

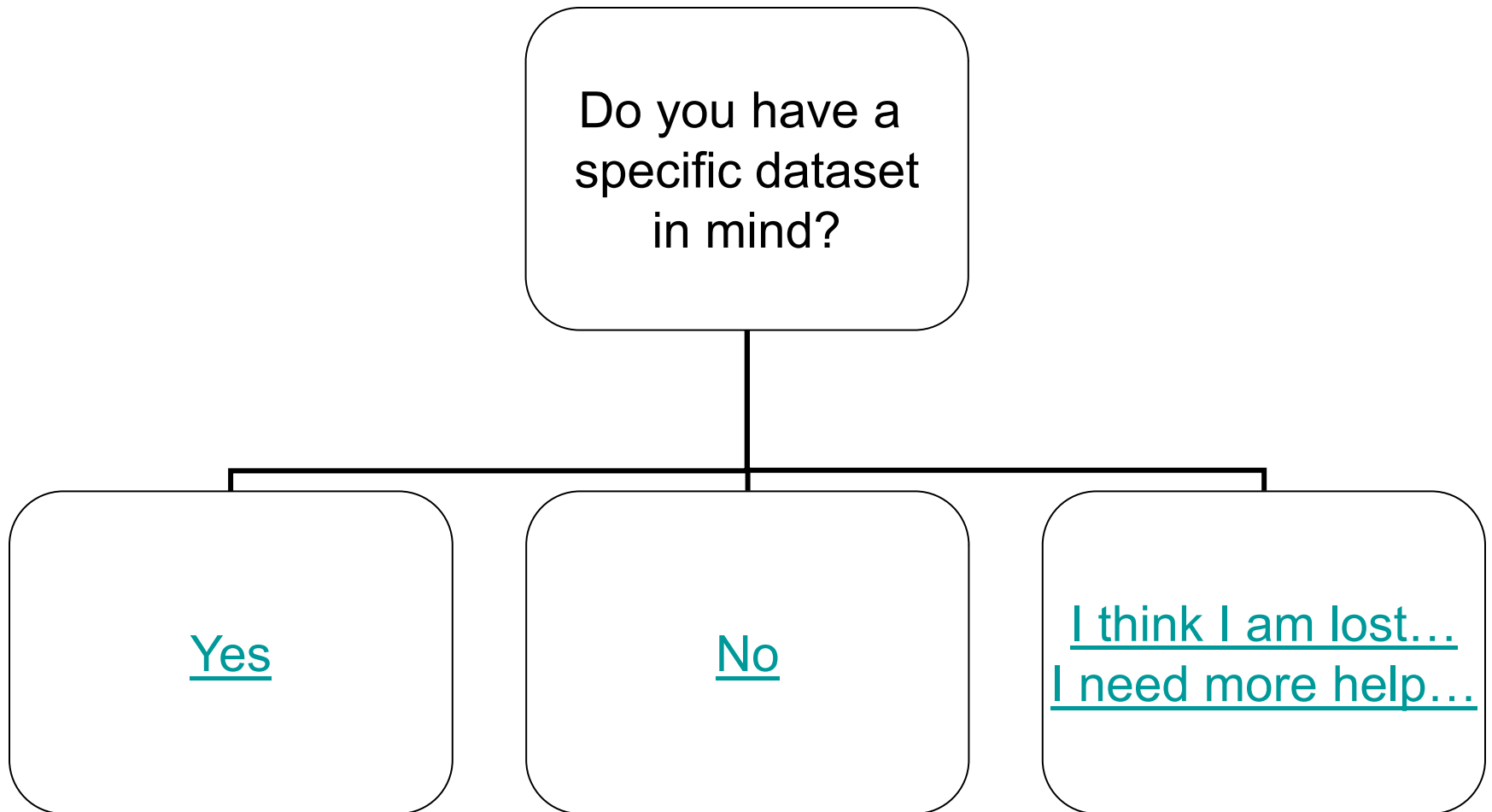


Finding data

(w/comments on crafting research questions)
(ver. 3.5)

Oscar Torres-Reyna
Data Consultant
otorres@princeton.edu

Question...



Now...

What kind of data
do you need?

```
graph TD; A[What kind of data do you need?] --> B[Micro<br/>(individual data, i.e. people, companies)]; A --> C[Macro<br/>(aggregate data, i.e. countries, states, counties)];
```

Micro

(individual data, i.e.
people, companies)

Macro

(aggregate data,
i.e. countries,
states, counties)

Topic/subject/geography...

Do you know the
topic/subject or
geographic area
for your work?

Yes

No

Planning your research...

If your advisor/professor gave you a research topic or a research question it is highly recommended to do a *literature review* focusing primarily on journal articles (obviously mostly quantitative).

For an extensive literature research you can use the “[Subject List of Databases](#)” (available in “[Articles & Databases](#)” at the library website) and check the **abstracts** for the subject area of interest (i.e. economics, politics, sociology, history, etc.). Here is the full link:

<http://library.princeton.edu/catalogs/articles.php?f=subjects>

Using your literature research...

[Back to front](#)

This is important because it can give you a sense of what others have done so you can either:

1. Try to replicate their work (to make sure they did it right) or
2. Use it as basis for your own research by defining topics, geographic coverage, and ***your research question***.

A very important reason is to ***find out the sources of their data***. Once you know the source or even the name of the data set you can go to the DSS page (<http://dss.princeton.edu> for micro data) or to “Articles & Databases” (<http://library.princeton.edu/catalogs/articles.php>) for macro data.

Bear in mind that there are lots of data sources so you may have to be **creative** with your keyword search to find the right source. This is why using the data sources suggested in journal papers or a book can save you lots of headaches.

By the way, it is OK to request data from the authors (just make sure you are not breaking any license agreement).

The research question...

If you have to come up with your own research question here are some steps that may help along the way.

1. It is important for you to pick a topic that you really, really, really like. You will be spending a lot of time doing research so you better enjoy it.
2. Once you pick your topic, it is time to brainstorm to convert the research topic into a research question. Remember that a research question is a statement of what you want to study.
3. Interesting research questions ***do not*** make papers, ***feasible*** research questions make papers. The research question has to be *doable*.

The research question...

To see whether your research question is feasible try the following “suggested feasibility test”:

1. Can you answer the question with a simple descriptive statistic (like an average, median, count, percentage, etc)? If so, then it is just too narrow, not much to research. Not a good research question.
2. Does the answer to your question has too many angles like for example: *what best practices have solved poverty?* Here you need to be more specific: what do you mean by ‘best practices’? Where? The US, the world? The question assumes that poverty has been solved somehow somewhere, does this make sense? Trying to find an answer here may take you everywhere. The question is not focused enough. Too broad.

The research question...

3. A feasible research question is answerable on time. You need to consider deadlines, whether the data is available immediately or not (and in the format you need), whether you need Institutional Review Panel ([IRP](#)) permission to do the research (i.e. would you be interviewing people) which takes at least a month. In sum, do you have enough time and the necessary resources to answer the question?
4. There is no magic procedure to craft a research question. It is a back-and-forth process between you, your advisor and your subject specialist at the library. In many cases it is a frustrating process in which you are expected to bear most of the responsibility. Just remember that your advisor and your subject specialist are there to support your learning process. The earlier you start the better.

The research question...

In a general sense, I can say that a good research question is feasible if...

...it is focused enough, not too narrow not too broad, clearly define the terms, may include the pros and cons of the issue and is answerable on time.

Example: Why some countries are rich while others are poor?

Answers to this question have gone from germs, geography, climate, religion, institutions, evil doers, etc.

Here are some links that may help you:

http://www.esc.edu/ESOnline/Across_ESC/WritersComplex.nsf/wholeshortlinks2/Research+Room+Menu

<http://globetrotter.berkeley.edu/DissPropWorkshop/nuts&bolts/question.pdf>

<http://www.socialresearchmethods.net/kb/strucres.php>

PU/DSS/OTR

Some things to consider...

Once you have your research question, when looking for data, you need to ask yourself:

What is my unit of analysis?

Here is a definition:

“...The unit of analysis is the major entity that you are analyzing in your study. For instance, any of the following could be a unit of analysis in a study:

- individuals
- groups
- artifacts (books, photos, newspapers)
- geographical units (town, census tract, state)
- social interactions (dyadic relations, divorces, arrests)

Why is it called the 'unit of analysis' and not something else (like, the unit of sampling)? Because *it is the analysis you do in your study that determines what the unit is*. For instance, if you are comparing the children in two classrooms on achievement test scores, the unit is the individual child because you have a score for each child. On the other hand, if you are comparing the two classes on classroom climate, your unit of analysis is the group, in this case the classroom, because you only have a classroom climate score for the class as a whole and not for each individual student. For different analyses in the same study you may have different units of analysis. If you decide to base an analysis on student scores, the individual is the unit. But you might decide to compare average classroom performance. In this case, since the data that goes into the analysis is the average itself (and not the individuals' scores) the unit of analysis is actually the group.” [Source:

<http://www.socialresearchmethods.net/kb/unitanal.php>]

Type of data...

The unit of analysis define the type of data you need. If you are studying individuals, households, companies, etc. then you are dealing with micro level data. (go to <http://dss.princeton.edu>)

If your unit of analysis are countries, states or some other form of aggregate data then you have macro level data. (go to <http://library.princeton.edu/catalogs/articles.php>).

The sources of data are different from one type to the other. At DSS we offer mostly micro level but we also provide links to some macro level data.

Final considerations...

If you are still not sure what to do, here are some tips:

1. Talk to your advisor/professor as many times as possible. Remember, he will grade your work.
2. Talk to your subject librarian (go to <http://library.princeton.edu/help/sub.php>)
3. Once you fine tuned your research but have not been able to find what you need, contact DSS at data@princeton.edu
4. If you have what you need but need assistance downloading, reading or analyzing it, contact DSS at data@princeton.edu
5. Keep a log of your research, record all the steps, sources and procedures you followed. This will come in handy when writing the final version of your paper (in particular, the methodological part)
6. For your analysis make sure you have a sound theory behind and a clear research question. This will make easy for you to define your outcome (y) and predictor (x) variable(s) and your model.

Use <http://dss.princeton.edu/> as a gateway to find **micro** level data. Find data by searching, by **subject**, by **geographic coverage** or by **source** (data providers).

Use these if you know your subject area or geographic coverage

[Back to front](#)

Type here the name of the dataset or the keyword(s) from the study.

I think I need to work more on my research project

Princeton University
Data and Statistical Services

Search DSS

☒ Data (DSS/ICPSR /Roper)
☐ DSS site only

Finding Data
• [Data](#)
• [Subject specialists](#)

Using Data
• [Help](#)
• [Software](#)
 • [Stata](#)
 • [SPSS](#)
 • [SAS](#)
 • [R](#)
• [Getting Started](#)
• [Consultants](#)

[About Us](#)

DSS lab
(Monday-Friday)

Sep 1- Nov 7	By appt. only
Nov 8- Dec 17	2 pm - 5 pm*, ot/times by appt.
Dec 18- Jan 30	By appt. only
Jan 31- May 3	2 pm - 6 pm*, ot/times by appt.
May 4- Aug 31	By appt. only

Email data@princeton.edu for an appt./questions.
*No appts. necessary during walk-in hrs.

ACCESS TO THESE DATA FILES IS RESTRICTED TO CURRENTLY ENROLLED/EMPLOYED MEMBERS OF PRINCETON UNIVERSITY

Find Data by Subject

- [Art & Culture](#)
- [Businesses](#)
- [Children & Families](#)
- [Community & Urban Studies](#)
- [Conflicts, Aggression, Wars, Terrorism](#)
- [Consumer Spending](#)
- [Crime & Justice](#)
- [Education](#)
- [Finance](#)
- [Frequently used data](#)
- [Health](#)
- [Housing](#)
- [Income & Employment](#)
- [Longitudinal Surveys](#)
- [Macroeconomic Statistics](#)
- [Politics, Elections, Governing Bodies](#)
- [Population, Demography, Fertility](#)
- [Public Opinion Polls](#)
- [Race, Ethnicity, Ethnic relations](#)
- [Religion](#)
- [Science, Technology, Computers, Internet](#)
- [Sexual Attitudes, Orientation & Behavior](#)
- [Social Attitudes & Values](#)
- [Transportation](#)
- [Women & Gender](#)
- [Young Adults](#)

Find Data by Region

- [Africa](#)
- [Asia](#)
- [Australia & New Zealand](#)
- [Canada](#)
- [Europe](#)
- [Latin America & the Caribbean](#)
- [Middle East](#)
- [Multinational/Crossnational](#)
- [Oceania](#)

Find Data by Major Providers
(you are leaving DSS sites)

- Princeton Library [Main Catalog](#)
- [ICPSR](#): Inter-university Consortium for Social and Political Research
- [Roper Center](#) for Public Opinion Research
- [Sociometrics](#) Social Science Electronic Data Library
- [IPUMS](#): Integrated Public Use Microdata Series
- [WRDS](#): Wharton Research Data Service

You can also use some other major data providers

Use <http://library.princeton.edu/catalogs/articles.php> as a gateway to find **macro** level data. Find data by searching, or by subject.

[Back to front](#)

NOTE: This link gives you also access to articles, journals and some other text databases.

[Princeton University](#) | [Contact the Library](#) | [Library Facts A-Z](#) | [Search This Site](#)


PRINCETON UNIVERSITY LIBRARY

Articles and Databases

[Library Home](#)
[Books, Articles, More](#)
[Research Help](#)
[Library Services](#)
[About the Library](#)

Use these if you know your subject area

Subject List of Databases (Art, Economics, History, Physics...)

or [See all subjects](#)

Reference Tools and more: (Dictionaries, Dissertations, Newspapers...)

or [See all tools](#)

Browse databases by title

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Try PUL Quicksearch:


PUL QuickSearch

Search multiple databases simultaneously

[ProQuest Research Library](#)
[Ebsco Academic Search](#)
[Wilson OmniFile](#)

[Connect from off-campus](#)
[Report a Problem](#)
[Trial databases](#)
[New and Featured Databases](#)

Find a database by title or keyword

Chat Live with a Librarian

Sun 6:00-11:00pm

Mon-Wed 5:00-11:00pm

Sat 1:00-5:00pm

AIM Get research help!

PU librarian is away.

Type your message below and a librarian will respond shortly.

Type here, press "Enter" to send.

Your Name:

[Get your widget at wimzi.com](#)

[Chat Information](#)

I think I need to work more on my research project

Type here the name of the database or the keyword(s) from the study.

Princeton University Library
 One Washington Road, Princeton, New Jersey 08544 USA
 609.258.1470 phone | 609.258.0441 fax | [Web Site Feedback](#)
[Princeton Library A-Z](#) | [Search This Site](#) | [Library Hours](#) | [For Staff](#)

PU/DSS/OTR

© 2008 The Trustees of Princeton University. All rights reserved.
 Last updated: November 13, 2008

15



Search

☒ Site ☐ Datasets [Advanced Search](#)



QUICK LINKS

DATA ACCESS

MEMBERSHIP

EDUCATION

RESEARCH

ABOUT THE CENTER



Search for Datasets

Browse the Center's on-line catalog of studies. Researchers interested in secondary analysis of survey data files can determine studies relevant to their areas of interest. Opinion data are available to members using the [RoperExpress](#) service. Non-members may [contact](#) the Center to place orders.

Search

Search

Reset

[Search Help](#)

Date Range

to

Country

All Countries 

Survey Organization

Type of Sample

☐ Search [RoperExpress](#) Studies Only

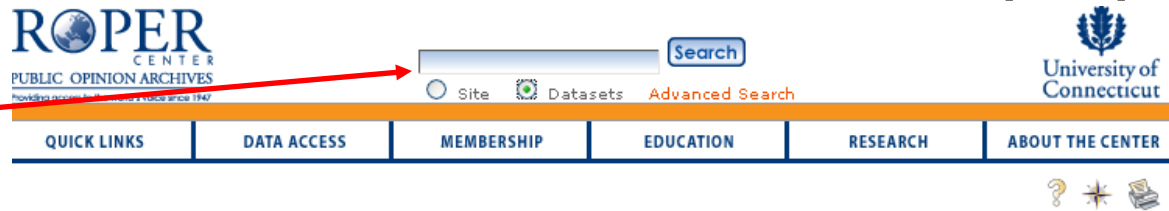
“The Roper Center for Public Opinion Research is one of the world's leading archives of social science data, specializing in data from surveys of public opinion. The data held by the Roper Center range from the 1930s, when survey research was in its infancy”

PU/ © 2008Roper Center, University of Connecticut. All Rights Reserved.

16

Roper, poll data

Search here in the home page for this dataset



Search Results

Title	Time Magazine/Abt SRBI Poll # 2008-4567: America by the Numbers [Study# USSRBI2008-4567]
Survey Firm	Abt SRBI, Inc. (Schulman, Ronca, & Bucuvalas, Inc.)
Survey Sponsor	Time Magazine
Field Dates	October 3-6, 2008
Sample	Registered likely voters
Sample Size	1,053
Sample Notes	Respondents were interviewed via landlines and cellular telephones.
Variables	136
Major Topics Covered	Voter history/intent (3); 2008 presidential election (1); Obama/Biden vs. McCain/Palin (4); rating political leaders/people (10); Sarah Palin vs. Joe Biden (5); Barack Obama vs. John McCain (3); source of news (15); George W. Bush job performance (1); direction of country (1); economy (2); social contract (2); comparing now and then (1); mortgage recovery plan (1); opinion on certain statements (10); war in Iraq (1); people in the news (5).

Metadata



Codebook in two formats



Documentation Download

Study documentation files are available for free download.

- PDF (186kb)
- Word (140kb)

RoperExpress

Datasets, two formats: ACII and SPSS



The following files are available only to **RoperExpress** Users and Members.

Data Sets

- ASCII (385kb)
- SPSS portable (496kb)

Study File Listing and Other Notes

- Text (3kb)

Data Tables/Frequencies

- PDF (54kb)

Marginals



NOTE: When data is not available in Stata, you can download the SPSS portable (*.por), open it using SPSS (available at the DSS lab) and saving it as Stata.

You can use this search page to access data at the ICPSR (<http://www.icpsr.umich.edu>)

ICPSR, survey data

Quick search for data here

[MyData Login/Create Account](#)

MyData options... ▼

☒ Search for Data ☐ Search Web site

Search

[Home](#) | [Find & Analyze Data](#) | [About ICPSR](#) | [Deposit Data & Findings](#) | [Digital Curation](#) | [Membership](#) | [Partners & Projects](#) | [Teaching & Learning](#) | [Help](#)

ICPSR
A PARTNER IN
SOCIAL SCIENCE
RESEARCH
INTER-UNIVERSITY
CONSORTIUM FOR
POLITICAL AND
SOCIAL RESEARCH

Our Mission

ICPSR provides leadership and training in data access, curation, and methods of analysis for a diverse and expanding social science research community.

Announcements

- [ICPSR Business Meeting Presentation - Recording Now Available!](#)
- [ICPSR Receives New Technology Grants](#)
- [New Releases through 2009-09-20](#)
- [2009 Undergraduate Research Paper Winners](#)
- [ICPSR Webinars in Early October - Register Today!](#)
- [2009 ICPSR Web Site - Navigating & Using the New Site!](#)
- [Funding opportunity for secondary data analysis of PreK-3rd data](#)

[See More Announcements...](#)

Featured Site ([view all](#))

Data Documentation Initiative (DDI)



The Data Documentation Initiative (DDI) is an emerging standard for social science metadata that is being developed by an international group called the DDI Alliance. Version 3.0 of the DDI documents the life cycle of research data from the start of a research project through data dissemination.

NOTE: You need to open an account to download datasets (is free)

“Established in 1962, ICPSR is the world's largest archive of digital social science data. We acquire, preserve, and distribute original research data and provide training in its analysis. We also offer access to publications based on our data holdings.”

[Find & Analyze Data](#)

[About ICPSR](#)

[Deposit Data & Findings](#)

[Digital Curation](#)

[Membership](#)

[Partners & Projects](#)

[Teaching & Learning](#)

You can use this page to access data at
IPUMS (<http://www.ipums.org/>)

IPUMS, demographic data



Home of the IPUMS and other data projects

The MPC is one of the world's leading developers of demographic data resources. We provide population data to thousands of researchers, policymakers, teachers, and students. All MPC data are available free over the internet.

Integrated Public Use Microdata Series



IPUMS-International

Harmonized data for 1960 forward, covering 279 million people in 130 censuses from around the world



IPUMS-USA

Harmonized data on people in the U.S. census and American Community Survey, from 1850 to the present



IPUMS-CPS

Harmonized data on people in the Current Population Survey, every March from 1962 to the present

Other MPC Projects



North Atlantic Population Project

Complete-count data from 1800s censuses of Canada, Great Britain, Norway, Sweden, and the U.S.



National Historical Geographic Information System

Tabular U.S. census data and GIS boundary files from 1790 to 2000



Integrated Health Interview Series

Annual harmonized data on people in the U.S. National Health Interview Survey from the 1960s to the present



American Time Use Survey-X

Annual harmonized data from 2003 forward on how U.S. adults divide their time among activities

Copyright © Regents of the [University of Minnesota](#). All rights reserved.
The University of Minnesota is an equal opportunity educator and employer.

You can use this page to access data at WRDS
(<http://wrds.wharton.upenn.edu/>)

You need to request an account (free to PU)

Wharton UNIVERSITY OF PENNSYLVANIA

wrds WHARTON RESEARCH DATA SERVICES
The Global Standard for Business Research

[HOME](#) [NEWS](#) [ABOUT](#) [DEMO](#)

[Home](#) → Wharton Research Data Services

Sign In **Register**

Username

Password

[Forgot your password?](#)

[Sign In >>](#)

Welcome to WRDS!

What is WRDS?

Wharton Research Data Services (WRDS) is a web-based business data research service from The Wharton School at the University of Pennsylvania. Developed in 1993 to support faculty research at Wharton, the service has evolved to become a common tool for research for over 250 institutions around the world.

WRDS is the de facto standard for business data, providing researchers worldwide with instant access to financial, economic, and marketing data through a uniform, web-based interface. This hosted data service has become the locus for quantitative data research and is recognized by the academic and financial research community around the world as the leading business intelligence tool.

WRDS provides access to COMPUSTAT, CRSP, IBES/First Call, NYSE-TAQ, Global Insight, OptionMetrics and other important business research databases.

For additional information, please see the About section.

World Development Indicators database (World Bank)

<http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators>

Once you finish selecting your data, follow these steps to download the file in a format suitable for panel data analysis. The default settings will not give you the data in the proper format.

The screenshot shows the World Bank DataBank interface with several steps highlighted for downloading data in a suitable format for panel data analysis:

- Explore. Create. Share** (1) - The 'Share' button is circled in green.
- Orientation** - The 'Custom' button is selected. The 'Time' dropdown is set to 'Row' (2), the 'Series' dropdown is set to 'Column' (3), and the 'Country' dropdown is set to 'Row' (4). The 'APPLY CHANGES' button is circled in green (5).
- Download Options** (6) - The 'Download Options' button is circled in green.
- Advanced Options** (7) - The 'Advanced Options' button is circled in green.
- Download Format** - The 'Excel' tab is selected. The 'Excel format' dropdown is set to 'xlsx' (8). The 'Data format' dropdown is set to 'Table' (9). The 'NA preference' dropdown is set to 'Blank' (10). The 'Apply scale & precision from rep.' checkbox is checked.
- Export range** - The 'Entire dataset' dropdown is selected.
- Variable format** - The 'Both codes & names' dropdown is selected.
- Metadata** - The 'No' dropdown is selected.
- Download** - The 'DOWNLOAD' button is circled in green (11).

For more on panel data see here:

Stata
<http://dss.princeton.edu/training/Panel101.pdf>

R
<http://dss.princeton.edu/training/Panel101R.pdf>

Set NA to 'Blank'

Some useful links...

[Back to front](#)

- DSS - <http://dss.princeton.edu>
- Articles & Databases (PU) - <http://library.princeton.edu/catalogs/articles.php>
- Libguides at PU - <http://libguides.princeton.edu/dss>
- ICPSR - <http://www.icpsr.umich.edu/>
- Roper Center - <http://www.ropercenter.uconn.edu/>
- IPUMS - <http://www.ipums.org/>
- WRDS - <http://wrds.wharton.upenn.edu/>
- Sociometrics - <http://infotrac.galegroup.com/itweb/prin77918?db=SSEDL>
- NBER - <http://www.nber.org/data/>
- IQSS Dataverse Network - <http://dvn.iq.harvard.edu/dvn/>
- Vanderbilt TV news archive - <http://tvnews.vanderbilt.edu/>
- CQPress Electronic Library - <http://library.cqpress.com/index.php>

THANK YOU!