Q11 and Q9</td>
</tr>
<tr>
<td><strong>Q14</strong></td>
<td><strong>Does R describe them self as [Q11 and Q9, RACE]?</strong></td>
</tr>
<tr>
<td></td>
<td>Do you describe yourself as? Se describe usted a si mismo/a como?</td>
</tr>
<tr>
<td><strong>Q11</strong></td>
<td><strong>Race R self-identifies</strong></td>
</tr>
<tr>
<td></td>
<td>What race do you consider yourself to be... (white, black, asian, native american, or some other race)? De que raza usted se considera ser. (blanca, negra, Asiática, Nativa Americana o alguna otra raza)? ... [view details]</td>
</tr>
</tbody>
</table>
### Race R self-identifies

**Question:**
What race do you consider yourself to be... (white, black, Asian, native American, or some other race)?

**Responses**

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>Unweighted Frequencies</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>White/Caucasian</td>
<td>1,692</td>
<td>60.4</td>
</tr>
<tr>
<td>2</td>
<td>Black/African American</td>
<td>309</td>
<td>11.0</td>
</tr>
<tr>
<td>3</td>
<td>Asian</td>
<td>252</td>
<td>9.0</td>
</tr>
<tr>
<td>4</td>
<td>Native American</td>
<td>26</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>Some other race (Please Specify Single Race)</td>
<td>149</td>
<td>5.3</td>
</tr>
<tr>
<td>6</td>
<td>Hispanic/Latino</td>
<td>273</td>
<td>9.8</td>
</tr>
<tr>
<td>7</td>
<td>Mixed (Please Specify Mutli Race)</td>
<td>67</td>
<td>2.4</td>
</tr>
<tr>
<td>8</td>
<td>Don’t Know</td>
<td>14</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>No Answer</td>
<td>18</td>
<td>0.6</td>
</tr>
</tbody>
</table>

### Does R describe themselves as (Q11 and Q9, RACE)?

**Question:**
Do you describe yourself as ethnicity fill from question Q11? To describe usted a si mismo/a como ethnicity fill from question Q11?

**Responses**

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>Unweighted Frequencies</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>2,341</td>
<td>83.6</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>412</td>
<td>14.7</td>
</tr>
<tr>
<td>8</td>
<td>Don’t Know</td>
<td>14</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>No Answer</td>
<td>9</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Recode of open-ended race question from Q11 and Q9

**Question:**
Recodes of open-ended race question from Q11 and Q9

**Responses**

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>Unweighted Frequencies</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>White/Caucasian</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>Black/African American</td>
<td>15</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>Asian</td>
<td>44</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>Native American</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>Some other race (Please Specify Single Race)</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>Hispanic/Latino</td>
<td>5</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>Mixed (Please Specify Mutli Race)</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>Don’t Know</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>9</td>
<td>No Answer</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Crosstab Creator

Select Variables (help)

- ADMINISTRATIVE VARIABLES
- WEIGHT VARIABLES
- DEMOGRAPHIC VARIABLES
- AMERICANIZATION VARIABLES
- PUBLIC OPINION VARIABLES
- RACE VARIABLES
- ORGANIZATIONAL MEMBERSHIP VARIABLES
- POLITICAL AFFILIATION VARIABLES


Depending upon your setup, you may be able to drag variables to the boxes in the form, or simply type the variable name into the box.

Row: 

Column: 

Control: 

Selection Filter(s): 

Example: age(18-50)

Weight: POP_WT - Population Weight to US Population over Age 18: 209,128,094

Percentaging: 
- [ ] Column
- [ ] Row
- [x] Total

N of cases to display: 
- [ ] Unweighted
- [x] Weighted

Summary statistics: 
- [ ] Show

Run the Table

Depending upon your setup, you may be able to drag variables to the boxes in the form, or simply type the variable name into the box.

Row: Q11 (required)

Column: Q144

Control: 

Selection Filter(s): Example: age(18-50)

Weight: POP_WT - Population Weight to US Population over Age 18: 209,128,034

Percentaging: [ ] Column [ ] Row [ ] Total

N of cases to display: [ ] Unweighted [ ] Weighted

Summary statistics: [ ] Show
<table>
<thead>
<tr>
<th>Q11</th>
<th>Q144 1: Male</th>
<th>Q144 2: Female</th>
<th>ROW TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: White/Caucasian</td>
<td>68.2%</td>
<td>67.9%</td>
<td>68.0%</td>
</tr>
<tr>
<td></td>
<td>60,565,494.2</td>
<td>81,669,511.5</td>
<td>142,235,005.8</td>
</tr>
<tr>
<td>2: Black/African American</td>
<td>9.7%</td>
<td>12.7%</td>
<td>11.4%</td>
</tr>
<tr>
<td></td>
<td>8,587,884.6</td>
<td>15,322,475.6</td>
<td>23,910,360.2</td>
</tr>
<tr>
<td>3: Asian</td>
<td>4.9%</td>
<td>2.9%</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>4,342,282.1</td>
<td>3,516,827.5</td>
<td>7,859,109.7</td>
</tr>
<tr>
<td>4: Native American</td>
<td>1.0%</td>
<td>1.1%</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>856,093.1</td>
<td>1,302,673.4</td>
<td>2,158,766.5</td>
</tr>
<tr>
<td>5: Some other race (Please Specify Single Race)</td>
<td>5.7%</td>
<td>5.2%</td>
<td>5.4%</td>
</tr>
<tr>
<td></td>
<td>5,050,064.8</td>
<td>6,196,523.6</td>
<td>11,246,588.4</td>
</tr>
<tr>
<td>6: Hispanic/Latino</td>
<td>6.6%</td>
<td>7.0%</td>
<td>6.8%</td>
</tr>
<tr>
<td></td>
<td>5,832,015.0</td>
<td>8,383,521.5</td>
<td>14,215,536.5</td>
</tr>
<tr>
<td>7: Mixed (Please Specify Multi Race)</td>
<td>2.8%</td>
<td>2.1%</td>
<td>2.4%</td>
</tr>
<tr>
<td></td>
<td>2,452,662.1</td>
<td>2,578,868.9</td>
<td>5,031,531.0</td>
</tr>
<tr>
<td>8: Don’t Know</td>
<td>0.8%</td>
<td>0.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>721,942.1</td>
<td>308,899.5</td>
<td>1,030,841.6</td>
</tr>
<tr>
<td>9: No Answer</td>
<td>0.5%</td>
<td>0.8%</td>
<td>0.7%</td>
</tr>
<tr>
<td></td>
<td>446,580.4</td>
<td>993,774.0</td>
<td>1,440,354.3</td>
</tr>
<tr>
<td>COL TOTAL</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>88,855,018.4</td>
<td>120,273,075.6</td>
<td>209,128,094.0</td>
</tr>
</tbody>
</table>
Crosstab Assignment Builder

**Dataset(s)**

**DS0: Study-Level Files**

Documentation: [Questionnaire.pdf (English)](#) [Questionnaire.pdf (Spanish)](#) [User guide.pdf](#)

**DS1: Public-Use Data**

Documentation: [Codebook.pdf](#)

Analyze Online: [SDA](#) [Build a Crosstab Assignment](#)
<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8</td>
<td>Country R identifies with most (first, second, or third mention in Q7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Hispanic or Latin origin or descent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11</td>
<td>Race R self-identifies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11X</td>
<td>Recode of open-ended race question from Q11 and Q9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13</td>
<td>Does R describe them self as (Q8, ANCESTID)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14</td>
<td>Does R describe them self as (Q11 and Q9, RACE)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15</td>
<td>Does R describe them self as an American?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16</td>
<td>How R thinks of them self most of the time - country of origin, race, or as an American</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q17</td>
<td>Importance of being (Q8, ANCESTID) to R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## DEMOGRAPHIC VARIABLES

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q10</td>
<td>Primary language spoken in R's home</td>
<td></td>
</tr>
<tr>
<td>Q123</td>
<td>Respondent's age</td>
<td></td>
</tr>
<tr>
<td>Q124</td>
<td>Highest grade of school or year of college R has completed</td>
<td></td>
</tr>
<tr>
<td>Q129</td>
<td>R's religious preference</td>
<td></td>
</tr>
<tr>
<td>Q130M1</td>
<td>Current status of employment? (multiple responses allowed)</td>
<td></td>
</tr>
<tr>
<td>Q130M2</td>
<td>Current status of employment? (multiple response allowed)</td>
<td></td>
</tr>
<tr>
<td>Q130M3</td>
<td>Current status of employment? (multiple response allowed)</td>
<td></td>
</tr>
<tr>
<td>Q131</td>
<td>Total amount of income, before taxes received, by all of the members in R's household during 2003 was above or below $45,000.</td>
<td></td>
</tr>
<tr>
<td>Q132</td>
<td>If below 45,000. Total amount of income, before taxes, received by all of the members in R's household during 2003.</td>
<td></td>
</tr>
<tr>
<td>Q133</td>
<td>If above $45,000. Total amount of income, before taxes, received by all of the members in your household during 2003.</td>
<td></td>
</tr>
<tr>
<td>INCOME</td>
<td>Combined Income</td>
<td></td>
</tr>
<tr>
<td>Q144</td>
<td>Gender of Respondent</td>
<td></td>
</tr>
<tr>
<td>STATE</td>
<td>State</td>
<td></td>
</tr>
<tr>
<td>CENR</td>
<td>Census Region</td>
<td></td>
</tr>
<tr>
<td>WHITE</td>
<td>GEO-CODE: Race of POP - White (%) based on current projections from CPS</td>
<td></td>
</tr>
<tr>
<td>BLACK</td>
<td>GEO-CODE: Race of POP - Black (%) based on current projections from CPS</td>
<td></td>
</tr>
<tr>
<td>HISPANIC</td>
<td>GEO-CODE: Race of POP - Hispanic (%) based on current projections from CPS</td>
<td></td>
</tr>
</tbody>
</table>

David webinar2013

Share this URL: www.icpsr.umich.edu/icpsrweb/instructors/tables/view/421

Variables

Row: Q11 - Race R self-identifies

Column: Q144 - Gender of Respondent

Type of Chart: Bar Chart

Weight: Population Weight to US Population over Age 18: 209,128,094
## Frequency Distribution

Cells contain:
- Column percent
- Weighted N

<table>
<thead>
<tr>
<th>Race Category</th>
<th>Q144 1: Male</th>
<th>Q144 2: Female</th>
<th>ROW TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q11 1: White/Caucasian</td>
<td>68.2%</td>
<td>67.9%</td>
<td>68.0%</td>
</tr>
<tr>
<td></td>
<td>60,565,494</td>
<td>81,669,512</td>
<td>142,235,006</td>
</tr>
<tr>
<td>Q11 2: Black/African American</td>
<td>9.7%</td>
<td>12.7%</td>
<td>11.4%</td>
</tr>
<tr>
<td></td>
<td>8,587,885</td>
<td>15,322,476</td>
<td>23,910,360</td>
</tr>
<tr>
<td>Q11 3: Asian</td>
<td>4.9%</td>
<td>2.9%</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>4,342,282</td>
<td>3,516,928</td>
<td>7,859,110</td>
</tr>
<tr>
<td>Q11 4: Native American</td>
<td>1.0%</td>
<td>1.1%</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>856,093</td>
<td>1,302,673</td>
<td>2,158,767</td>
</tr>
<tr>
<td>Q11 5: Some other race (Please Specify Single Race)</td>
<td>5.7%</td>
<td>5.2%</td>
<td>5.4%</td>
</tr>
<tr>
<td></td>
<td>5,050,065</td>
<td>6,196,524</td>
<td>11,246,588</td>
</tr>
<tr>
<td>Q11 6: Hispanic/Latino</td>
<td>6.6%</td>
<td>7.0%</td>
<td>6.8%</td>
</tr>
<tr>
<td></td>
<td>5,832,015</td>
<td>8,383,522</td>
<td>14,215,537</td>
</tr>
<tr>
<td>Q11 7: Mixed (Please Specify Mutli Race)</td>
<td>2.8%</td>
<td>2.1%</td>
<td>2.4%</td>
</tr>
<tr>
<td></td>
<td>2,452,662</td>
<td>2,578,869</td>
<td>5,031,531</td>
</tr>
<tr>
<td>Q11 8: Don’t Know</td>
<td>0.8%</td>
<td>0.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>721,942</td>
<td>308,899</td>
<td>1,030,842</td>
</tr>
<tr>
<td>Q11 9: No Answer</td>
<td>0.5%</td>
<td>0.8%</td>
<td>0.7%</td>
</tr>
<tr>
<td></td>
<td>446,580</td>
<td>993,774</td>
<td>1,440,354</td>
</tr>
<tr>
<td>COL TOTAL</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>88,855,018</td>
<td>120,273,076</td>
<td>209,128,094</td>
</tr>
</tbody>
</table>
Online Analysis
Assistance is available

- Our user support staff are available from 8:00 a.m. to 5:00 p.m., ET, Monday to Friday
  - Phone: 734-647-2200
  - netmail@icpsr.umich.edu
References


Thank you!

Q&A