

Introduction to the NLSY Dataset

Introduction

The **National Longitudinal Surveys (NLS)** are a set of surveys conducted by the US Department of Labor's Bureau of Labor Statistics, designed to gather information at multiple points in time on labor market activities and other significant life events of several population samples of US citizens.

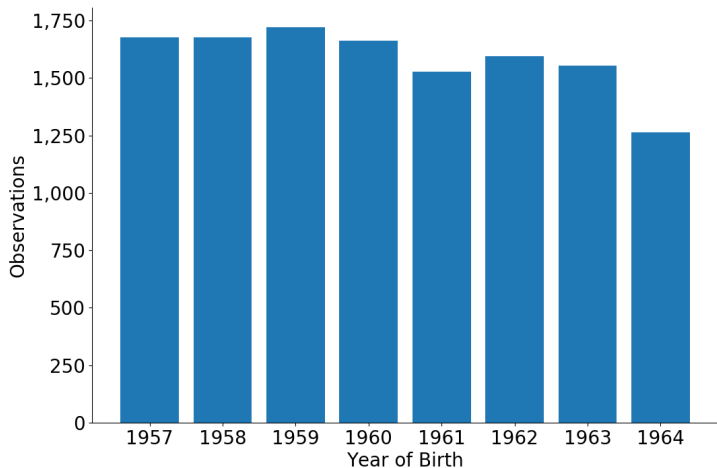
Surveys

Cohort	Birth Years	Interview Years
Older Men	1906-1921	1966-1990
Mature Women	1922-1937	1967-2003
Young Men	1941-1952	1966-1981
Young Women	1943-1953	1968-2003
NLS Youth - 1979 cohort	1957-1964	1979-present
NLS Youth - 1997 cohort	1980-1984	1997-present
NLSY Child and Young Adult	1975-present	1986-present

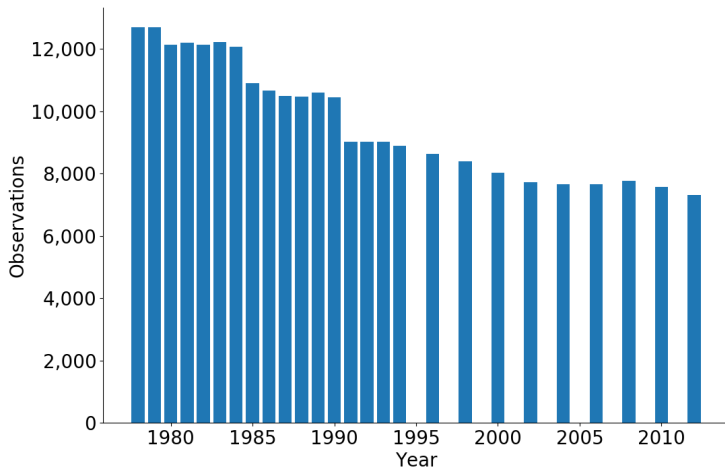
National Longitudinal Survey of Youth, 1979 Cohort (NLSY79)

- ▶ Nationally representative sample of original **12,686** young men and women
- ▶ First interviews in 1979 with respondents **aged 14 to 22**
- ▶ Annual interviewing from 1979 to 1994, **biennial interviews since 1996**
- ▶ Most **recent interviews in 2014** with respondents aged 49 to 58 and a sample size of 7,071

Year of Birth



Number of Observations

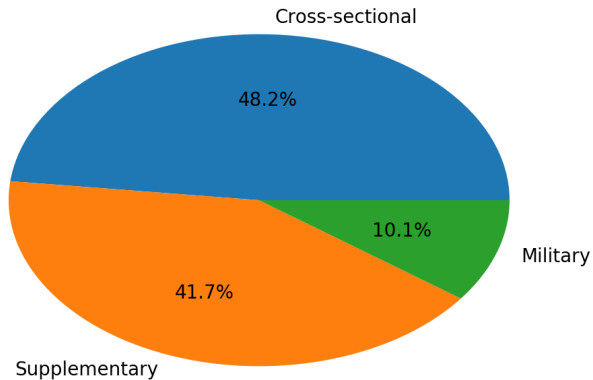


Subsample Design

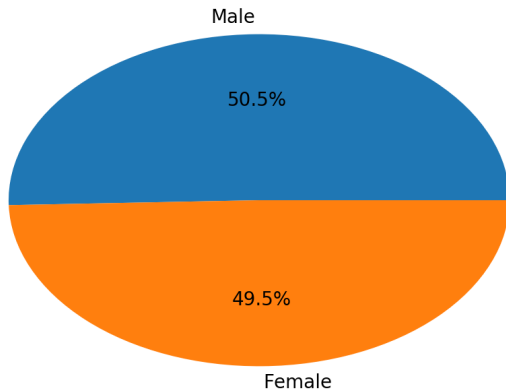
Three subsamples

- ▶ **Cross-Sectional Sample:** 6,111 individuals representing the civilian segment of young people in the US
- ▶ **Supplemental Sample:** 5,295 civilian Hispanic or Latino, black, and economically disadvantaged non-black/non-Hispanic youth
- ▶ **Military Sample:** 1,280 individuals representing the population serving in the U.S. military as of September 30, 1978

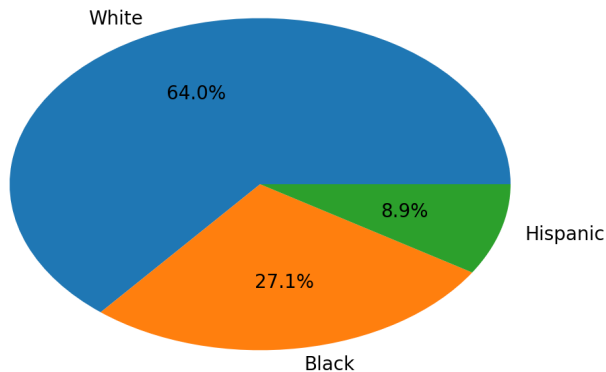
Subsample Design



Gender Composition in the Initial Survey Year



Race/ Ethnicity in the Initial Survey



Variables in the NLSY79

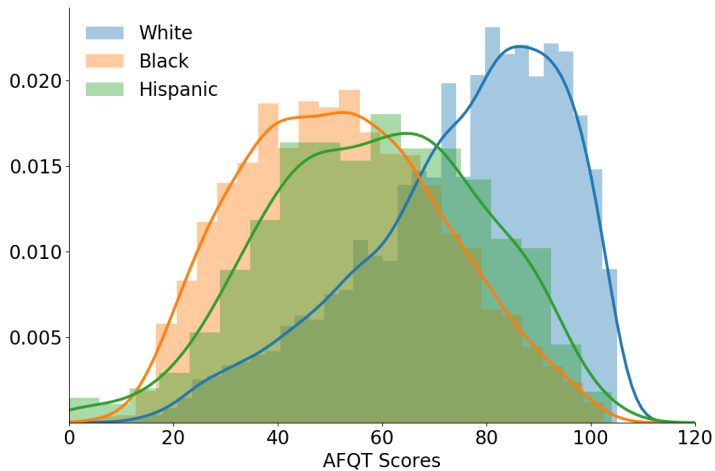
Types of Information gathered in the NLSY79

- ▶ **Education, Training, and Achievement Scores:** Information about high school and college experience, types of training, and cognitive test scores
- ▶ **Labor market behavior:** Extensive history of employment over time, including wages, hours, occupation, industry, and job search
- ▶ **Health:** Health status, substance use, and health insurance
- ▶ **Attitudes, Expectations, and Non-Cognitive Tests:** personality traits, expectations about the future, locus of control
- ▶ And many more...

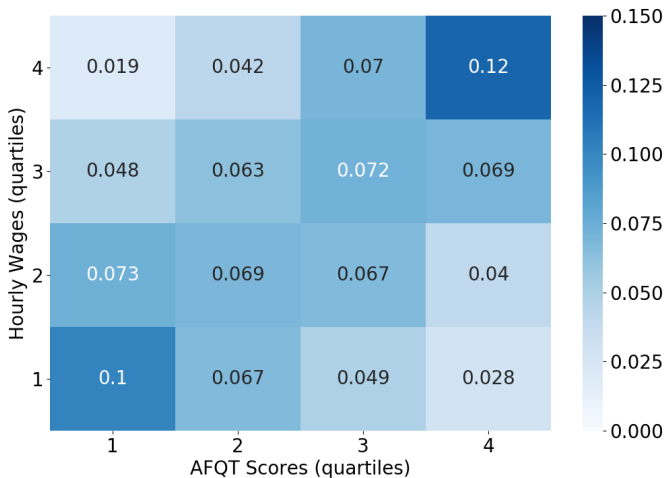
Variables of Cognitive and Non-cognitive Skills

- ▶ **AFQT Scores:** Measure of cognitive skills based on the ASVAB (10 tests in the areas of arithmetic reasoning, math knowledge, word knowledge, coding speed, etc.)
- ▶ **Rotter Score:** Internal-External Locus of Control Scale (Rotter, 1966)
- ▶ **Rosenberg Score:** Degree of approval or disapproval toward oneself (Rosenberg, 1965)

AFQT Score Distribution



Do more able individuals get higher wages?



Longitudinal Availability of Variables

- ▶ Note that only a small share of variables is available for all points in time.
- ▶ Go to <https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/asterisk/i-labor-market-experience-variables> to check for which time points a certain variable is available.
- ▶ Helpful to do this quite early in order to avoid data availability problems at later stages.

Summing up

Pros & Cons

Pros:

- ▶ Covers a long time period
- ▶ Large part of the data set is freely available
- ▶ Lots of helpful internet resources (use them!)

Cons

- ▶ Significant reduction of the sample size over the years
- ▶ Not the whole data set is freely available (might be a problem for replication)

Access to the NLSY Dataset

Access to the NLSY Dataset

- ▶ **Cleaned version** of the NLSY available on https://github.com/OpenSourceEconomics/nlsy_dataset
- ▶ To download the **raw data**, visit <https://www.nlsinfo.org/investigator/pages/login.jsp>
 - ▶ You can either create an account for free or search for data as a guest.
 - ▶ Setting up an account only requires an email address and a username, not necessarily your real name.

Selecting a Data Source

NLS Investigator

Select the study you want to work with:

(Choose One)	▼
(Choose One)	
NLSY79 (National Longitudinal Survey of Youth 1979)	
NLSY79 Child & Young Adult	
NLSY97 (National Longitudinal Survey of Youth 1997)	
Original Cohorts - Mature Women & Young Women	
Original Cohorts - Older Men & Young Men	

[NLS Home](#) | [NLS Bibliography](#) | [Privacy Policy](#)

- ▶ Select the data source that you want to use.
- ▶ For this example, we will work with NLSY79.

Go to the "Review selected Variables" Tab

NLS Investigator

Select the study you want to work with:

NLSY79 (National Longitudinal Survey of Youth 1979) ▼

Released March 01, 2017

Additional Resources:

[Errata, Documentation](#) (user's guide, questionnaires and other materials)
[Custom Weights](#)

To start a new search [click here](#)

Choose Tagsets

Variable Search

Review Selected Variables (5)

Codebook

Save / Download

Options

Refresh List

	<input checked="" type="checkbox"/>	RNUM ▲	QUESTION NAME	VARIABLE TITLE
1	<input checked="" type="checkbox"/>	A0002600	VERSION_R26	VERSION_R26_1
2	<input checked="" type="checkbox"/>	R0000100	CASEID	IDENTIFICATION CODE
3	<input checked="" type="checkbox"/>	R0173600	SAMPLE_ID	SAMPLE IDENTIFICATION CODE
4	<input checked="" type="checkbox"/>	R0214700	SAMPLE_RACE	R'S RACIAL/ETHNIC COHORT FROM SCREENER
5	<input checked="" type="checkbox"/>	R0214800	SAMPLE_SEX	SEX OF R

Note that by default five variables will be added to your data:

- ▶ ID (can not remove this)
- ▶ Gender
- ▶ Race/ethnicity
- ▶ Sample identification code
- ▶ Public release identification number

"Variable search" tab

NLS Investigator

Select the study you want to work with:

NLSY79 (National Longitudinal Survey of Youth 1979) ▼

Released March 01, 2017

Additional Resources:

[Errata, Documentation](#) (user's guide, questionnaires and other materials)
[Custom Weights](#)

To start a new search [click here](#)

Choose Tagsets

Variable Search

Review Selected Variables (5)

Codebook

Save / Download

Browse Index

Browse Index with Search

Search

Index of Selected Variables

- + Education, Training & Achievement Scores (896)
- + Employment (47873)
- + Health (2513)
- + Survey Methodology (180)
- + Household, Geography & Contextual Variables (2696)
- + Marriage & Cohabitation (664)
- + Sexual Activity, Pregnancy & Fertility (1169)
- + Children (3508)
- + Parents, Family Process & Childhood (540)
- + Income, Assets & Program Participation (3351)
- + Attitudes, Expectations & Non-cognitive Tests (847)
- + Crime & Substance Use (365)

Options 

Please browse the index on the left to display variables.
This index contains a set of NLSY79 variables commonly used in research and is not the full data set.

[NLS Home](#) | [NLS Bibliography](#) | [Privacy Policy](#)

Search for Data in the "Search" Tab

NLS Investigator

Welcome, Patrizia | [SEARCH](#) | [Downloads](#) | [Account](#) | [Help](#) | [Logout](#)

Select the study you want to work with:

NLSY79 (National Longitudinal Survey of Youth 1979) ▼

Additional Resources:

[Errata](#) [Documentation](#) (user's guide, questionnaires and other materials)
[Custom Weights](#)

Released March 01, 2017

To start a new search [click here](#)

[Choose Tagsets](#) | [Variable Search](#) | [Review Selected Variables \(5\)](#) | [Codebook](#) | [Save / Download](#)

[Browse Index](#) | [Browse Index with Search](#) | [Search](#)

Create search criteria below:

Include only intersecting (AND) ▼

Area of Interest (pick from list) equals BETWEEN JOB GAPS Remove
Word in Title (pick from list) is MONTH (5511) Add Reset

126 Variables [Display Variables](#)

Options	Showing 126 of 126 filtered by Primary Variables	
RNUM	QUESTION NAME	VARIABLE TITLE
1	R0279200 Q7-10_1_M	MONTH BEGAN 1ST PERIOD NOT EMPLOYED SINCE LAST INT
2	R0279300 Q7-11_1_M	MONTH ENDED 1ST PERIOD NOT EMPLOYED SINCE LAST INT
3	R0280100 Q7-10_2_M	MONTH BEGAN 2ND PERIOD NOT EMPLOYED SINCE LAST INT
4	R0280200 Q7-11_2_M	MONTH ENDED 2ND PERIOD NOT EMPLOYED SINCE LAST INT
5	R0281000 Q7-10_3_M	MONTH BEGAN 3RD PERIOD NOT EMPLOYED SINCE LAST INT
6	R0281100 Q7-11_3_M	MONTH ENDED 3RD PERIOD NOT EMPLOYED SINCE LAST INT
7	R0281900 Q7-10_4_M	MONTH BEGAN 4TH PERIOD NOT EMPLOYED SINCE LAST INT
8	R0282000 Q7-11_4_M	MONTH ENDED 4TH PERIOD NOT EMPLOYED SINCE LAST INT
9	R0282800 Q7-10_5_M	MONTH BEGAN 5TH PERIOD NOT EMPLOYED SINCE LAST INT

The general search offers a variety of options. From the dropdown menu in the search tab select the type of search you want.

Codebook Tab

Choose Tagsets

Variable Search

Review Selected Variables (5)

Codebook

Save / Download

R02792.00 [Q7-10_1_M]

PRIMARY VARIABLE

MONTH BEGAN 1ST PERIOD NOT EMPLOYED SINCE LAST INT

ORIGINAL QUESTION NAME: S0901837

I: DRAW LINES ON ROW C TO REPRESENT PERIODS DURING WHICH THERE ARE NO LINES IN ROW A OR B. USE DATES ENTERED IN ROWS A & B TO INDICATE IN ROW C DATES R BEGAN AND ENDED EACH PERIOD OF NON-EMPLOYMENT. ENTER THE DATES FOR EACH PERIOD INTO BOX A. MOST RECENT PERIOD FIRST.

1 THRU 12 ACTUAL MONTH

UNIVERSE: R >= 16; period not employed since DLI

0	0: < 1
391	1
1174	2
1500	3
698	4
506	5
321	6
286	7
941	8
563	9
272	10
283	11
475	12
0	13
0	14
0	15
0	16
0	17 TO 99999: 17+

7490

Refusal(-1)

0

Don't Know(-2)

0

Invalid Skip(-3)

2

TOTAL ----->

7492

VALID

SKIP(-4)

4649

NON-INTERVIEW(-5)

545

Min:

1

Max:

12

Mean:

5.45

Lead In:

R02791.00[Default]

Default Next Question:

R02792.10

Lead In

[R02791.00 \[Default\]](#)

Default Next

[R02792.10](#)

Selected Variables

R02792.00

◀

▶

↔

⏪

⏩

Stats as Graphs

Questionnaire Links

[1980 Questionnaire](#)

Area of Interest

BETWEEN JOB GAPS

Here you can get some additional information about your variables along with some basic statistics.

26 / 27

Downloading Data

NLS Investigator

Select the study you want to work with:

NLSY79 (National Longitudinal Survey of Youth 1979) ▼

Released March 01, 2017

Additional Resources:

[Errata](#), [Documentation](#) (user's guide, questionnaires and other materials)

[Custom Weights](#)

To start a new search [click here](#)

Choose Tagsets	Variable Search	Review Selected Variables (10)	Codebook	Save / Download
Save Tagset	Basic Download	Advanced Download	Manage Downloads	

Customize your advanced download:

- ☒ **Create Download of Data**
 - ☒ Tagset (list of selected variables)
 - ☐ SAS® control file (includes the datafile of selected variables)
 - ☐ SPSS® control file (includes the datafile of selected variables)
 - ☒ STATA® dictionary file of selected variables
 - ☐ R® Source code (includes the datafile of selected variables)
 - ☒ Codebook of selected variables
 - ☒ Short Description File
 - ☒ Comma-delimited datafile of selected variables (to be read in Excel, etc.)
Column headers -- Use ☒ Use ☐ Reference Number ☐ Question Name (does not guarantee uniqueness)
- ☐ **Create Frequency / Table**
- ☐ **Apply Universe Restrictors** ([How to use Universe Restrictors](#))
- ☐ Notify me by email when download is complete.

Filename:

Filename must only contain alpha, numeric, hyphen or underscore characters.

Download status appears under "Manage Downloads" tab.
Downloads may be deleted after 10 days of inactivity.

[NLS Home](#) | [NLS Bibliography](#) | [Privacy Policy](#)

Select the data format you want (here Stata and comma-delimited). Download the **short description file** for a text-file containing descriptions of the selected variables!

Appendix

Further Information

Further Information

- ▶ For detailed information on all NLS cohorts:
<https://www.bls.gov/nls/handbook/nlshndbk.htm>
- ▶ For an extensive user guide to the NLSY79:
<https://www.bls.gov/nls/79guide/2001/nls79g0.pdf>
- ▶ For an extensive user guide to the NLSY97:
<https://www.bls.gov/nls/97guide/rd5/nls97ugall.pdf>

Further Information (2)

- ▶ For information on how to use NLSY in STATA:

<http://dss.princeton.edu/training/NLS101.pdf>

- ▶ For a tutorial on how to construct comparable sample across NLSY79 and NLSY97:

<https://www.nlsinfo.org/content/getting-started/intro-to-the-nls/tutorials/tutorial-constructing-comparable-samples-across>

- ▶ For information on sample weights/clustering adjustments:

<https://www.nlsinfo.org/content/cohorts/nlsy79/using-and-understanding-the-data/sample-weights-clustering-adjustme>

References

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1–28.