Introduction to NLS Investigator

(v. 1.5)

Oscar Torres-Reyna

otorres@princeton.edu

August 2013

http://www.princeton.edu/~otorres/

Introduction

This document offers a quick introduction to the NLS Investigator. It follows a basic approach and focus on searching, downloading and putting the data into Stata

If you are not familiar at all with the site, I strongly recommend to follow the example in this document.

It is important to clarify that this document does not cover all the complexities of the NLS site. For more details I suggest to look at the following links:

Getting Started: How to Get the Most from This Site

<u>https://www.nlsinfo.org/content/getting-started</u>

How to Use the NLS Investigator

<u>https://www.nlsinfo.org/InvestigatorGuide/investigator_guide_TOC.html</u>

To start using the NLS Investigator, please go to the following page:

• <u>https://www.nlsinfo.org/investigator/</u>

NLS investigator

https://www.nlsinfo.org/investigator/

NLS Investigator

Welcome, Guest | LOGIN | Register | Search | Help

Log In	Welcome to Investigator
Username:	Sponsored by the Bureau of Labor Statistics, the National Longitudinal Surveys (NLS) are a family of surveys dedicated to tracking the labor market and other life experiences of American men and women.
Password:	Login w account The seven NLS cohorts are: National Longitudinal Survey of Youth 1997 (NLSY97) National Longitudinal Survey of Youth 1979 (NLSY79) NLSY79 Child and Young Adult Older Men Mature Women Mature Women
	Young Men Young Women To access data for any of the seven NLS cohorts use the login box to the left or <u>begin searching</u> as guest.
If you hav account lo	e an NLSY User-Initiated Questions: We're soliciting suggestions for new questions to add to the NLSY97, NLSY79, and child/young adult surveys. Please visit the NLSY User-Initiated Questions page to learn how to make an informal suggestion or submit a formal proposal.
nere	New with updated NLSY79 release: Beta version of Employer History roster now available.
	Attention In the event that Investigator does not appear to be working correctly, first please try to clear your browser cache. If you continue to have issues, please contact <u>usersvc@chrr.osu.edu</u>

NLS Home | NLS Bibliography | Privacy Policy

Selecting a data source



Tagsets tab

NLS Investigator Welcome, Guest Login Register SEARCH Help v Select the study you want to work with: Additional Resources: NLSY97 1997-2010 (rounds 1-14) Errata, Documentation (user's guide, questionnaires and other materials) Custom Weights Released September 17, 2012 To start a new search click here Save / Download Choose Tagsets Variable Search **Review Selected Variables (6)** Codebook Required / Recommended Variables: Required ID Variable - PUBID will always be selected (1 variable) After selecting the study you will see a series of tabs. The Recommended Demographic Variables (5 variables) first one is to choose a *tagset*. Tagsets are basically saved searches, Saved Tagsets (on server): None Available if you have not save them or it is your fist time you will not need this. Upload Tagset (from PC): Browse... No file selected. Upload Notice that by the default six variables will be added to your data: id (can't remove this) and the following demographics (optional): gender, age, race/ethnicity and birthday (month and year). Select "Variable Search" and go to the next slide

"Variable Search" tab

NLS Invest	igator	Welcome, Guest	Login Register	SEARCH Help
Select the study you want to work with: NLSY97 1997-2010 (rounds 1-14) Released Sep	Additional Resources: Errata, Documentation (user's guide, questionnaires and other materials) Custom Weights			×
Choose Tagsets Variable Searc	To start a new search <u>click here</u> Review Selected Variables (6) Codebook Save / Download			
Browse Index Browse Index v Index of Selected Variables Index of Selected Variables Education, Training & Achievement Scores (16972) Education, Training & Achievement Scores (16972) Employment (21210) Household, Geography & Contextual Variables (7047) Dating, Marriage & Cohabitation (1281) Sexual Activity, Pregnancy & Fertility (1427) Children (1508) Parents, Family Process & Childhood (169) Income, Assets & Program Participation (3959) Health (1313) Attitudes, Expectations & Non-cognitive Tests (372) Crime & Substance Use (5863) Survey Methodology (268)	Vith Search Search Options Please browse the index on the left to display variables. This index contains a set of NLSY97 variables commonly used in research and is not the full of the index and a general search. "Variable Search" offers three search modes: by top the index and a general search. Notice that the site does not offer access to the full above). For this exercise we will use the "Search" tab, go to	^{data set.} pics (index), l data sets (se o the next slic	by searching ee the note de	g

Searching for data (1)

The general search offers a variety of options. From the dropdown menu select the type of search you want. See the next slide **NLS** Investigator Welcome, Guest | Login | Register | SEARCH | Help Select the study you want to work with: Additional Resources: NLSY97 1997-2010 (rounds 1-14) Errata, Documentation (user's guide, guestionnaires and other materials) Released September 17, 2012 Custom Weights To start a new search click here Variable Search Review Sele ted Variables (6) Codebook Save / Download Choose Ta isets Browse Index with Search Search Browse Index Create search criteria below: (Choose One) ~ Ŧ Add (Choose One) Area of Interest (pick from list) 0 Variables **Display Variables** Word in Title (pick from list) Word in Title (enter search term) Question Text (enter search term) Question Name (pick from list) Question Name (enter search term) y variables. Reference Number (pick from list) Reference Number (enter search term) Survey Year (pick from list) Codebook (enter search term) Variable Type (pick from list)

 \geq

Reset

Searching for data (2)

NLS Investig	gator			Welcome, Guest <u>Login</u> <u>Register</u>	SEARCH Hel
Select the study you want to work with: NLSY97 1997-2010 (rounds 1-14) Released Septem	v ber 17, 2012 Review Selected	Additional Resources: Errata, Documentation Custom Weights To start a new search o	(user's guide, questionnaires and other mate	You can do conditi search with "AND"	onal '/"OR"
Browse Index Browse Index with Create search criteria below:	Search Search	This is a ' for 'age'	nested' search, we ar within 'demographic i	re looking indicators' Include only intersection	ting (AND) 🔻
Area of Interest (pick from list)	equals	DEMOGRAPHIC INDICATO	ors 🔍	R	emove
Word in Title (enter search term)	contains	age		R	emove
Options Showing 14 of 14 filtered by	Yo All Variables	ou can add m -	ore terms if needed	14 Variables Display	/ Variables
			<u>VARIABLE IIILE</u>		1007
2 2 R25535.00 CV AGE INT DATE F					1998
3 V R38763.00 CV AGE INT DATE F 4 V R54537.00 CV AGE INT DATE F 5 V R72160.00 CV AGE INT DATE F 6 V S15314.00 CV AGE INT DATE F 7 V S20010.00 CV AGE INT DATE F 8 V S38011.00 CV AGE INT DATE F 9 V S54010.00 CV AGE INT DATE F 10 V S75012.00 CV AGE INT DATE F 11 V 100085.00 CV AGE INT DATE F	RS AGE AT INTERVI RS AGE AT INTERVI	EW DATE	Results will show he will see a "+" next to meaning that there available (as a subse Check the square ne select all variables o need.	ere, sometimes you o the 'RNUM' are more variables et of responses). ext to 'RNUM' to or check the ones you	1999 2000 2001 2002 2003 2004 2005 2006 2007
12 V T20111.00 CV AGE INT DATE F	RS AGE AT INTERV	IEW DATE			2008
13 Image: 136015.00 CV AGE INT DATE F 14 Image: 152014.00 CV AGE INT DATE F	RS AGE AT INTERV	IEW DATE			2009

Searching for data (3)

NLS Investig	ator	R11941.00 [CV_AGE_INT_DATE] Surve PRIMARY VARIABLE	y Year: 1997
Select the study you want to work with: NLSY97 1997-2010 (rounds 1-14) Released September Choose Tagsets Variable Search Browse Index Browse Index with Search Create search criteria below:	Additional Resources: Errata, Documentation (user's guide, Custom Weights To start a new search <u>click here</u> Review Selected Variables (20) Codebook arch Search	RS AGE AT INTERVIEW DATE Age as of interview date. 0 0 TO 11: LESS THAN 12 1169 12 1726 13 1858 14 1877 15 1719 16 614 17 21 18 0 19 TO 999: GREATER THAN 18 8984 Refusel(-1) 0	
Area of Interest (pick from list) equilibrium Word in Title (enter search term) control	uals DEMOGRAPHIC INDICATORS age	Don't Know(-2) 0 TOTAL ====> 8984 VALID SKIP(-4) 0 NON-INTERVI	EW(-5) 0
Word in Title (enter search term)	Il Variables	Hard Minimum: [0] Hard Maximum: [25] Lead In: R11940.00[Default] Default Next Question: R11980.00	
RNUM QUESTION NAME		VARIABLE TITLE	YEAR
1	AGE AT INTERVIEW DATE		1997
2 2 R25535.00 CV AGE INT DATE BS (AGE AT INTERVIEW DATE		1998
3 🔽 R38763.00 CV AGE INT DATE RS A			1999
4 🔽 R54537.00 CV AGE INT DATE RS A	AGE AT INTERVIEW DATE		2000
5 7 R72160.00 CV AGE INT DATE RS A	AGE AT INTERVIEW DATE	u hover the cursor over the question	2001
6 S15314.00 CV AGE INT DATE RS A		e a window will pop-up with detail	2002
7 7 \$20010.00 CV AGE INT DATE RS A		mation about that variable	2003
8 S38011.00 CV AGE INT DATE RS A			2004
9 S54010.00 CV AGE INT DATE RS4			2004
10 S75012.00 CV AGE INT DATE RSA			2006
11 V T00085.00 CV AGE INT DATE RSA			2000
12 7 T20111 00 CV AGE INT DATE RS 4			2007
13 V T36015.00 CV AGE INT DATE RS 4			2000
			2010

9

"Codebook' tab

NLS Investi	gator	Welcome, Guest <u>Login</u> <u>Register</u> SEARCH <u>Help</u>
Select the study you want to work with: NLSY97 1997-2010 (rounds 1-14) Released Septer	Additional Resources:	ionnaires and other materials)
	To start a new search <u>click here</u>	
Choose Tagsets Variable Search	Review Selected Variables (20) Codebook Sa	ave / Download
R00001.00 [PUBID] PRIMARY VARIABLE	Survey Year: 1997	Lead In R72976.00 [Default] Default Next
COMMENT: YOUTH CASE IDENTIFICATIO	NN CODE	<u>R05363.00</u>
0 0 998 1 TO 999 999 1000 TO 1999		Selected Variables ☑ R00001.00 ↔ ↔ ↔)>
997 2000 TO 2999 996 3000 TO 3999 998 4000 TO 4999 996 5000 TO 5999 994 6000 TO 6999 994 7000 TO 7999 994 7000 TO 7999	Here you can get additional information on your variables along with some basic stats	Stats as Graphs
23 9000 TO 9999 		NLSY97 Round 1 Parent and Screener Questionnaires <u>NLSY97 Round 1 Youth Questionnaire</u>
Refusal(-1) 0 Don't Know(-2) 0 TOTAL =====> 8984 VALID	SKIP(-4) 0 NON-INTERVIEW(-5) 0	Areas of Interest SYMBOLS TYPE: SYMBOLS
Min: 1 Max:	9022 Mean: 4504.3	
Hard Minimum: [0] Hard Maximum:	[99999939]	
Lead In: <u>R72976.00</u> [Default] Default Next Question: <u>R05363.00</u>		

Downloading data (1)

Once you are satisfied with your search, you can save it as a "tagset'. You can keep the defaults, select a name (in this case we choose 'Age') and click on "Save".

Go to "Advanced Download", see next slide

NLS Investigator	Welcome, Guest	Login	<u>Register</u>	SEARCH	Hel
Select the study you want to work with: Additional Resources: NLSY97 1997-2010 (rounds 1-14) Released September 17, 2012 Custom Weights Additional Resources:					⊻
To start a new search click here Choose Tagsets Variable Search Review Selected Variables (20) Codebook Save / Download					
Save Tagset Basic Download Advanced Download Manage Downloads Choose where to save the tagset of your selected variables: Save to PC Save on our server Tagset Type: By Rnum By Qname with Year 					
Filename: Age Filename must only contain alpha, numeric, Save hyphen or underscore characters. Tagsets saved on your PC or our server can be reloaded at a later time through the 'Choose Tagsets' tab. Tagsets stored on our server with no activity may be deleted after 90 days.					

NLS Home | NLS Bibliography | Privacy Policy

For help, email usersvc@chrr.osu.edu

Downloading data (2)

NLS Investigator Welcome, Guest | Login | Register | SEARCH | Help Y Select the study you want to work with: Additional Resources: NLSY97 1997-2010 (rounds 1-14) Errata, Documentation (user's guide, guestionnaires and other materials) Released September 17, 2012 Custom Weights To start a new search click here **Choose Tagsets** Variable Search **Review Selected Variables (20)** Codebook Save / Download Save Tagset Basic Download Advanced Download Manage Downloads Customize your advanced download: Select the data format you want. For this Create Download of Data example we will select Stata and comma-Tagset (list of selected variables) delimited along with the codebook. SAS® control file (includes the datafile of selected variables) 1 SPSS® control file (includes the datafile of selected variables) If you are an R user, R can easily read STATA® dictionary file of selected variables Codebook of selected variables Stata, SPSS or comma-delimited files. Short Description File Comma-delimited datafile of selected variables (to be read in Excel, etc.) Column headers -- Use Reference Number Question Name (does not guarantee uniqueness) Create Frequency / Table Apply Universe Restrictors (How to use Universe Restrictors) Filename: Age Download 2 Select a name for the file and click "Download" Filename must only contain alpha, numeric. 2 hyphen or underscore characters. Download status appears under 'Manage Downloads' tab. Downloads may be deleted after 10 days of inactivity.

Downloading data (3)

If you set your browser to ask you where to save, a pop-up window will prompt you to select a folder location for the zip file, click "OK" and select the folder.

If it gets downloaded automatically, then the file should be in the "Downloads" folder or the default place for downloads.

NLS Investigator	Welcome, Guest <u>Login</u> <u>Register</u> SEARCH <u>Help</u>
Select the study you want to work with: Additional Resources: NLSY97 1997-2010 (rounds 1-14) Released September 17, 2012 Errata, Documentation (user's guid Custom Weights) To start a new search click here	Le, questionnaires and other materials)
Choose Tagsets Variable Search Review Selected Variables (20) Codebook	Save / Download
Save Tagset Basic Download Advanced Download Manage Downloads	Opening Age.zip
Download Status: All downloads are available. Please click a download link below to begin downloading. All Available Downloads:	You have chosen to open: Age.zip which is a: Compressed (zipped) Folder from: https://www.nlsinfo.org What should Firefox do with this file?
Date Study Name Size Download 1 NLSY97 Age 152.5K download Delete Selected Files Image: Comparison of the selected files	 <u>Open with</u> WinZip (default) <u>Save File</u> Do this <u>a</u>utomatically for files like this from now on.
Click here to start downloading the data	OK Cancel
NLS Home NLS Bibliography Privacy Policy	For help, email usersvc@chrr.osu.edu

Unzipping the files downloaded from the NLS site



Two files are of importance, the *.dct which is the dictionary file that has the layout to read the data. And the do-file (*.do) which has additional commands to format the data once in Stata.

The *.csv file can be read directly by Excel, Stata, SPSS, SAS or R.

Reading the NLS data into Stata

After running infile, the

 variables window will populate with information about your dataset

📆 Stata/SE 12.0 - [Results]	(
File Edit Data Graphics Statistics User Wi	ndow Help			e e
📂 🚽 🖶 🗐 🖸 🕶 📖 - 1 🗹 - 1 🛒 🚯 🔲 1 🛛	8			•
Review T 7 ×	2. New update available; type -update a	all-	 Variables 	т т ×
# Command _rc		Change the working directory	Variable	Label
1 cd "H:\MyData\NLS"	. cd "H:\MyData\NLS"		R0000100	PUBID - YTH ID CODE 1997
2 infile using Age.dct	H: \MyData\NLS	to the location of the NLS files	R0536300	KEY!SEX (SYMBOL) 1997
	. infile using Age.dct		R0536401	KEY!BDATE M/Y (SYMBOL) 1997
		(type in the command window	R0536402	KEY!BDATE M/Y (SYMBOL) 1997
	infile dictionary {	h a lass A	R1194100	CV_AGE_INT_DATE 1997
	R0000100 "PUBID - YTH IN CODE 1997"	Delow)	R1235800	CV_SAMPLE_TYPE 1997
	R0536300 "KEY!SEX (SYMBOL) 1997"		R1482600	KEY!RACE_ETHNICITY (SYMBO
	R0536401 "KEY!BDATE M/Y (SYMOL) 1997"		R2553500	CV_AGE_INT_DATE 1998
	R1194100 "CV AGE INT DATE 1997"	To read the data use the	R3876300	CV_AGE_INT_DATE 1999
	R1235800 "CV SAMPLE TYPE 1997"	to read the data use the	R5453700	CV_AGE_INT_DATE 2000
Review window	R1482600 "KEY!RACE_ETHNICITY (SYMBOL) 1997"	infile command using the	R/216000	CV_AGE_INT_DATE 2001
	R2553500 "CV_AGE_INT_DATE 1998"		51551400	CV_AGE_INT_DATE 2002
(anything typed in	R3876300 "CV_AGE_INT_DATE 1999"	dictionary file (*.dct)	53801100	CV_AGE_INT_DATE 2003
	R7216000 "CV_AGE_INI_DATE 2000"		\$5401000	CV_AGE_INT_DATE 2005
the command	S1531400 "CV_AGE_INT_DATE 2002"	downloaded from the NLS site	\$7501200	CV_AGE_INT_DATE 2006
window will	S2001000 "CV_AGE_INT_DATE 2003" S3801100 "CV_AGE_INT_DATE 2004"	(type in the command window	Properties	ч ×
	S5401000 "CV AGE INT DATE 2005"	(type in the command window	≘ 🔒 i ← 🌩	
appear here)	S7501200 "CV_AGE_INT_DATE 2006"	below)	Variables	
	T0008500 "CV_AGE_INT_DATE 2007"	DCIOWJ	Name	
	T2011100 "CV_AGE_INT_DATE 2008"		Label	
	T3601500 "CV_AGE_INT_DATE 2009"		Type	
	15201400 "CV_AGE_INI_DATE 2010"		Value Label	
	(*) · · · · · · · · · · · · · · · · · ·	Output window	Notes	
	(8984 observations read)	Output window	Data	
			- Label	
	1		Notes	
	Command		Variables	20
	Command	· · · ·	Observations	8,984
	Co	mmand window	Size	701.88K
HANA (Data) NILS			wemory	
The (WyData (14E)				CAP NOM OVR



Opening Stata's do-file editor



Adding value labels to the original variables

Once the file Age-value-labels.do is open, select all the commands starting with "label..." and run them by clicking on the last icon at the top.

🛃 Do	ile Editor - Age-value-labels.do	
File	Edit Tools View Click here to run the selecte	d
	· □ ● A & a a a o o o e e e e e e e e e e e e e e	ч
A	e-value-labels.do Untitled.do Commands	• ×
1	label define v1R0000100 0 "0" Execute Selection (do)	
2	label values R0000100 v1R0000100	
3	label define vlR0536300 1 "Male" 2 "Female" 0 "No Information"	
4	label values R0536300 v1R0536300	
5	label define vlR0536401 1 "1: January" 2 "2: February" 3 "3: March" 4 "4: April" 5 "5: May" 6 "6: J	lune"
6	label values R0536401 vlR0536401	
7	label define vlR1194100 12 "12" 13 "13" 14 "14" 15 "15" 16 "16" 17 "17" 18 "18"	
8	label values R1194100 vlR1194100	
9	label define vlR1235800 1 "Cross-sectional" 0 "Oversample"	
10	label values R1235800 vlR1235800	
11	label define vlR1482600 1 "Black" 2 "Hispanic" 3 "Mixed Race (Non-Hispanic)" 4 "Non-Black / Non-Hispa	nic"
12	Tabel values R1482600 vIR1482600	=
13	Tabel define VIX2553500 12 "12" 13 "13" 14 "14" 15 "15" 16 "16" 1/ "1/" 18 "18"	
14	Tabel Values K2553500 VLK2553500	
15	Tabel deline vikso/6300 12 "12" 13 "13" 14 "14" 15 "15" 16 "16" 1/ "1/" 16 "16"	
17	Tabel Values K30/0500 VLK30/0500	
18	label deline vikatatano vikatatano	
19	label verices version version of the "14" 15 "15" 16 "16" 17 "17" 18 "18" 19 "19" 20 "20" 21 "21"	
20		
21	label define v1S1531400 14 "14" 15 "15" 16 "16" 17 "17" 18 "18" 19 "19" 20 "20" 21 "21"	
22	label values S1531400 v1S1531400	
23	label define v1S2001000 18 "18" 19 "19" 20 "20" 21 "21" 22 "22" 23 "23"	
24	label values S2001000 vlS2001000	
25	label define v153801100 18 "18" 19 "19" 20 "20" 21 "21" 22 "22" 23 "23"	
26	label values \$3801100 v1\$3801100	
27	label define v155401000 20 "20" 21 "21" 22 "22" 23 "23" 24 "24" 25 "25" 26 "26"	
28	label values 55401000 v155401000	
29	label define v1S7501200 20 "20" 21 "21" 22 "22" 23 "23" 24 "24" 25 "25" 26 "26"	
30	label values S7501200 v1S7501200	
31	label define v1T0008500 20 "20" 21 "21" 22 "22" 23 "23" 24 "24" 25 "25" 26 "26"	
32	label values T0008500 v1T0008500	
33	Tabel define vII2011100 23 "23" 24 "24" 25 "25" 26 "26" 27 "27" 28 "28" 29 "29: Age 29 or greater"	-
34		
35	Table1 define virisorisov 27 "24" 25 "25" 26 "26" 2/ "2/" 26 "26" 29 "29"	
30		
38		
39	/* Crossvalk for Reference number & Question name	
40	* Uncomment and edit this RENAME statement to rename variables for ease of use.	
41	* This command does not guarentee unigueness	-
•	m .	•
Execut		IM OVR

Renaming variables

In the file Age-value-labels. do you need to remove the "/*" in row 44 and "*/" in row 65. This will uncomment the commands between them.

One important edit is adding the years to the names of the variables that change over time. In the example below, rows 49, 52-64 had originally the same name except that now each has its corresponding year, this makes them unique. Other minor edits were done in rows 46-48 and 51 (compare to the original)

File	Edit Tools View
D 🖻	🛃 💭 (슈) 💥 🗈 🛍 🔊 🗠 (표 '중 .윤) 🖧 🕒 💽 📮
Ag	-value-labels.do* Untitled.do*
34	label values T2011100 v1T2011100
35	label define v1T3601500 24 "24" 25 "25" 26 "26" 27 "2
36	label values T3601500 v1T3601500
37	label define v1T5201400 25 "25" 26 "26" 2 "27" 28 "2
38	label values T5201400 v1T5201400
39	<pre>/* Crosswalk for Reference number & Question name</pre>
40	* Uncomment and edit this RENAME statement to rename vari
41	* This command does not guarentee uniqueness
42	*/
43	/* *start* */
44	
45	rename R0000100 PUBID // PUBID
46	rename R0536300 GENDER // KEY!SEX
47	rename R0536401 BDATE M // KEY!BDATE M
40	TENAME RUSSONUZ DUALE_I // ALI:DUALE_I
19	TENAME RIISTIO CV_AGE_INI_DAILISS/ // CV_AGE_INI_DAIL
51	rename R1233000 CV_SAMPLE_TIPE // CV_SAMPLE_TIPE
52	rename R2553500 CV AGE INT DATE1998 // CV AGE INT DATE
53	rename R3876300 CV AGE INT DATE1999 // CV AGE INT DATE
54	rename R5453700 CV AGE INT DATE2000 // CV AGE INT DATE
55	rename R7216000 CV AGE INT DATE2000 // CV AGE INT DATE
56	rename S1531400 CV AGE INT DATE2002 // CV AGE INT DATE
57	rename S2001000 CV AGE INT DATE2003 // CV AGE INT DATE
58	rename S3801100 CV AGE INT DATE2004 // CV AGE INT DATE
59	rename S5401000 CV AGE INT DATE2005 // CV AGE INT DATE
60	rename S7501200 CV AGE INT DATE2006 // CV AGE INT DATE
61	rename T0008500 CV AGE INT DATE2007 // CV AGE INT DATE
62	rename T2011100 CV AGE INT DATE2008 // CV AGE INT DATE
63	rename T3601500 CV AGE INT DATE2009 // CV AGE INT DATE
64	rename T5201400 CV_AGE_INT_DATE2010 // CV_AGE_INT_DATE
65	
66	/* *end* */
7	/* To convert variable names to lower case use the TOLOWE

1 Before

Va	riables	T	џ	×
	Variable	Label		A
	R0000100	PUBID - YTH ID CODE 1997		
	R0536300	KEY!SEX (SYMBOL) 1997		
	R0536401	KEY!BDATE M/Y (SYMBOL) 1997		
	R0536402	KEY!BDATE M/Y (SYMBOL) 1997		
	R1194100	CV_AGE_INT_DATE 1997		
	R1235800	CV_SAMPLE_TYPE 1997		=
	R1482600	KEY!RACE_ETHNICITY (SYMBO		
	R2553500	CV_AGE_INT_DATE 1998		
	R3876300	CV_AGE_INT_DATE 1999		
	R5453700	CV_AGE_INT_DATE 2000		
	R7216000	CV_AGE_INT_DATE 2001		
	S1531400	CV_AGE_INT_DATE 2002		
	S2001000	CV_AGE_INT_DATE 2003		
	S3801100	CV_AGE_INT_DATE 2004		
	S5401000	CV_AGE_INT_DATE 2005		
	S7501200	CV_AGE_INT_DATE 2006		Ŧ

3 After

Variables	T	д	×
Variable	Label		*
PUBID	PUBID - YTH ID CODE 1997		
GENDER	KEY!SEX (SYMBOL) 1997		
BDATE_M	KEY!BDATE M/Y (SYMBOL) 1997		
BDATE_Y	KEY!BDATE M/Y (SYMBOL) 1997		
CV_AGE_INT_DATE	CV_AGE_INT_DATE 1997		
CV_SAMPLE_TYPE	CV_SAMPLE_TYPE 1997		=
RACE_ETHNICITY	KEY!RACE_ETHNICITY (SYMBO		
CV_AGE_INT_DATE1988	CV_AGE_INT_DATE 1998		
CV_AGE_INT_DATE1999	CV_AGE_INT_DATE 1999		
CV_AGE_INT_DATE2000	CV_AGE_INT_DATE 2000		
CV_AGE_INT_DATE2001	CV_AGE_INT_DATE 2001		
CV_AGE_INT_DATE2002	CV_AGE_INT_DATE 2002		
CV_AGE_INT_DATE2003	CV_AGE_INT_DATE 2003		
CV_AGE_INT_DATE2004	CV_AGE_INT_DATE 2004		
CV_AGE_INT_DATE2005	CV_AGE_INT_DATE 2005		
CV_AGE_INT_DATE2006	CV_AGE_INT_DATE 2006		Ŧ

DSS/OT

2

For more on rename see http://dss.princeton.edu/training/StataTutorial.pdf#page=33

Looking at the data

If you type browse in the command line you will see the data set. As it is now, each row represents one individual. While you can start working with this format, it is not ideal for panel data analysis.

)ata Edito	or (Browse) - [U	ntitled]												x
File	e Edit	View Data	Tools												
2			7 😫 😤 🛛	ń -											
				1											_
a [CENDER					BACE ETUNE V	CV ACE 1088	CV ACE 1000	<i>c</i> .	V	rishler		1
≊ v	1	PUBID	GENDER	BUATE_M	1001	CV_AGE_INT~E	CV_SAMPLE_~E	KACE_ETHNI~T	CV_AGE_~1988	CV_AGE_~1999	ĥ	3	Eilter variak	les here	
	- 2	2		7	1987	13	1		16	10	-			nes nere	-
sh	- 3	- 3	2	, 9	1983	13	1	2	15	16			Variable	Label	
5	4	4	2	2	1981	15	1	2	17	18			PUBID	PUBID -	Y
	5	5	- 1	10	1982	15	1	2	16	17			GENDER	KEY!SEX	(5
	6	6	2	1	1982	15	1	2	16	17			BDATE_M	KEY!BDA	1
	7	7	1	4	1983	14	1	2	15	16			BDATE_Y	KEY!BDA	1
	8	8	2	6	1981	16	1	4	17	18			CV_AGE_INT_	CV_AGE	<u> </u>
	9	9	1	10	1982	15	1	4	16	17			CV_SAIVIPLE	CV_SAM	P
	10	10	1	3	1984	14	1	4	14	15			CV ACE INT	KEY!KAU	E.
	11	11	2	6	1982	15	1	2	16	17			CV_AGE_INT		
	12	12	1	10	1981	15	1	2	17	18			CV_AGE_INT		-
	13	13	1	11	1984	12	1	2	13	15			CV_AGE_INT		."
	14	14	1	7	1980	17	1	2	18	-5					þ.
	15	15	2	1	1983	15	1	2	15	17		Pr	operties		
	16	16	1	2	1982	15	1	2	16	17		Ξ	Variables		
	17	17	2	11	1981	15	1	2	17	18	Na		Name	PUBID	
	18	18	1	2	1982	15	1	1	16	17	,	Label	PUBID - YT	Н	
	19	19	1	4	1984	12	1	1	14	15			Туре	float	-
	20	20	1	12	1980	16	1	1	17	19			Format	%9.0g	-
	21	21	1	8	1982	14	1	2	16	17		Value Lab	Value Label		-
	22	22	1	6	1982	14	1	2	16	17			Notes Data		-1
	23	23	2	1	1983	14	1	2	15	16			Filename		-1
	24	24	1	6	1984	12	1	2	14	15			Label		-L
	25	25	2	3	1983	13	1	2	15	16			Notes		
	26	26	1	10	1980	16	1	1	18	19	-		Variables	20	
	•										F.		Observations	8.984	۰
l Read	ły							Vars: 20 Ord	er: Dataset Ob	s: 8,984 Filter: O	ff	Мо	de: Browse	8.984 CAP NU	M

Preparing the data for panel analysis

To run panel regression you need to reshape the data so it looks like the example in this document: https://www.princeton.edu/~otorres/Panel101.pdf

For details on how to reshape data see slide 7 in the above documen:

Since we already have a unique id for this dataset, in the command line we can just type

reshape long CV_AGE_INT_DATE, i(PUBID) j(YEAR)

```
. reshape long CV_AGE_INT_DATE, i(PUBID) j(YEAR)
(note: j = 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010)
```

Data	wide	->	long			
Number of obs.	8984	->	125776			_
Number of variables	20	->	8			
j variable (14 values)		->	YEAR			
xij variables:						
CV_AGE_INT_DATE1997 CV_AGE_INT_D	ATE1998	C	V_AGE_INT_DA	ATE2010->CV	_AGE_INT_	DATE

Notice that reshape only applies to variables that are observed over time (i.e. have a year suffix), in this case CV_AGE_INT_DATE. If you have other variables you can add them to the list, for example:

Looking at the reshaped data

If you type browse in the command line you will see that the dataset has only one CV_AGE_INT_DATE variable and all the years are in rows. Here, each row represents an individual per year. Data for individual 1 ends at row 14, data for individual 2 starts at row 15. You can analyze the data using the panel data techniques shown in this document https://www.princeton.edu/~otorres/ Panel101.pdf

🔝 Data Editor (Browse) - [Untitled]														
File	File Edit View Data Tools													
2		1 🖬 🔂 1	7 😫 🕾 🖬	_										
	var9[11]											_		
		PUPTD	VEAD	GENDER	PDATE M	PDATE V		PACE ETHNT-N	CV AGE INT. E				Variables	ņ
S	1	1	1997	dender 2	BUATE_M	1981	CV_3AMPEE_~E	ACE_ETHNI~1	CV_AGE_1N1~E				 Filter variah 	ler here
nap	2	1	1998	2	9	1981	1	4	17					
not	3	- 1	1999	2	9	1981	1	4	18			Ŀ		Label
°	4	1	2000	2	9	1981	1	4	19					PORID - ATH
	5	1	2001	2	9	1981	1	4	20					
	6	1	2002	2	9	1981	1	4	21					
	7	1	2003	2	9	1981	1	4	22			1		
	8	1	2004	2	9	1981	1	4	23					
	9	1	2005	2	9	1981	1	4	24					
	10	1	2006	2	9	1981	1	4	25					
	11	1	2007	2	9	1981	1	4	26					
	12	1	2008	2	9	1981	1	4	27					
	13	1	2009	2	9	1981	1	4	28					
	14	1	2010	2	9	1981	1	4	29				•	•
	15	2	1997	1	7	1982	1	2	14			1	Properties	 д
	16	2	1998	1	7	1982	1	2	16			E	Variables	A
	17	2	1999	1	7	1982	1	2	17				Name	
	18	2	2000	1	7	1982	1	2	18				Label	
	19	2	2001	1	7	1982	1	2	19				Type	
	20	2	2002	1	7	1982	1	2	20				Value Label	=
	21	2	2003	1	7	1982	1	2	21				Notes	
	22	2	2004	1	7	1982	1	2	22			Ē	T Data	
	23	2	2005	1	7	1982	1	2	23			E	∃ Filename	
	24	2	2006	1	7	1982	1	2	-5				Label	
	25	2	2007	1	7	1982	1	2	-5				Notes	
	26	2	2008	1	7	1982	1	2	26			-	Variables	8
	<												Observations	125.776 <u>21</u> 🔻
Read	ły							Vars: 8 Orde	r: Dataset 🔰 Obs: 🛛	125,776 Fil	ter: Off	M	lode: Browse	CAP NUM

Adding value labels (...again, part 1)

Notice that reshaping the data removed the value labels added before (type tab GENDER to check it). If you go back to the do-file Age-value-labels.do, select and copy (Ctrl-C) rows 3-12 (or until you see categories). Copy the code at the end of the do-file (see next slide)

🛃 Do	o-file Editor - Age-value-labels.do	2
File	Edit Tools View	
A	ge-value-labels.do Untitled.do*	• ×
1	label define vlR0000100 0 "0"	
2	label values R0000100 vlR0000100	
3	label define vlR0536300 1 "Male" 2 "Female" 0 "No Information"	
4	label values R0536300 vlR0536300	
5	label define vlR0536401 1 "1: January" 2 "2: February" 3 "3: March" 4 "4: April" 5 "5: May" 6 "6: June"	
6	label values R0536401 vlR0536401	
7	label define vlR1194100 12 "12" 13 "13" 14 "14" 15 "15" 16 "16" 17 "17" 18 "18"	
8	label values R1194100 vlR1194100	
9	label define vlR1235800 1 "Cross-sectional" 0 "Oversample"	
10	label values R1235800 vlR1235800	
11	label define vlR1482600 1 "Black" 2 "Hispanic" 3 "Mixed Race (Non-Hispanic)" 4 "Non-Black / Non-Hispanic"	=
12	label values R1482600 vlR1482600	
13	label define vlR2553500 12 "12" 13 "13" 14 "14" 15 "15" 16 "16" 17 "17" 18 "18"	

Adding value labels (...again, part 2)

Notice that the lines starting with "label values..." have the variables' old names.

To match the old and new names, look at the 'rename' section, rows 46-48, 50, 51.

Manually replace the first name appearing in the lines starting with "label values..." (in this example start with 'R') with the new names as they appear in the rename section (remember that Stata is case sensitive). Select the code (see the 'After' column below) and run it by clicking on the last icon in the do-file.

Before

```
label define vlR0536300 1 "Male" 2 "Female" 0 "No Inf(
label values R0536300 vlR0536300
label define vlR0536401 1 "1: January" 2 "2: February"
label values R0536401 vlR0536401
label define vlR1194100 12 "12" 13 "13" 14 "14" 15 ":
label values R1194100 vlR1194100
label define vlR1235800 1 "Cross-sectional" 0 "Oversamy
label values R1235800 vlR1235800
label define vlR1482600 1 "Black" 2 "Hispanic" 3 "Mix(
label values R1482600 vlR1482600
```



After

tab (PNDED				label	define	v1R0536300	1	"Mal	e" 2	"Fem	ale"	0 "	No I	infor
. CAD GENDER				label	values	GENDER vlR	0536	300						
KEY!SEX				label	define	vlR0536401	1	"1:	Janua:	ry"	2 "2:	Feb	ruar	Ω n
(SYMBOL) 1997	Freq.	Percent	Cum.	label	values	BDATE_M vl	R053	5401						
				label	define	vlR1194100	12	2 "12	" 13	"13"	14	"14"	15	15
Male	64,386	51.19	51.19	label	values	BDATE_Y vl	R1194	4100						
remare	01,390	40.01	100.00	label	define	vlR1235800	1	"Cro	ss-se	ction	al"	0 "0	vers	ampl
Total	125,776	100.00		label	values	CV_SAMPLE_	TYPE	vlR1	23580	0				
				label	define	vlR1482600	1	"Bla	ck" 🛛	2 "Hi	spani	LC"	3 "M	lixed
				label	values	RACE ETHNI	CITY	vlR1	48260	0				

DSS/OTR

For more on adding value labels see slide 34 here https://www.princeton.edu/~otorres/StataTutorial.pdf

The End

Do not forget to save the datafile by either using the menu, go to 'File'->'Save As' or typing:

```
save name-of-your-file, replace
```

This will save the Stata file in the working directory specified at the beginning, it will have extension *.dta. The first time the 'replace' option is not necessary but if after saving you make changes to the dataset you will need to use it to update the file.

Now the data is ready for analysis, see here

https://www.princeton.edu/~otorres/Panel101.pdf

https://www.princeton.edu/~otorres/StataTutorial.pdf

Once again, for more details I suggest to look at the following links:

Getting Started: How to Get the Most from This Site

<u>https://www.nlsinfo.org/content/getting-started</u>

How to Use the NLS Investigator

<u>https://www.nlsinfo.org/InvestigatorGuide/investigator_guide_TOC.html</u>