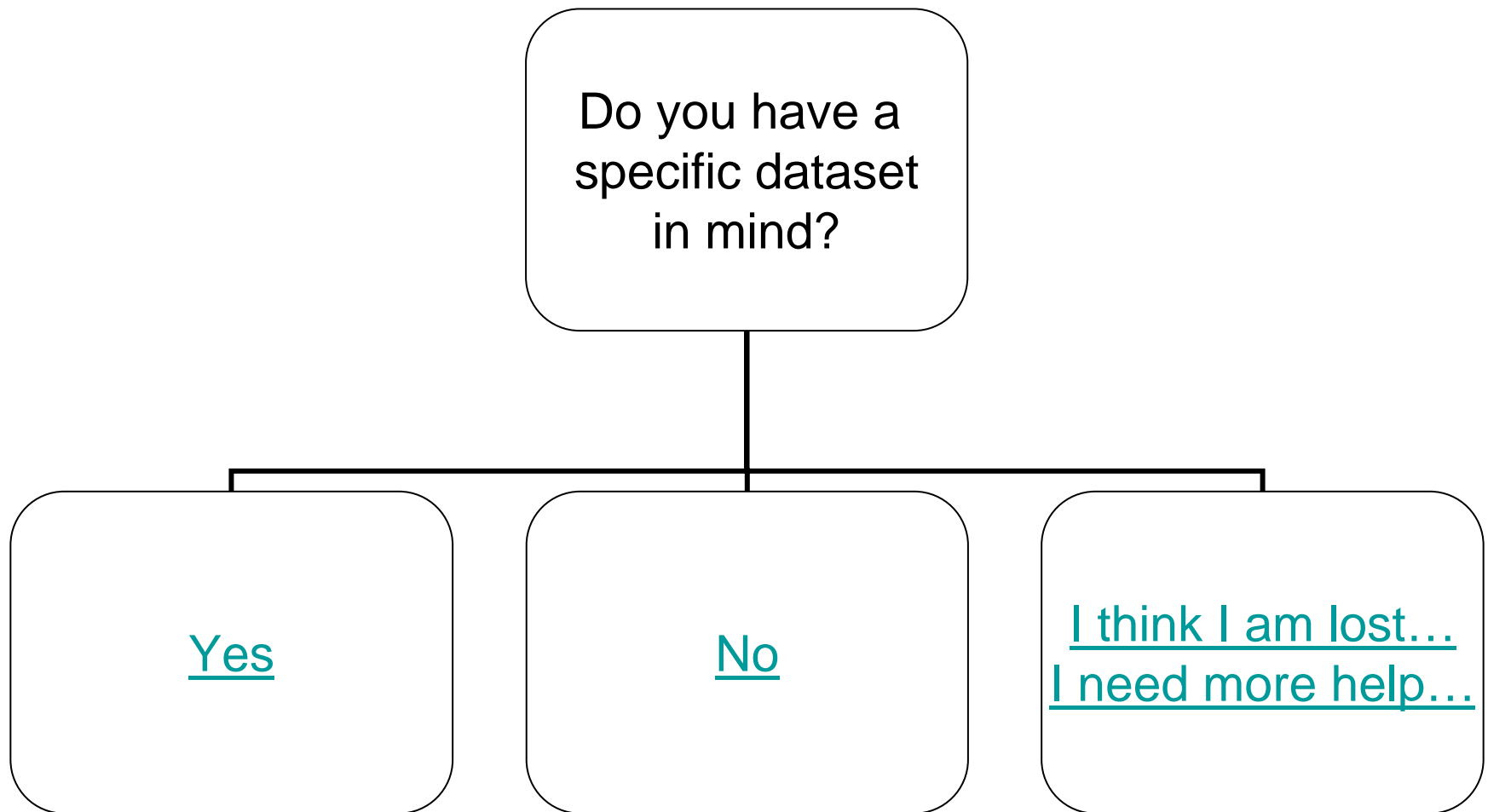


First question...



Now...

What kind of data
do you need?

```
graph TD; A[What kind of data do you need?] --> B[Micro]; A --> C[Macro];
```

Micro

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

Macro

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

Very well...

You think you are
lost because...?

```
graph TD; A[You think you are lost because...?] --- B[ ]; B --- C[I need to organize and analyze my data]; B --- D[I do not know how to define my research topic and/or how to find data];
```

I need to organize
and analyze my
data

I do not know how to define
my research topic and/or
how to find data

Second question...

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

Yes

No

Preparing data...

Your data should have the following structure:

- Variables in columns
- Cases/individuals in rows

Where 'id' could be 'country name', 'persons', 'households', 'states', 'counties', etc.

You need this format to run regressions, frequencies, crosstabulations, summary statistics, and all kinds of statistical analysis you may need.

[Click here](#) for definitions of different type of data

OR

id	var1	var2	var3
1	7.30	32.27	0.10
2	8.28	40.68	0.56
3	3.35	5.62	0.55
4	4.08	62.80	0.83
5	9.09	22.76	0.26
6	8.15	90.85	0.23
7	7.59	54.94	0.42

id	year	var1	var2	var3
1	2000	7	74.03	0.55
1	2001	2	4.60	0.44
1	2002	2	25.56	0.77
2	2000	7	59.52	0.05
2	2001	2	16.95	0.94
2	2002	9	1.20	0.08
3	2000	9	85.85	0.50
3	2001	3	98.85	0.32
3	2002	3	69.20	0.76

Cleaning your variables...

Before doing any serious data analysis but right after you explore your data, you need to prepare your variables for analysis, this is particularly important when dealing with public opinion data or any [categorical](#) data (i.e. non-response) or in the case of continuous data with outliers.

In the example below responses go from 'very well' to 'refused', with code 1 to 6. Lets assume you want to use this variable in a regression. You need to do two things: get rid of non-responses and reorder their valence (from positive-to-negative TO negative-to-positive) to ease the interpretation of the coefficients. This second step is optional but highly recommended for interpretation purposes

```
. tab ecostatu
```

Status of Nat'l Eco	Freq.	Percent	Cum.
Very well	149	10.85	10.85
Fairly well	670	48.80	59.65
Fairly badly	348	25.35	85.00
Very badly	191	13.91	98.91
Not sure	12	0.87	99.78
Refused	3	0.22	100.00
Total	1,373	100.00	

=

```
. tab ecostatu, nolabel
```

Status of Nat'l Eco	Freq.	Percent	Cum.
1	149	10.85	10.85
2	670	48.80	59.65
3	348	25.35	85.00
4	191	13.91	98.91
5	12	0.87	99.78
6	3	0.22	100.00
Total	1,373	100.00	

Cleaning your variables...

[Back to front](#)

Lets clean the variable `ecostatu`. First, **never** work on the original variable, *always keep originals original*.

The command `recode` in Stata lets you create a new variable without modifying the original variable. You could also make a copy of the variable (by typing `clonevar ecostatu_copy=ecostatu`) and work on the copy.

Lets use `recode` in this case:

```
recode ecostatu (1=4 "Very well") (2=3 "Fairly well") (3=2 "Fairly badly") (4=1 "Very badly") (else=.), gen(ecostatu_rec) label(ecostatu_rec)
```

Get frequencies of both variables: `ecostatu` and `ecostatu_rec` to verify:

```
. tab ecostatu
```

Status of Nat'l Eco	Freq.	Percent	Cum.
Very well	149	10.85	10.85
Fairly well	670	48.80	59.65
Fairly badly	348	25.35	85.00
Very badly	191	13.91	98.91
Not sure	12	0.87	99.78
Refused	3	0.22	100.00
Total	1,373	100.00	

```
. tab ecostatu_rec
```

RECODE of ecostatu (Status of Nat'l Eco)	Freq.	Percent	Cum.
Very badly	191	14.06	14.06
Fairly badly	348	25.63	39.69
Fairly well	670	49.34	89.03
Very well	149	10.97	100.00
Total	1,358	100.00	

Now you can use `ecostatu_rec` in a regression since it is a ordinal variable where higher values mean positive opinions.

For additional help on data management, analysis and presentation please check:

<http://dss.princeton.edu/training/>

<http://dss.princeton.edu/>

Defining your research...

It is highly recommended to do a *literature research* 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 [here](#) to find it.

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 headaches.

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

[Hold on, do not go far, we are not over yet...](#)

Some things to consider...

Data comes in different flavors and shapes. So, 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. Even though you had data at the student level, you use aggregates in the analysis. In many areas of social research these hierarchies of analysis units have become particularly important and have spawned a whole area of statistical analysis sometimes referred to as **hierarchical modeling**. This is true in education, for instance, where we often compare classroom performance but collected achievement data at the individual student level.” [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. ([click here for some sources for this kind of data](#))

If your unit of analysis are countries, states or some other form of aggregate data then you have macro level data. ([click here for some sources of this kind of data](#)).

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...

[Back to front](#)

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 ([click here to find him/her](#))
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 found 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 dependent and independent variable(s) and your model.

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

Use these if you know your subject area or geographic coverage

[Back to front](#)

Princeton University **Data and Statistical Services**

Search DSS

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

Finding Data

- [Data](#)
- [Contacts](#)
- [About DSS Data](#)

Using Data

- [Lab](#)
- [Help](#)
- [Software](#)
 - [Stata](#)
 - [SPSS](#)
 - [SAS](#)
- [Getting Started](#)
- [Consultants](#)

About Us

Finding Data: Data

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

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](#)

Data by Region

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

Data by Major Providers
(you are leaving the 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

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

[Got what I need now I need help organizing, analyzing my data](#)

[I think I need to work more on my research project](#)

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 search, or subject.

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

[Back to front](#)

[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...)

Select subject Go

or [See all subjects](#)

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

Select tool Go

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

Got what I need now I need help organizing, analyzing my data

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.

Try 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

Go

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: aimguest55876237

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

[Chat Information](#)




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](#)


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

You can use this search page to access public opinion data at the Roper Center ([click here to access it](#))

Roper, downloading data



☒ 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 Help](#)

Date Range

to

Country



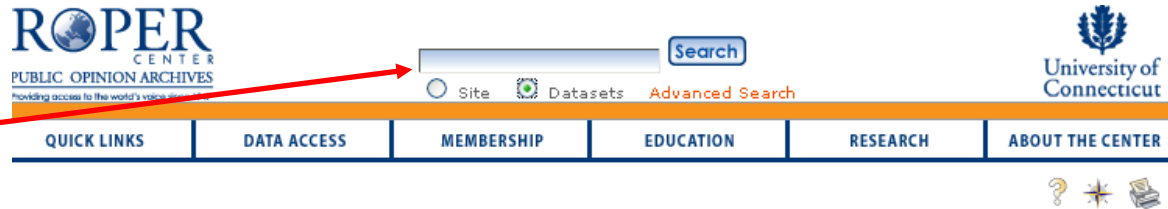
Survey Organization

Type of Sample

☐ Search [RoperExpress](#) Studies Only

Roper, downloading data

Search here in the home page for this dataset



Search Results

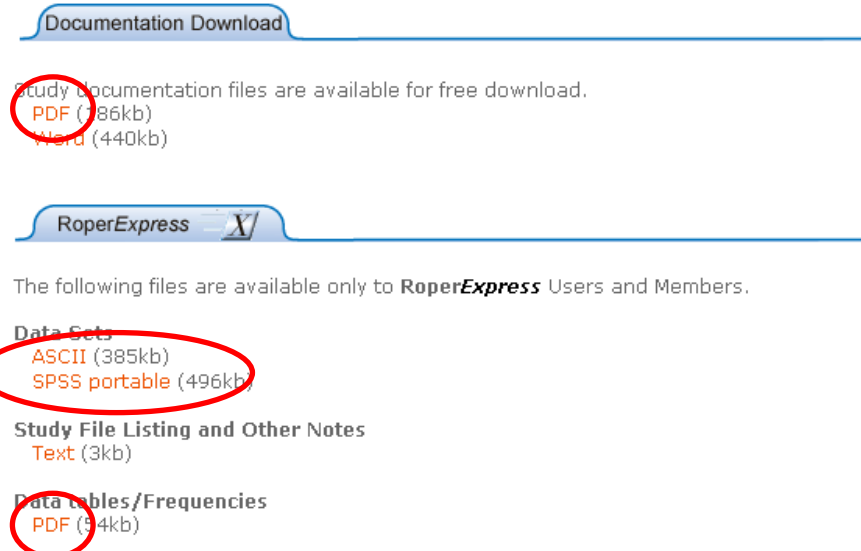
Title	Time Magazine/Abt SRBI Poll # 2008-4567: America by the Numbers [Study# USSRBI2008-4567]
Survey Firm	Abt SRBI, Inc. (Schulman, Ronca, & Bucuvalas, Inc.)
Survey Sponsor	Time Magazine
Field Dates	October 3-6, 2008
Sample	Registered likely voters
Sample Size	1,053
Sample Notes	Respondents were interviewed 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

Datasets, two formats: ASCII and SPSS

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.

ICPSR, downloading data

You can use this search page to access data at the ICPSR ([click here to access it](#))

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR
POLITICAL AND SOCIAL RESEARCH

[MyData Login/Create Account](#)

MyData options... ▼

Data

Courses &
Learning Tools

Our
Research

Membership

About
ICPSR

Help

Home



A partner in
social science
research

Click here to browse data by subject or region

SEARCH

in all fields ▼

☐ Web Site ☒ Data Holdings ☐ Publications

[Advanced Search](#) [Searching Tips](#)

[based on our data](#)

Uncheck the option
"Web Site"

QUICK LINKS

Select a link ▼

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.

NOTE: To download datasets you will need to open an account (which is free)

ANNOUNCEMENTS

2008.11.30
[New data releases...](#)

2008.11.24
[ICPSR undergraduate research paper competition...](#)

2008.11.24
[ICPSR undergraduate internship...](#)

2008.11.21
[Digital preservation management workshop hosted by ICPSR...](#)

2008.11.20
[New strategic plan and mission statement...](#)

2008.11.19
[Webinar recording and slides available...](#)

2008.11.13
[Site unavailable on Dec. 3...](#)

2008.10.06
[ICPSR Online Learning Center - opening doors to quantitative literacy...](#)

[more >>](#)

Feature:

[Data Preservation Alliance for the Social Sciences \(Data-PASS\)](#)



Funded by the Library of Congress, Data-PASS acquires and preserves data from opinion polls, voting records, large-scale surveys, and other social science studies, many of which are considered at risk of being lost.


[view all ICPSR sites](#)

[Careers](#) | [Contact Us](#) | [Privacy Policy](#) | [Official Representative Site](#) | [Accessibility](#)


© 2007 Regents of the University of Michigan. ICPSR is part of the [Institute for Social Research](#) at the University of Michigan.

IPUMS, downloading data

You can use this page to access data at IPUMS ([click here to access it](#))




[MPC HOME](#) | [IPUMS INTERNATIONAL](#) | [IPUMS USA](#) | [IPUMS CPS](#) | [NAPP](#) | [NHGIS](#) | [IHIS](#)




Integrated Public Use Microdata Series International

IPUMS-International is an integrated series of census microdata samples from 1960 to the present. At this time, the series includes 111 samples drawn from 35 countries, with more scheduled for release in the future.



Integrated Public Use Microdata Series USA

IPUMS-USA is an integrated series of representative samples drawn from the U.S. censuses of 1850 to 2000 and the Puerto Rican censuses of 1910 to 2000. IPUMS-USA also includes American Community Survey (ACS) data from 2000 to 2006.



Integrated Public Use Microdata Series CPS

IPUMS-CPS provides integrated data and documentation from the March Current Population Survey (CPS) from 1962 to 2006. The harmonized CPS data is also compatible with the data from IPUMS-USA.

:: Data Project News ::

New IPUMS-International Data

IPUMS-International recently added 32 new samples from 15 countries. The new data release also includes approximately 100 new internationally harmonized variables and 2000 unharmonized variables specific to individual samples.

[more information >>](#)

New IPUMS-USA Data

IPUMS-USA recently added 14 new samples, including data from the Puerto Rican censuses of 1910-2000 and the Puerto Rican Community Survey. The new data release includes nearly 100 new variables, along with improvements to every existing IPUMS sample.

[more information >>](#)

Use it for GOOD - never for EVIL

"IPUMS est un projet absolument extraordinaire...époustouflante [mind-blowing]!"
Liens-Socio (French portal for the social sciences)

"One of the great archival projects of the past two decades"
Journal of American History

North Atlantic Population Project

The North Atlantic Population Project provides access to late-nineteenth century complete-count census microdata from Canada, Great Britain, Iceland, Norway, Sweden, and the United States.

[NAPP Homepage >>](#)

National Historical Geographic Information System

The National Historical Geographic Information System (NHGIS) provides U.S. aggregate census data and electronic boundary files for tracts and counties between 1790 and 2000.

[NHGIS Homepage >>](#)

Integrated Health Interview Series

The Integrated Health Interview Series (IHIS) provides integrated data and documentation from the National Health Interview Survey (NHIS). IHIS currently contains over 1000 integrated variables covering the years 1969 to 2006.

[IHIS Homepage >>](#)

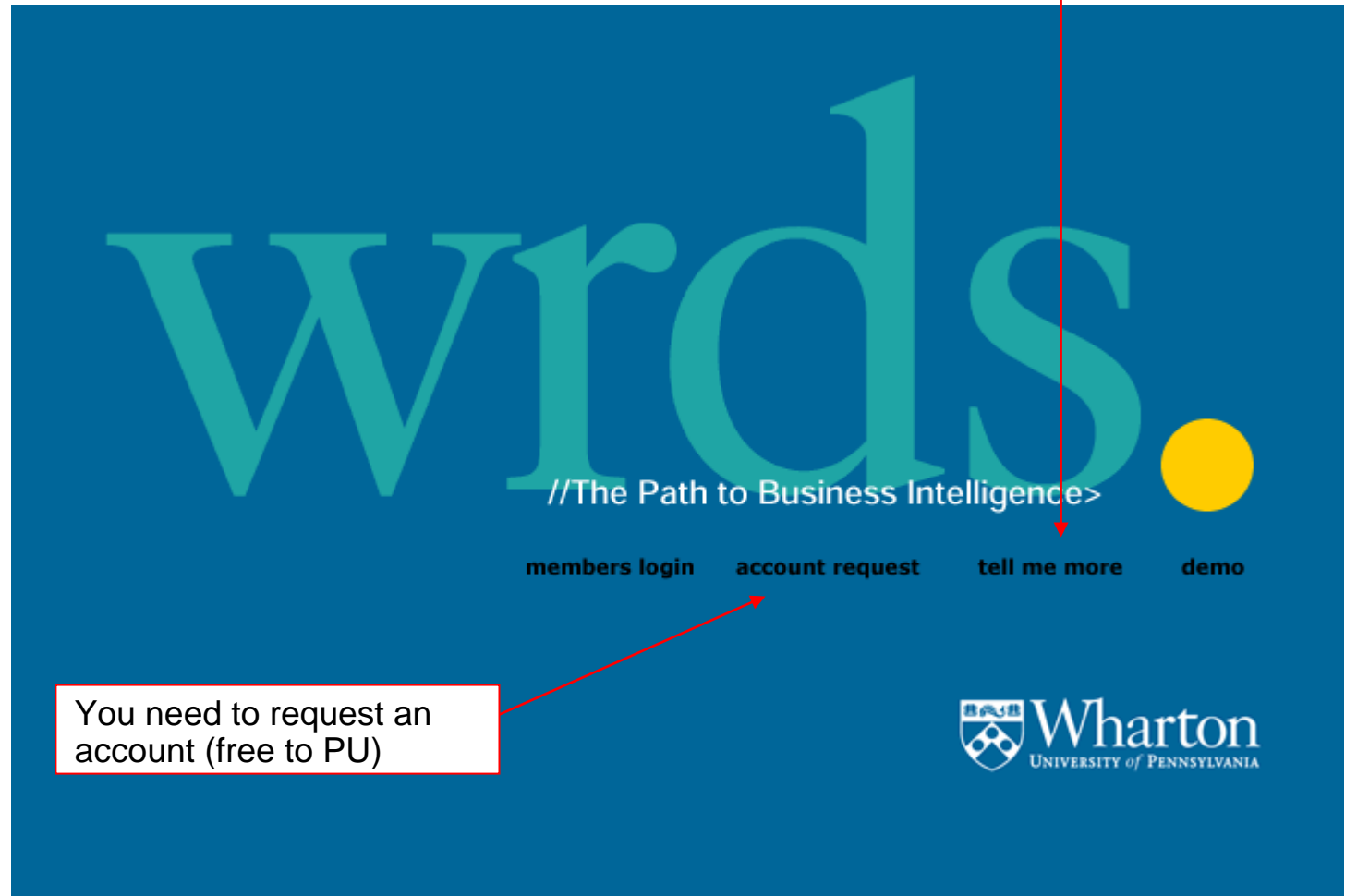
[MPC HOME](#) | [IPUMS INTERNATIONAL](#) | [IPUMS USA](#) | [IPUMS CPS](#) | [NAPP](#) | [NHGIS](#) | [IHIS](#)

Copyright © Minnesota Population Center, University of Minnesota

WRDS, downloading data

You can use this page to access data at WRDS ([click here to access it](#))

Click here for more info on data available at WRDS



You need to request an account (free to PU)

World Development Indicators is available through PU library homepage in “[Articles & Databases](#)”. You can use this link for other macro level data sources.

WDI, downloading data

The World Bank Group

WDI Online

World Development Indicators

Select English

Select Variables Show Report

Country | Series | Time

Classification

- Aggregates
- East Asia & Pacific
- Europe & Central Asia
- Latin America & Caribbean
- Middle East & North Africa
- North America
- South Asia
- Sub-Saharan Africa

Search

Available - 227

Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia

Select Select All

Deselect Deselect All

Selected - 0

Next >>

Select Variables Show Report

Report | Chart | Map

Info Help ?

Use this to export your data to excel or as csv format (Stata reads csv format, SPSS reads both excel and csv)

© 2007 The World Bank Group, All Rights Reserved.

After selecting your countries, series and time, make sure you set 'country' in rows and 'series' in columns. This will ease your analysis when you download your data

Row	Column	Scale	Precision	Notes	Export
Country	Series	Units	0	No	Select
Time	1960				

Show Report

	GDP (constant 2000 US\$)	GDP (constant LCU)	GDP (current LCU)	GDP (current US\$)	GDP deflator (base year varies by country)	GDP growth (annual %)
Argentina	108322324480	105261342720	0

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/>
- 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>



Finding data

(ver. 1.0 *beta*)

Oscar Torres-Reyna
Data Consultant
otorres@princeton.edu

