

Using Data in Research

Library Data Archive Fall Quarter 2017

Psych M144/Stats M154 Fall 2017

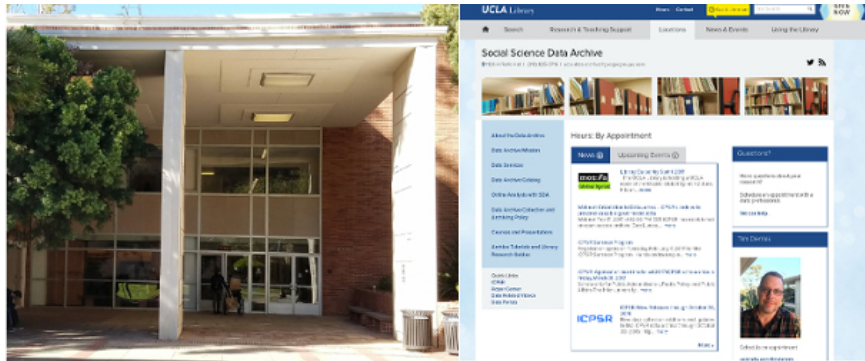
Today's Topics

- UCLA Library Data Archives
- The role of data in your research
- How to search for and access data for research
- Resources, tools and support available to students at UCLA
- Publishing and Data Management Plans

UCLA Library Data Archives

(formally the Social Science Data Archive)

- On the UCLA campus since 1961 [About the Data Archive](#)
- Available to all UCLA faculty and students
- Holdings include public opinion polls, census data, crime stats, political and social behavior, etc.



<https://www.library.ucla.edu/location/social-science-data-archive>

Psych M144/Stats M154
Fall 2017

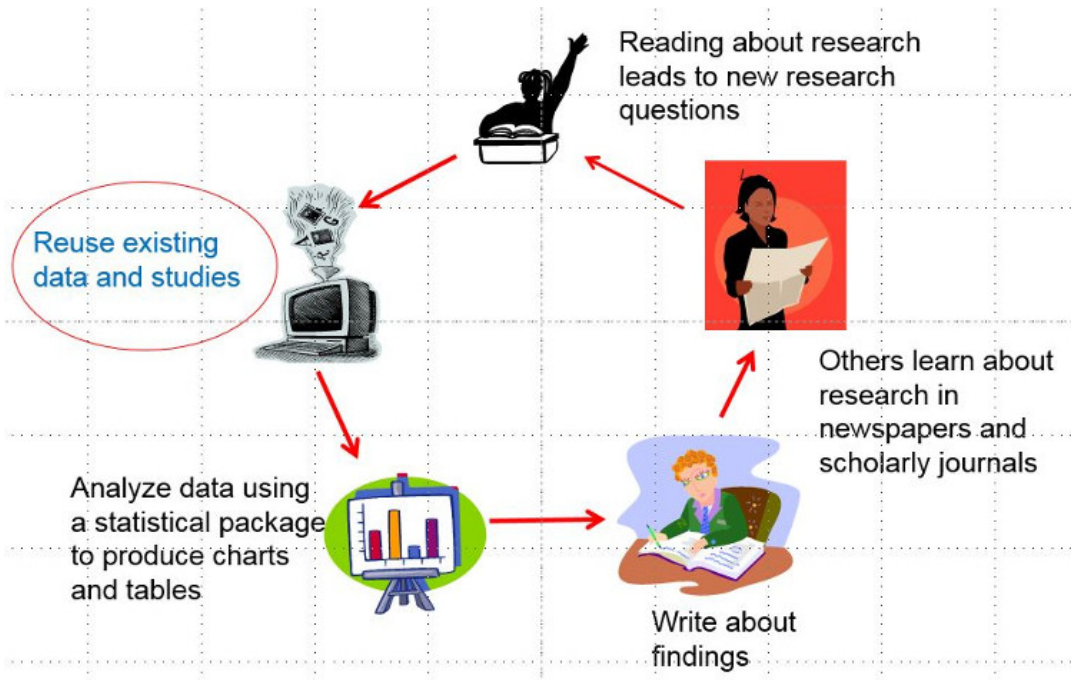
Data Archive: Our services

- Reference: assist students to find and use data
- Data Management Plans
- Metadata
- Provide training: [Software/Library carpentry](#) / [Software Carpentry R-language](#)



- Host events and lectures: [Data Archive News and Upcoming Events](#)

Role of Data in Research



Finding Out About Data Sources

- UCLA Library Data Archive
- Colleagues
- Public media (newspapers, news)
- Social media (twitter, etc)
- Scholarly literature
- Government reports and documents
- Archives and project websites
- Codebooks, questionnaires
- Exploration of datasets

Search Strategies and Defining Your Research

- General to specific; broad to narrow
- Names of investigators; names of studies
- Studies described in literature, news and websites

Geography	Physical areas	Government units	Political boundaries	Other – school districts, etc.	
Time period	Current		Historical		Over time
Population & Units of analysis	Gender, age, etc.	Race, ethnicity, ancestry, etc.	Persons	Households	Other – schools, hospitals, etc.
Context	What is your hypothesis? What will the data tell you? How will you interpret it?				
Variables in the dataset	What variables do you need to answer your research question?				

Psych
M144/Stats
M154 Fall
2017

How do I know this study useful?

- Raw data are **not** eye-readable
- To decipher data you will use **Questionnaires and Codebooks**

Dump of Data -- This is an example of a raw data file.

```
U C L A  CAMPUS COMPUTING NETWORK
-----3068----- 1 0 289----- $0000214200000581 000000000000000000000
000021420000000000000000000000000000000000000000000000000022070
-00000000000001-0000000000000100000000003290-00000000000001 00000000002035
000000000022070-00000000000001 -00000000000001 00000000003285-0000000
0000001 0000000000203500001 140000002500000250000094 000001 2800000350000026
000002 1000002 70 0000 15000001 2000001 10000000000043000001 70000001 390000104
000002 70000001 100000050000001 7000000000001 150000020000003000000880000106
0000025000003100000030000001 700000021 000001500000190000001 6000005700000157
00000151000001 000000023000000130000007000000200000 14000000200000040000000
0000003000000200000000000001 000000000000030000002000000310000002000001
0000000000000000000000000020000002 0000004000000300000001 00000010000002
000000000000001 00000001 00000006000000200000000000020000001 0000000000002083
00000023000000000001000000 17000004500000000000 024 000001300001900000460
000004000004 200000 140000160000005000001 00000000000000000000000000000004
0000000000000200000000000000004360000 00600000 54 000004360000 105700000038
0000032000000600000006000001 70000000500000000000 00000004 000000120000000
00000001 000000 100000000000000000 100000020000 00500001 05000001 50000000
00000020000002100000000000006000001 300000070000015000000300000000000000
```


What is a Questionnaire?

- A **questionnaire** is a *research instrument* consisting of a series of *questions* and other prompts for the purpose of gathering information from respondents.
- Questionnaires cover separate topics such as:
 - Preferences (e.g. political party)
 - Behaviors (e.g. food consumption)
 - Facts (e.g. gender)
- Questionnaires can include indexes or scales:
 - Latent traits (e.g. personality traits)
 - Attitudes (e.g. towards immigration)
 - An index (e.g. Social Economic Status)

Questionnaire: ABC New Poll

TNS Job #175374
Horsham, PA July 23, 2008
FINAL

Date of poll

ABC News/Discovery Channel/Stanford University Poll #1067: Global Warming Poll

Title of Poll

Hello, I'm (NAME), calling for the ABC News public opinion poll. We're not selling anything, just doing an opinion poll on interesting subjects for the news. Out of all the people age 18 or older who are AT HOME RIGHT NOW, can I please speak to the YOUNGEST (male/female)? (IF NO MALE/FEMALE IN THE HOUSEHOLD, ASK: Then may I please speak to the YOUNGEST (female/male)?)

Question text

1. How would you rate the condition of the natural environment in the world today – excellent, good, fair, poor, or very poor?

1 Excellent
2 Good
3 Fair
4 Poor

5 Very poor
DK/No opinion
NA/Refused

Response categories and values

What is a Codebook?

- An overview of the study or dataset
- Variable names and descriptions
- Column locations (mostly this is for raw data)

May also include:

- Test of questions if the data comes from a survey
- Example of the Questionnaire
- Detail on who responded to the survey

Codebook: General Social Survey

Variable name

Question text

Source: General Social Survey, 1972-2004

HLTHINFO R SOUGHT HEALTH INFORMATION PAST YEAR					
Text of this Question or Item					
792. In the past year, that is, between (INSERT CURRENT MONTH) 1999/2001 and (INSERT CURRENT MONTH) 2000/02 have you looked for information about a health concern or medical problem?					
% Valid	% All	N	Value	Label	
49.8	3.5	1,628	0	NOT AT ALL	
20.6	1.5	675	1	1 OR 2 TIMES	
17.1	1.2	559	2	3-5 TIMES	
12.5	0.9	408	3	6 OR MORE TIMES	
	0.0	13	8	DK	
	0.1	27	9	NA	
	92.9	43,200	.	(No Data)	
100.0	100.0	46,510		Total	
Properties					
Data type:		numeric			
Minimum code defined as valid:		0			
Maximum code defined as valid:		3			
Missing-data codes:		-1,8,9			
Record/column:		1/3360			

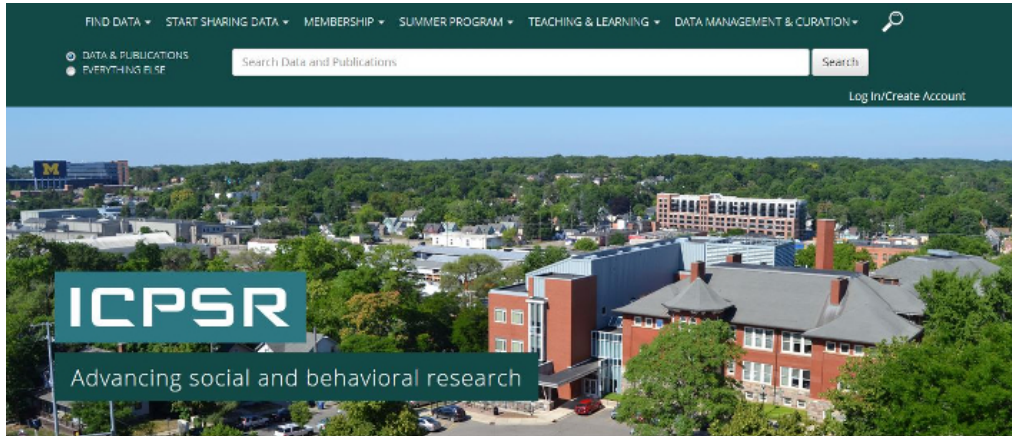
Number of responses

Values and labels

ICPSR Data Site

Inter-university Consortium for Political and Social Research

<http://www.icpsr.umich.edu>



News

More Headlines

[view all news](#)

Psych M144/Stats M154 Fall 2017

This is ICPSR



ICPSR: Find Data

<http://www.icpsr.umich.edu/icpsrweb/ICPSR>

FIND DATA ▾ START SHARING DATA ▾ MEMBERSHIP ▾ SUMMER PROGRAM ▾ TEACHING & LEARNING ▾ DATA MANAGEMENT & CURATION ▾ 🔍

ICPSR

 Find & Analyze Data Log In/Create Account

FIND DATA SEARCH/COMPARE VARIABLES DATA-RELATED PUBLICATIONS RESOURCES FOR STUDENTS HELP

Find Data

More Search Options

- [Browse by topic](#)
- [Browse by series](#)
- [Browse by geography](#)
- [Browse by investigator](#)
- [Browse by thematic collection](#)
- [Browse international data](#)
- [View all studies](#)
 - [View all studies for which online analysis is available](#)
 - [View all replication datasets \(more info\)](#)
 - [View all studies that have learning guides](#)

Restricted-Use Data

ICPSR ensures respondent confidentiality within public use versions of the datasets. The vast majority of ICPSR data holdings are public-use files with no access restrictions. Sometimes the protective measures taken to reduce disclosure risk would significantly degrade the research potential of the data. In these cases, ICPSR provides access to restricted use versions that retain confidential data by imposing stringent requirements for accessing them. [\(more\)](#)



Search Tips

Select a link below to reveal search tips relevant to your search.

[I'm looking for datasets related to a specific topic.](#)

[I'm looking for datasets that contain specific variables.](#)

[I'm looking for datasets that answer a particular research question.](#)

[I'm looking for a specific dataset; I know the title and/or principal investigator.](#)

ICPSR: Register to Access Data

Log into ICPSR

Some functions of the web site, such as downloading and uploading data, require you to log in.

Login via Facebook/Google/LinkedIn

Sign in with

 Google

 LinkedIn

 Facebook

Login via ORCID

[Sign In Using ORCID >](#)

ORCID [®] provides a persistent digital identifier that distinguishes researchers and supports automated linkages between their professional activities ensuring that their work is recognized.

Login via MyData

As an alternative to Facebook, Google, LinkedIn and ORCID, you can create a MyData account. Creating a MyData account is free and does not require you to be associated with an ICPSR member institution.

[Returning User](#)

Email:

Password:

[Log In](#)

[Request your password](#) be sent to you via email.

[New User](#)

Creating a free MyData account will permit you to access features such as notification when data that interest you become available. Further, if you are from a [member institution](#), creating an account permits you to download data available only to ICPSR members.

[Create Account](#)

Note: You must set your browser to accept cookies in order for login to succeed. Your login session will last for nine hours (or until you logout), after which you will be asked to log in again.

ICPSR: Search by Topics

Browse by Topic

Users have two options for identifying studies related to a particular topic.

- [Topic Classifications](#) address topics broadly, and are meant to be a good starting point for exploring the collection.
- The [Thesaurus](#) contains both broad and narrow terms that are more specific to the subject matter of individual studies.
- Our [Thematic Collections](#) are full websites that focus on particular subject areas.

Topic Classifications

I. [Census Enumerations: Historical and Contemporary Population Characteristics](#)

A. [United States](#)

1. [Decennial Censuses](#)

a. [1790-1960 Censuses](#)

b. [1970 Census](#)

c. [1980 Census](#)

d. [1990 Census](#)

e. [2000 Census](#)

f. [2010 Census](#)

2. [American Housing Survey Series](#)

3. [Current Population Survey Series](#)

4. [Other Census, Including County and City Data Books](#)

B. [Nations Other Than the United States](#)

II. [Community and Urban Studies](#)

A. [Studies of Local Politics](#)

1. [United States](#)

2. [Nations Other Than the United States](#)

B. [Detroit Area Studies](#)

C. [Historical Urban and Community Structure Studies](#)

III. [Conflict, Aggression, Violence, Wars](#)

A. [Conflict and Stability Within Nations](#)

B. [Conflict Between and Among Nations](#)

IV. [Economic Behavior and Attitudes](#)

A. [Foundations Series of Consumer Surveys](#)

ICPSR: Search for and Compare Variables

Search/Compare Variables

The Social Science Variables Database (SSVD) enables ICPSR users to examine and compare variables and questions across studies or series. The SSVD currently includes over 4 million variables, representing about 76% of ICPSR's holdings that have quantitative data described in statistical syntax.

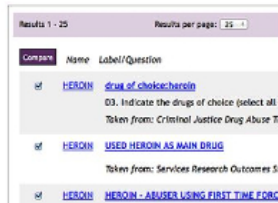
The SSVD does not include our entire collection. For a comprehensive search of all datasets, use our [main data search](#), which indexes all of the documentation available for our studies, including the variable descriptions.

college major

- [Browse by Study](#)
- [Browse by Series](#)

Comparing Variables

Compare variables using the checkboxes and the "Compare" button on the search results page.



Compare	Name	Label/Question
<input type="checkbox"/>	HEROIN	drug_of_choice:heroin D3. Indicate the drugs of choice (select all that apply). Taken from: Criminal Justice Drug Abuse 79
<input type="checkbox"/>	HEROIN	USED HEROIN AS MAIN DRUG Taken from: Services Research Outcomes Study
<input type="checkbox"/>	HEROIN	HEROIN - ABUSER USING FIRST TIME FORC...

Search Tips

- 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

- 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

- A Boolean "and" is implied in the search.
- The search automatically does stemming ([more info](#)); there's no need to type in an asterisk. It's also

ICPSR: Compare Variables

Compare Variables

Compare (3) Clear All (3) Back to search results

<input checked="" type="checkbox"/> MAJOR1	<input checked="" type="checkbox"/> MAJOR2	<input checked="" type="checkbox"/> OC12														
General Social Survey, 1972-2014 (Cumulative File)	General Social Survey, 2012 Merged Data, Including a Cultural Module (United States)	Longitudinal Study of American Youth, 1987-1994, 2007-2011: LSAY Merged Cohort File (Base File)														
COLLEGE MAJOR 1	COLLEGE MAJOR 2	COLLEGE MAJOR														
<p>Question</p> <p>"What was your major or field of study when you received your (RESPONDENT'S COLLEGE DEGREE) degree? IF THE RESPONDENT RECEIVED MORE THAN ONE GRADUATE LEVEL DEGREE, ASK ABOUT THE HIGHEST DEGREE OBTAINEDCOUNT UP TO TWO MENTIONS."</p>	<p>Question</p> <p>What was your major or field of study (RESPONDENT'S COLLEGE DEGREE) degree?</p> <p>Added Text: IF THE RESPONDENT RECEIVED MORE THAN ONE GRADUATE LEVEL DEGREE, ASK ABOUT THE HIGHEST DEGREE OBTAINED RATHER THAN AN M.B.A.). (COUNT UP TO 2 MENTIONS.)</p>	<p>Question</p> <p>COLLEGE MAJOR</p>														
<p>Responses</p> <table border="1"> <thead> <tr> <th>VALUE</th> <th>LABEL</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>IAP</td> </tr> </tbody> </table>	VALUE	LABEL	0	IAP	<p>Responses</p> <table border="1"> <thead> <tr> <th>VALUE</th> <th>LABEL</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>Inapplicable</td> </tr> </tbody> </table>	VALUE	LABEL	0	Inapplicable	<p>Responses</p> <table border="1"> <thead> <tr> <th>VALUE</th> <th>LABEL</th> <th>UN' FRQ</th> </tr> </thead> <tbody> <tr> <td>.99</td> <td>Not asked</td> <td>87</td> </tr> </tbody> </table>	VALUE	LABEL	UN' FRQ	.99	Not asked	87
VALUE	LABEL															
0	IAP															
VALUE	LABEL															
0	Inapplicable															
VALUE	LABEL	UN' FRQ														
.99	Not asked	87														

ICPSR: Study Page and Data

THE DATA SEARCHING VARIABLES DATA-RELATED PUBLICATIONS RESOURCES FOR STUDENTS HELP

High School Longitudinal Study, 2009-2013 [United States] (ICPSR 36423)

Alternate Title: HSLS09

Funded by...
This study is provided by NIMBL.

Principal Investigator(s): United States Department of Education, Institute of Education Sciences, National Center for Education Statistics

Quick Download -
(It takes 4,717,143 KB)

Table of Contents

- Access Notes
- Dataset(s)
- Study Description
- Creation
- Scope of Study
- Methodology
- Variables
- ICPSR
- ICPSR SUPPORT
- Download Statistics

Summary:
The High School Longitudinal Study of 2009 (HSLS09) is a nationally representative, longitudinal study of 9th graders who were followed through their secondary and postsecondary years, with an emphasis on understanding students' trajectories from the beginning of high school into postsecondary education, the workforce, and beyond. Since students decide to pursue various jobs and how well they prepare for this path. The study focuses on answering the following questions: How do parents, teachers, counselors, and students various choices... from 1994

Access Notes

- These data are available to the general public.

Dataset(s)

WARNING: This study is over 115MB in size and may take several minutes to download into a local browser connection.

DS0: Study-Level Files

Documentation: [Documentation.pdf](#) [Documentation.pdf \(Fastest Download\)](#)
[Documentation.pdf \(Cachefirst\)](#) [Documentation.xls \(table guide\)](#)
[Questionnaire.pdf](#) [Report of Information](#) [User Guide.pdf](#)

DS1: Base Year School File - Download All Files (80.1 MB)

Documentation: [Cachefirst.pdf](#)

Download: SAS SPSS Stata B ASCI ExcelCSV
ASCI + SAS Setup SPSS Setup Stata Setup Other

Analyze: [simple cross tab frequency](#) [EQS](#)

Define:

DS2: Base Year, First Year Follow-up and 2013 Update Student File - Download All Files (4,717,143 MB) Large File

Documentation: [Cachefirst.pdf](#)

Download: SAS SPSS Stata B ASCI ExcelCSV
ASCI + SAS Setup SPSS Setup Stata Setup Other

Analyze: [simple cross tab frequency](#) [EQS](#)

Define:

Quick Tables: [Student Demographics by Demographic Category in Act, Related Activities with Students](#)
[Students' Post-Secondary Plans by Act, Related Postsecondary Activities](#)

Study Description

Citation

Psych M144/Stats M154 Fall 2017

ICPSR: Reading About Data

FIND DATA SEARCH/COMPARE VARIABLES DATA-RELATED PUBLICATIONS RESOURCES FOR STUDENTS HELP

About the Bibliography

The *ICPSR Bibliography of Data-related Literature* is a searchable database that as of 2016 contains over 70,000 citations of published and unpublished works resulting from analyses of data held in the ICPSR archive. Developed with support from the National Science Foundation (SES-9977984), the *Bibliography* represents over 50 years of scholarship in the

Data-Related Publications

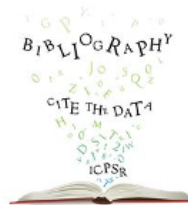
The *ICPSR Bibliography of Data-related Literature* is a continuously updated database of thousands of citations of [works using data held in the ICPSR archive](#). The works include journal articles, books, book chapters, government and agency reports, working papers, dissertations, conference papers,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

Please [take a short survey](#) to help us improve our data-related literature resources.

- [View all citations](#)
- [Browse by author](#)
- [Browse by journal](#)
- [Browse by study](#)

Search Tips

- Our citation search indexes only the literal citations, not the full text



ICPSR: Publications linked to Data

The screenshot shows the ICPSR search interface. At the top, there is a navigation bar with links: FIND DATA, SEARCH/COMPARE VARIABLES, DATA-RELATED PUBLICATIONS, RESOURCES FOR STUDENTS, and HELP. Below this is a yellow banner with a message: "Please take a short survey to help us improve our data-related literature resources." On the left side, there are filter options: "Filter Results", "Pub. Year", "Author", and "Export Results". The main content area is titled "Search Results" and contains a search box with the text "optimism pessimism mental health" and a "REVISE SEARCH" button. Below the search box, there are tabs for "Studies", "Variables", and "Publications (1)". The text "Your query returned 1 citations." is displayed. A "Sort by:" dropdown menu is set to "Relevance". There are "Export" and "Clear All" buttons. The search results are presented in a table with columns for "Pub. Year" and "Citation".

Pub. Year	Citation
1992	Plomin, Robert, Scheier, Michael F., Bergeman, C.S., Pedersen, N.L., Nesselroade, J.R., McClearn, G.E. . Optimism, pessimism and mental health: A twin/adoption analysis . <i>Personality and Individual Differences</i> . 13, (8), 921-930. Full Text Options: DOI Worldcat Google Scholar Export Options: RIS EndNote XML

Results 1 - 1

ICPSR: Publications Linked to Data

[FIND DATA](#) [SEARCH/COMPARE VARIABLES](#) [DATA-RELATED PUBLICATIONS](#) [RESOURCES FOR STUDENTS](#) [HELP](#)

[Please take a short survey](#) [↗] to help us improve our data-related literature resources.

Citation

The following citation is from the [ICPSR Bibliography of Data-related Literature](#):

[Plomin, Robert](#), [Scheier, Michael F.](#), [Bergeman, C.S.](#), [Pedersen, N.L.](#), [Nesselroade, J.R.](#), [McClearn, G.E.](#) . Aug 1992. **Optimism and mental health: A twin/adoption analysis.** *Personality and Individual Differences*. 13, (8), 921-930.

doi:10.1016/0191-8869(92)90009-E

Full Text Options: [DOI](#) [↗] [WorldCat](#) [↗] [Google Scholar](#) [↗]

Export Citation: [RIS](#) [EndNote XML](#)

Related Studies

This publication is related to the following dataset(s):

- [Swedish Adoption/Twin Study on Aging \(SATSA\), 1984, 1987, 1990, 1993, 2004, 2007, and 2010](#)

Data Sources

- ICPSR: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>
- UCLA Library Data Archive Catalog: https://dataverse.harvard.edu/dataverse/ssda_ucla
- Government Sites, Census Data: <https://www.census.gov/>
- IPUMS (Integrated Public Use Microdata Series): <https://www.ipums.org/>
- Public Policy Institute of CA: www.ppic.org
- NBER (National Bureau of Economic Research): <http://www.nber.org/data/>
- CDC/NCHS (Centers for Disease Control and Prevention/National Center for Health Statistics): <https://www.cdc.gov/nchs/surveys.htm>
- Pew Research Center: <http://www.pewresearch.org/>
- Data.gov: <https://www.data.gov/>

Data Management Plans

- What is a Data Management Plans
- Why do I need this?
- What tools are available to help?

What is a Data Management Plan?

A data management plan is a document that describes what you will do with your data *during* your research and *after* you complete your research.

Why do you need this?

- Funding Agencies
- Publishing <https://dataden.library.ucla.edu/jspui/> ORCID:<https://orcid.org/>
- Sharing and Archiving Data
- Regardless of the requirements, good data management is an essential skill for researchers.

Getting Help with Data Management Plans

- Data Archive
- ICPSR

Questions?



Psych M144/Stats M154 Fall 2017

Problems ?

The Data Archive is here to help

1120-H Rolfe
ucla-data-archive@googlegroups.com
310-825-0716

Tim Dennis, Director of the Data Archive
Jamie Jamison, Archive Staff