

Metrics

Andriy Koval

Tuesday, June 24, 2014

Contents

Metrics: labeling factors and exploring scales	1
Data preliminaries	1
Labeling Factor Levels	2
Time metrics : Age, Period, Cohort	4
Mapping Church Attendance	6

Metrics: labeling factors and exploring scales

Report explains how the response categories from NLSY97 questionnaire are labeled and demonstrates application of labeled factors in data operations and graphing.

Data preliminaries

This section introduces data space.

Data In Initial point of departure - the [databox](#) of the selected variables, described in the [Methods](#) chapter.

VARIABLE TITLE	Units	Codename	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
CV_SAMPLE_TYPE	0/1	sample	1997														
PUBID, YOUTH CASE IDENTIFICATION CODE	integers	id	1997														
KEYISEX, RS GENDER	m/f	sex	1997														
KEYIRACE_ETHNICITY, COMBINED RACE AND ETHNICITY	b/h/m/o	race	1997														
KEYIBDATE, RS BIRTHDATE MONTH/YEAR	01-12	bmonth	1997														
KEYIBDATE, RS BIRTHDATE MONTH/YEAR	years	byear	1997														
HOW OFTEN PR ATTEND CHURCH IN LAST YEAR?	1-8	attendPR	1997														
WHAT IS PRS CURRENT RELIGIOUS PREFERENCE?	1-8	relprefPR	1997														
WHAT RELIGION WAS PR RAISED IN?	1-8	elraisedPR	1997														
RS AGE IN MONTHS AS OF INTERVIEW DATE	months	agemon	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
RS AGE AT INTERVIEW DATE	years	ageyear	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
# DAYS PER WEEK TYPICALLY FAMILY DOES SOMETHING RELIGIOUS	days	famrel	1997	1998	1999	2000											
HOW OFTEN R ATTENDED WORSHIP SERVICE IN PAST 12 MONTHS	0-7	attend				2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
R DOES NOT NEED RELIGION FOR GOOD VALUES	y/n	values						2002			2005			2008			2011
GOD NOTHING TO DO HAPPENS TO R	y/n	todo						2002			2005			2008			2011
R BELIEVES RELIGIOUS TEACHINGS ARE TO BE OBEYED EXACTLY AS WRITTEN	y/n	obeyed						2002			2005			2008			2011
R PRAYS MORE THAN ONCE A DAY	y/n	pray						2002			2005			2008			2011
R ASKS GOD HELP MAKE DECISIONS	y/n	decisions						2002			2005			2008			2011
WHAT IS R'S CURRENT RELIGIOUS PREFERENCE?	cats:35	relpref									2005			2008			2011
R A BORN-AGAIN EVANGELICAL CHRISTIAN?	y/n	bornagain												2008			2011
IMPORTANCE OF RELIGIOUS FAITH IN DAILY LIFE	0-7	faith												2008			2011
HOW OFTEN R FELT CALM AND PEACEFUL IN PAST MONTH	1-4	calm				2000		2002		2004		2006		2008		2010	
HOW OFTEN R FELT DOWN OR BLUE IN PAST MONTH	1-4	blue				2000		2002		2004		2006		2008		2010	
HOW OFTEN R HAS BEEN A HAPPY PERSON IN PAST MONTH	1-4	happy				2000		2002		2004		2006		2008		2010	
HOW OFTEN R DEPRESSED IN LAST MONTH	1-4	depressed				2000		2002		2004		2006		2008		2010	
HOW OFTEN R HAS BEEN A NERVOUS PERSON IN PAST MONTH	1-4	nervous				2000		2002		2004		2006		2008		2010	
HOW MANY HOURS PER WEEK DOES R WATCH TELEVISION	cats:6	tv						2002					2007	2008	2009	2010	2011
HOW MANY HOURS PER WEEK DOES R USE A COMPUTER	cats:6	computer						2002					2007	2008	2009	2010	2011
CURRENTLY HAVE ACCESS TO INTERNET?	y/n	internet						2002	2003	2004	2005	2006	2007	2008	2009	2010	2011

Figure 3.2 Databox slice of variables selected from NLSY97 for analyses

This [databox](#) corresponds to the dataset **dsL** produced by [Derive_dsL_from_Extract](#) report, given in the Appendix.

```
dsL<-readRDS("./Data/Derived/dsL.rds")
```

Time Invariant									Time Variant																					
sample	id	sex	race	bmonth	byear	attendPR	relprefPR	relraisedPR	year	agemon	ageyear	famrel	attend	values	todo	obeyed	pray	decisions	relpref	bornagain	faith	calm	blue	happy	depressed	nervous	tv	computer	internet	
1	1	2	4	9	1981	7	21	21	1997	198	15	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
1	1	2	4	9	1981	7	21	21	1998	206	17	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
1	1	2	4	9	1981	7	21	21	1999	219	18	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
1	1	2	4	9	1981	7	21	21	2000	231	19	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	3	3	3	3	3	NA	NA	NA	
1	1	2	4	9	1981	7	21	21	2001	243	20	NA	6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
1	1	2	4	9	1981	7	21	21	2002	256	21	NA	2	1	1	1	0	1	NA	NA	NA	4	2	3	2	1	5	1	NA	
1	1	2	4	9	1981	7	21	21	2003	266	22	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	
1	1	2	4	9	1981	7	21	21	2004	279	23	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	4	1	4	1	1	NA	NA	0	
1	1	2	4	9	1981	7	21	21	2005	290	24	NA	1	0	1	0	0	1	21	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	
1	1	2	4	9	1981	7	21	21	2006	302	25	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	4	1	4	1	1	NA	NA	1	
1	1	2	4	9	1981	7	21	21	2007	313	26	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2	6	1
1	1	2	4	9	1981	7	21	21	2008	325	27	NA	1	0	1	0	0	1	21	NA	3	3	3	3	3	3	NA	NA	1	
1	1	2	4	9	1981	7	21	21	2009	337	28	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2	6	1
1	1	2	4	9	1981	7	21	21	2010	350	29	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	3	3	3	3	3	1	6	1	
1	1	2	4	9	1981	7	21	21	2011	360	29	NA	1	0	1	0	0	1	21	NA	1	NA	NA	NA	NA	NA	NA	6	1	
1	2	1	2	7	1982	NA	NA	NA	1997	178	14	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
1	2	1	2	7	1982	NA	NA	NA	1998	196	16	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	

Figure 3.3 Generic dataset used in the current study, view for one [respondent](#)

Labeling Factor Levels

Review of the item reference [cards](#) shows that initially, all items were recorded on some discrete scale, either counting occasions or assigning an integer to a category of response. However, data were saved as numerical values or integers

```
ds<- dsL[,1:(ncol(dsL)/2)]# selects the first half of variables
str(ds)
```

```
'data.frame': 134760 obs. of 30 variables:
 $ sample : int 1 1 1 1 1 1 1 1 1 1 ...
 $ id : int 1 1 1 1 1 1 1 1 1 1 ...
 $ sex : int 2 2 2 2 2 2 2 2 2 2 ...
 $ race : int 4 4 4 4 4 4 4 4 4 4 ...
 $ bmonth : int 9 9 9 9 9 9 9 9 9 9 ...
 $ byear : int 1981 1981 1981 1981 1981 1981 1981 1981 1981 1981 ...
 $ attendPR : int 7 7 7 7 7 7 7 7 7 7 ...
 $ relprefPR : int 21 21 21 21 21 21 21 21 21 21 ...
 $ relraisedPR : int 21 21 21 21 21 21 21 21 21 21 ...
 $ year : int 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 ...
 $ agemon : num 190 206 219 231 243 256 266 279 290 302 ...
 $ ageyear : num 15 17 18 19 20 21 22 23 24 25 ...
 $ famrel : num NA NA NA NA NA NA NA NA NA NA ...
 $ attend : num NA NA NA 1 6 2 1 1 1 1 ...
 $ values : num NA NA NA NA NA 1 NA NA 0 NA ...
 $ todo : num NA NA NA NA NA 1 NA NA 1 NA ...
 $ obeyed : num NA NA NA NA NA 1 NA NA 0 NA ...
 $ pray : num NA NA NA NA NA 0 NA NA 0 NA ...
 $ decisions : num NA NA NA NA NA 1 NA NA 1 NA ...
 $ relpref : num NA NA NA NA NA NA NA NA 21 NA ...
 $ bornagain : num NA NA NA NA NA NA NA NA NA NA ...
 $ faith : num NA NA NA NA NA NA NA NA NA NA ...
 $ calm : num NA NA NA 3 NA 4 NA 4 NA 4 ...
 $ blue : num NA NA NA 3 NA 2 NA 1 NA 1 ...
 $ happy : num NA NA NA 3 NA 3 NA 4 NA 4 ...
 $ depressed : num NA NA NA 3 NA 2 NA 1 NA 1 ...
 $ nervous : num NA NA NA 3 NA 1 NA 1 NA 1 ...
 $ tv : num NA NA NA NA NA 2 NA NA NA NA ...
 $ computer : num NA NA NA NA NA 5 NA NA NA NA ...
 $ internet : num NA NA NA NA NA NA 1 0 1 1 ...
```

`LabelingFactorLevels.R` sourced at the end of [Derive_dsL_from_Extract](#) matches numeric values with response labels from the questionnaire and adds to `dsL` copies of the variables, saved as labeled factors. For estimations routines such as `lm4` or graphing functions such as `ggplot`, the data type (string, numeric, factor) is a meaningful input, so a quick access to both formats frequently proves useful. It is convenient to think that `dsL` contains only

```
ncol(dsL)/2
```

```
[1] 30
```

variables, but each of them has a double, a labeled factor.

```
str(dsL)
```

```
'data.frame':  134760 obs. of  60 variables:
 $ sample      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ id          : int  1 1 1 1 1 1 1 1 1 1 ...
 $ sex         : int  2 2 2 2 2 2 2 2 2 2 ...
 $ race        : int  4 4 4 4 4 4 4 4 4 4 ...
 $ bmonth      : int  9 9 9 9 9 9 9 9 9 9 ...
 $ byear       : int  1981 1981 1981 1981 1981 1981 1981 1981 1981 ...
 $ attendPR    : int  7 7 7 7 7 7 7 7 7 7 ...
 $ relprefPR   : int  21 21 21 21 21 21 21 21 21 21 ...
 $ relraisedPR : int  21 21 21 21 21 21 21 21 21 21 ...
 $ year        : int  1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 ...
 $ agemon      : num  190 206 219 231 243 256 266 279 290 302 ...
 $ ageyear     : num  15 17 18 19 20 21 22 23 24 25 ...
 $ famrel      : num  NA NA NA NA NA NA NA NA NA NA ...
 $ attend      : num  NA NA NA 1 6 2 1 1 1 1 ...
 $ values      : num  NA NA NA NA NA 1 NA NA 0 NA ...
 $ todo        : num  NA NA NA NA NA 1 NA NA 1 NA ...
 $ obeyed      : num  NA NA NA NA NA 1 NA NA 0 NA ...
 $ pray        : num  NA NA NA NA NA 0 NA NA 0 NA ...
 $ decisions    : num  NA NA NA NA NA 1 NA NA 1 NA ...
 $ relpref     : num  NA NA NA NA NA NA NA NA NA 21 NA ...
 $ bornagain    : num  NA NA NA NA NA NA NA NA NA NA NA ...
 $ faith       : num  NA NA NA NA NA NA NA NA NA NA NA ...
 $ calm        : num  NA NA NA 3 NA 4 NA 4 NA 4 ...
 $ blue        : num  NA NA NA 3 NA 2 NA 1 NA 1 ...
 $ happy       : num  NA NA NA 3 NA 3 NA 4 NA 4 ...
 $ depressed    : num  NA NA NA 3 NA 2 NA 1 NA 1 ...
 $ nervous     : num  NA NA NA 3 NA 1 NA 1 NA 1 ...
 $ tv          : num  NA NA NA NA NA 2 NA NA NA NA ...
 $ computer     : num  NA NA NA NA NA 5 NA NA NA NA ...
 $ internet    : num  NA NA NA NA NA NA 1 0 1 1 ...
 $ sampleF     : Ord.factor w/ 2 levels "Cross-Sectional"<...: 1 1 1 1 1 1 1 1 1 1 ...
 $ idF         : Factor w/ 8984 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ sexF        : Ord.factor w/ 3 levels "Male"<"Female"<...: 2 2 2 2 2 2 2 2 2 2 ...
 $ raceF       : Ord.factor w/ 4 levels "Black"<"Hispanic"<...: NA NA NA NA NA NA NA NA NA NA ...
 $ bmonthF     : Ord.factor w/ 12 levels "Jan"<"Feb"<"Mar"<...: 9 9 9 9 9 9 9 9 9 9 ...
 $ byearF      : Factor w/ 5 levels "1980","1981",...: 2 2 2 2 2 2 2 2 2 2 ...
 $ attendPRF   : Ord.factor w/ 8 levels "Never"<"Once or Twice"<...: 7 7 7 7 7 7 7 7 7 7 ...
 $ relprefPRF  : Ord.factor w/ 33 levels "Catholic"<"Baptist"<...: 21 21 21 21 21 21 21 21 21 21 ...
 $ relraisedPRF: Ord.factor w/ 33 levels "Catholic"<"Baptist"<...: 21 21 21 21 21 21 21 21 21 21 ...
```

```

$ yearF      : Factor w/ 15 levels "1997","1998",...: 1 2 3 4 5 6 7 8 9 10 ...
$ agemonF    : Factor w/ 244 levels "146","147","148",...: 45 61 74 86 98 111 121 134 145 157 ...
$ ageyearF   : Factor w/ 21 levels "12","13","14",...: 4 6 7 8 9 10 11 12 13 14 ...
$ famrelF    : Factor w/ 8 levels "0","1","2","3",...: NA NA NA NA NA NA NA NA ...
$ attendF    : Ord.factor w/ 8 levels "Never"<"Once or Twice"<...: NA NA NA 1 6 2 1 1 1 ...
$ valuesF    : Ord.factor w/ 2 levels "FALSE/less Religious"<...: NA NA NA NA NA 2 NA NA 1 NA ...
$ todoF      : Ord.factor w/ 2 levels "FALSE/less Religious"<...: NA NA NA NA NA 2 NA NA 2 NA ...
$ obeyedF    : Ord.factor w/ 2 levels "FALSE/less Religious"<...: NA NA NA NA NA 2 NA NA 1 NA ...
$ prayF      : Ord.factor w/ 2 levels "FALSE/less Religious"<...: NA NA NA NA NA 1 NA NA 1 NA ...
$ decisionsF : Ord.factor w/ 2 levels "FALSE/less Religious"<...: NA NA NA NA NA 2 NA NA 2 NA ...
$ relprefF   : Ord.factor w/ 33 levels "Catholic"<"Baptist"<...: NA NA NA NA NA NA NA NA 21 NA ...
$ bornagainF : Ord.factor w/ 2 levels "NO"<"YES": NA NA NA NA NA NA NA NA NA ...
$ faithF     : Ord.factor w/ 5 levels "Exrtremely"<"Very"<...: NA NA NA NA NA NA NA NA NA NA ...
$ calmF      : Ord.factor w/ 4 levels "All of the time"<...: NA NA NA NA NA NA NA NA NA NA ...
$ blueF      : Ord.factor w/ 4 levels "All of the time"<...: NA NA NA NA NA NA NA NA NA NA ...
$ happyF     : Ord.factor w/ 4 levels "All of the time"<...: NA NA NA NA NA NA NA NA NA NA ...
$ depressedF : Ord.factor w/ 4 levels "All of the time"<...: NA NA NA NA NA NA NA NA NA NA ...
$ nervousF   : Ord.factor w/ 4 levels "All of the time"<...: NA NA NA NA NA NA NA NA NA NA ...
$ tvF        : Ord.factor w/ 6 levels "less than 2"<...: NA NA NA NA NA 2 NA NA NA NA ...
$ computerF  : Ord.factor w/ 6 levels "None"<"less than 1"<...: NA NA NA NA NA 5 NA NA NA NA ...
$ internetF  : Ord.factor w/ 2 levels "No"<"Yes": NA NA NA NA NA NA 2 1 2 2 ...

```

This give a certain flexibility to assemble needed dataset quickly and have access to factor labels. One can alternate between the raw metric and labeled factor by adding “F” suffix to the end of the variable name:

```

selectCols<-c("year","id","byear","attend","attendF") # select the columns with these names
ds<-dsL[,selectCols] # select all rows for the columns listed selectCols
print(ds[ds$id==1,]) # print all available data for respondent with ID 1

```

	year	id	byear	attend	attendF
1	1997	1	1981	NA	<NA>
2	1998	1	1981	NA	<NA>
3	1999	1	1981	NA	<NA>
4	2000	1	1981	1	Never
5	2001	1	1981	6	About once/week
6	2002	1	1981	2	Once or Twice
7	2003	1	1981	1	Never
8	2004	1	1981	1	Never
9	2005	1	1981	1	Never
10	2006	1	1981	1	Never
11	2007	1	1981	1	Never
12	2008	1	1981	1	Never
13	2009	1	1981	1	Never
14	2010	1	1981	1	Never
15	2011	1	1981	1	Never

Having quick access to factor labels will be especially useful during graph production. For the grammar rules of operations with relevant data see [Data Manipulation Guide](#).

Time metrics : Age, Period, Cohort

NLSY97 sample includes individuals from five cohorts, born between 1980 and 1984. The following graphics shows how birth cohort, age of respondents, and round of observation are related in NSLY97.

Wide age	Age in years																			
	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Born in 1980					1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	
1981				1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011		
1982			1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011			
1983		1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011				
1984	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011					
	Wave																			

Wide wave	Waves of measurement														
	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Born in 1980	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1981	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1982	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1983	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1984	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	Age														

Long wave	Wave	Born in				
		1980	1981	1982	1983	1984
	1997	17	16	15	14	13
	1998	18	17	16	15	14
	1999	19	18	17	16	15
	2000	20	19	18	17	16
	2001	21	20	19	18	17
	2002	22	21	20	19	18
	2003	23	22	21	20	19
	2004	24	23	22	21	20
	2005	25	24	23	22	21
	2006	26	25	24	23	22
	2007	27	26	25	24	23
	2008	28	27	26	25	24
	2009	29	28	27	26	25
	2010	30	29	28	27	26
	2011	31	30	29	28	27
		Age				

Long age	Age years	Born in				
		1980	1981	1982	1983	1984
	13					1997
	14				1997	1998
	15			1997	1998	1999
	16		1997	1998	1999	2000
	17	1997	1998	1999	2000	2001
	18	1998	1999	2000	2001	2002
	19	1999	2000	2001	2002	2003
	20	2000	2001	2002	2003	2004
	21	2001	2002	2003	2004	2005
	22	2002	2003	2004	2005	2006
	23	2003	2004	2005	2006	2007
	24	2004	2005	2006	2007	2008
	25	2005	2006	2007	2008	2009
	26	2006	2007	2008	2009	2010
	27	2007	2008	2009	2010	2011
	28	2008	2009	2010	2011	
	29	2009	2010	2011		
	30	2010	2011			
	31	2011				
		Wave				

Figure 3.1 Overview of the age-period-cohort structure of NLSY97

There are several indicators of age in NLSY97 that vary in precision. Birth cohort (**byear**) is the most general one, it was recorded once. Two age variables were recorded at each interview: age at the time of the interview in months (**agemon**) and in years (**ageyear**). Those are not derivatives of each other, but are closely related. The variable **ageyear** records the full number of years a respondent reached at the time of the interview. Due to difficulties of administering the survey, time intervals between the waves could differ. For example, for one person **id** = 25 the age was recorded as 21 years for both 2003 and 2004 (see **ageyear**). However, when you examine age in months (**agemon**) you can see this rounding issue disappears once a more

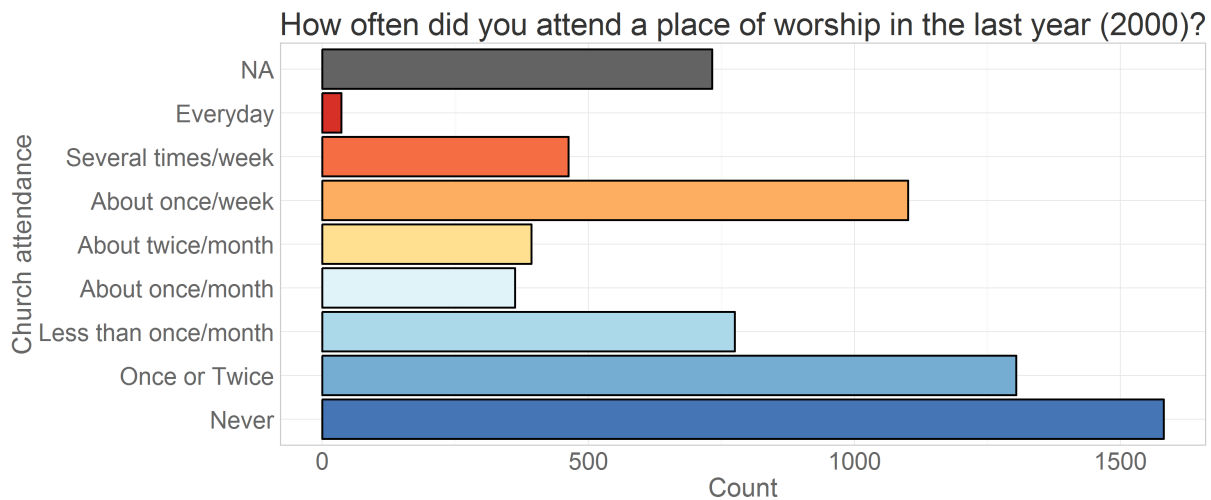
precise scale is used. To avoid this potentially confusing peculiarity, age in years will be calculated as (**age** = **year** - **byear**) or as (**ageALT** = **agemon**/12).

```
ds<-dsL[dsL$year %in% c(2000:2011),c('id',"byear","year","attend","ageyear","agemon")]
ds<- ds[ds$id %in% c(25),]
ds$age<-ds$year-ds$byear
ds$ageALT<- ds$agemon/12
print(ds)
```

	id	byear	year	attend	ageyear	agemon	age	ageALT
364	25	1983	2000	5	17	214	17	17.83
365	25	1983	2001	7	18	226	18	18.83
366	25	1983	2002	7	19	236	19	19.67
367	25	1983	2003	2	21	254	20	21.17
368	25	1983	2004	7	21	261	21	21.75
369	25	1983	2005	5	22	272	22	22.67
370	25	1983	2006	7	23	284	23	23.67
371	25	1983	2007	5	24	295	24	24.58
372	25	1983	2008	7	25	307	25	25.58
373	25	1983	2009	7	26	319	26	26.58
374	25	1983	2010	7	27	332	27	27.67
375	25	1983	2011	7	28	342	28	28.50

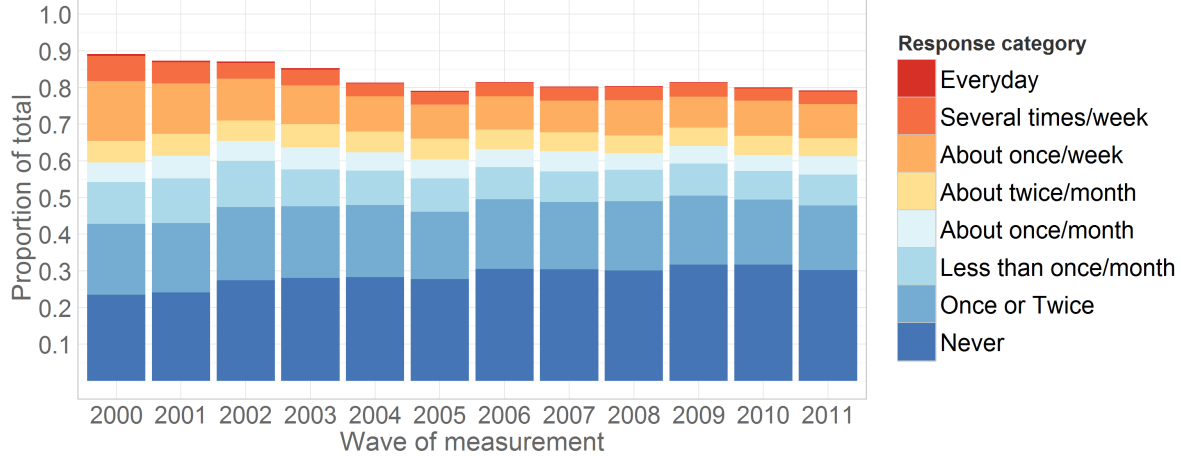
Mapping Church Attendance

The focal variable of interest is **attend**, an item measuring church attendance in the current year. The questionnaire recorded the responses on the ordinal scale.



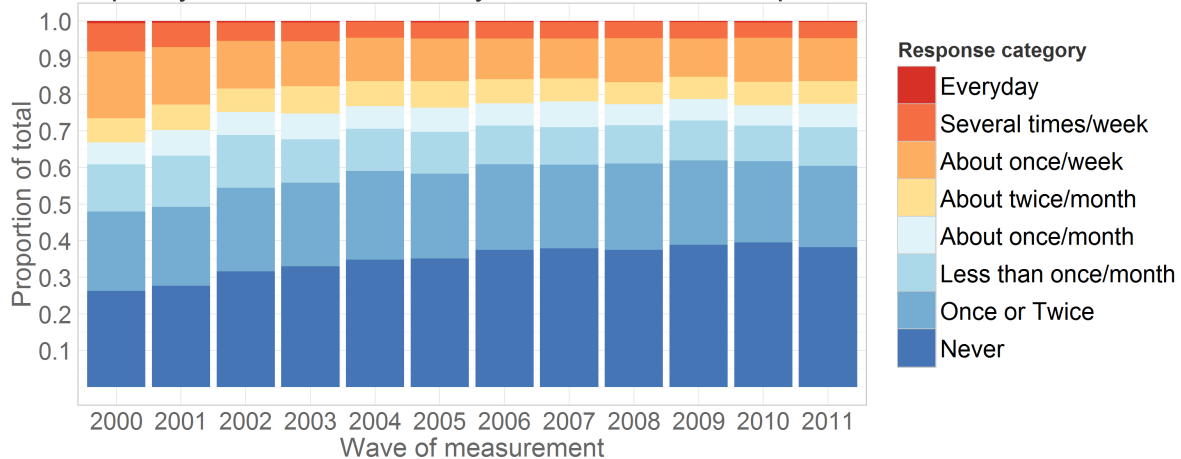
Creating frequency distributions for each of the measurement wave we have:

In the past year, how often have you attended a worship service?



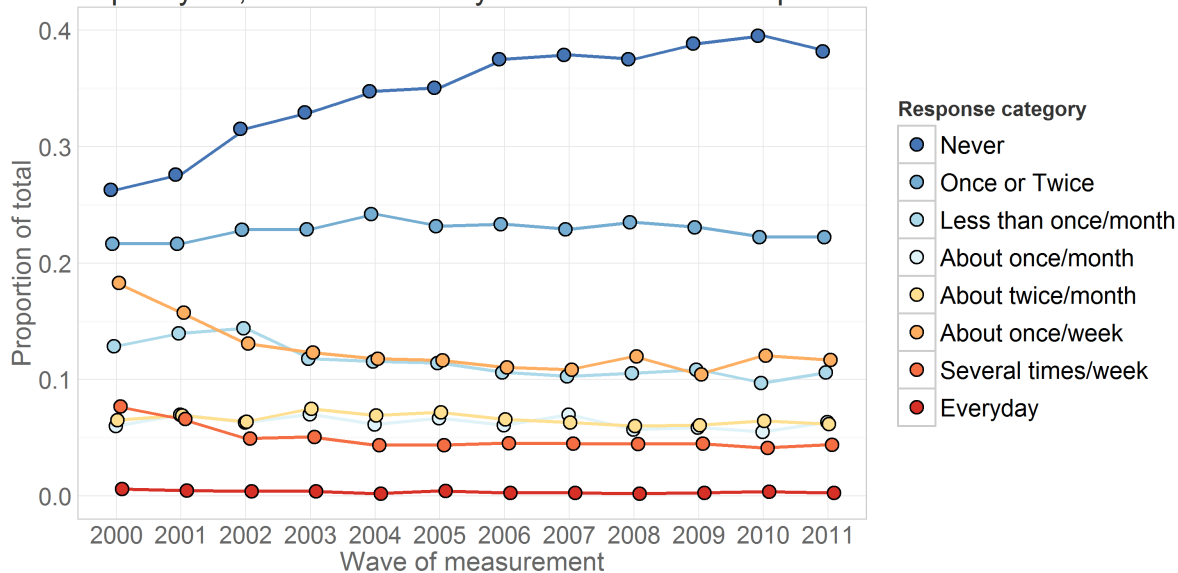
Missing values are used in the calculation of total responses to show the natural attrition in the study. Assuming that attrition is not significantly associated with the outcome measure, we can remove missing values from the calculation of the total and look at prevalence of endorsements over time.

In the past year, how often have you attended a worship service?



Tracing the rate of change of prevalence in a line graph, we see more clearly which categories increase over time (e.g. "Never"), which decline (e.g. "About once/week"), and which stay relatively stable (e.g. "About twice/month")

In the past year, how often have you attended a worship service?



Graphs above shows change in the cross-sectional distribution of responses over the years. Modeling the change in these response frequencies is handled well by Markov models. LCM, however, works with longitudinal data, modeling the trajectory of each individual and treating attendance as a continuous outcome.

To demonstrate mapping of individual trajectories to time, let's select a dataset that would include personal identifier (**id**), cohort indicator (**byear**), wave of measurement (**year**) and the focal variable of interest - worship attendance (**attend**).

```
ds<- dsL %>% dplyr::filter(year %in% c(2000:2011), id==47) %>%
  dplyr::select(id, byear, year, attend, attendF)
print(ds)
```

	id	byear	year	attend	attendF
1	47	1982	2000	5	About twice/month
2	47	1982	2001	2	Once or Twice
3	47	1982	2002	4	About once/month
4	47	1982	2003	2	Once or Twice
5	47	1982	2004	3	Less than once/month
6	47	1982	2005	2	Once or Twice
7	47	1982	2006	2	Once or Twice
8	47	1982	2007	3	Less than once/month
9	47	1982	2008	2	Once or Twice
10	47	1982	2009	1	Never
11	47	1982	2010	1	Never
12	47	1982	2011	1	Never

The view above lists attendance data for subjust with id = 47. Mapping his attendance to time we have



where vertical dimension maps the outcome value and the horizontal maps the time. There will be a trajectory for each of the

```
length(unique(dsL$id))
```

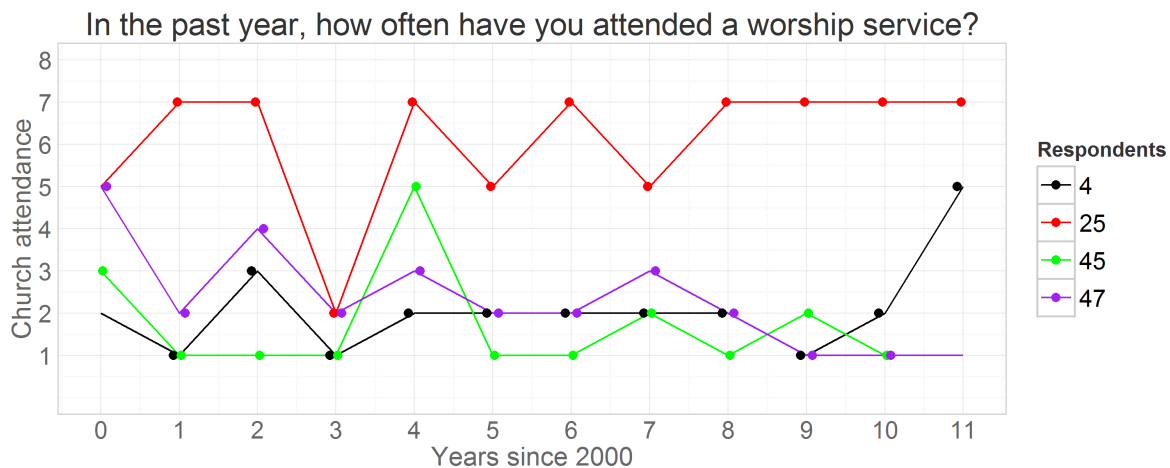
```
[1] 8984
```

subjects in total. Unless specified otherwise, only individuals from the cross-sample will be used in the model to increase external validity.

```
ds<- dsL %>% dplyr::filter(sample==1)
```

Each of such trajectories imply a story, a life scenario. Why one person grows in his religious involvement, while other declines, or never develops an interest in the first place? To demonstrate how interpretations of trajectories can vary among individuals consider the following scenario.

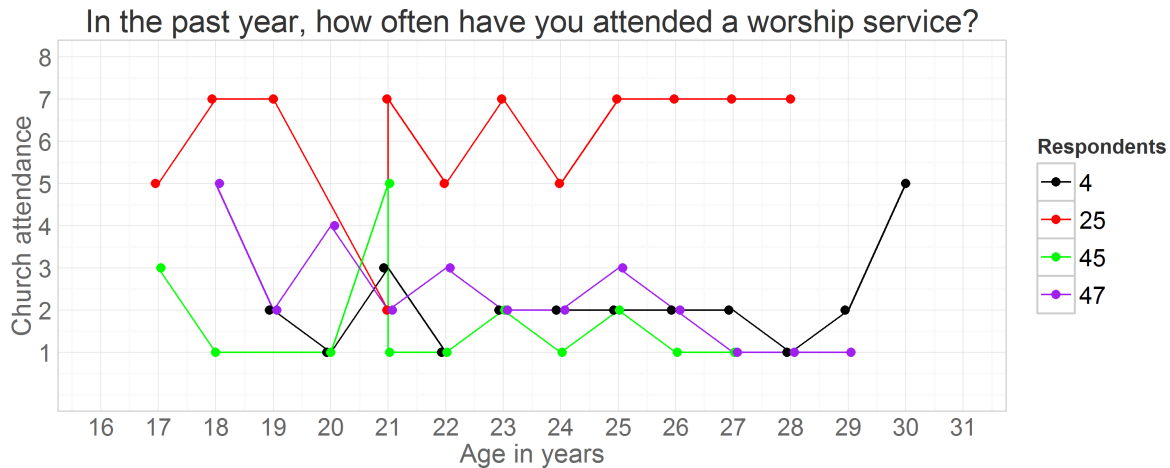
Attendance trajectories of subjects with **ids** 4, 25, 35, and 47 are plotted in the next graph



The respondent **id**=35 reported attending no worship services in any of the years, while respondent **id**=25 seemed to frequent it, indicating weekly attendance in 8 out of the 12 years. Individual **id**=47 started as a fairly regular attendee of religious services in 2000 (5= "about twice a month"), then gradually declined his involvement to nill in 2009 and on. Respondent **id**=4, on the

other hand started off with a rather passive involvement, reporting attended church only “Once or twice” in 2000, maintained a low level of participation throughout the years, only to surge his attendance in 2011. Latent curve models will describe intraindividual trajectories of change, while summarizing the interindividual similarities and trends.

Previous research in religiosity indicated that age might be one of the primary factors explaining interindividual differences in church attendance. To examine the role of age, we change the metric of time from waves of measurement, as in the previous graph, to biological age.

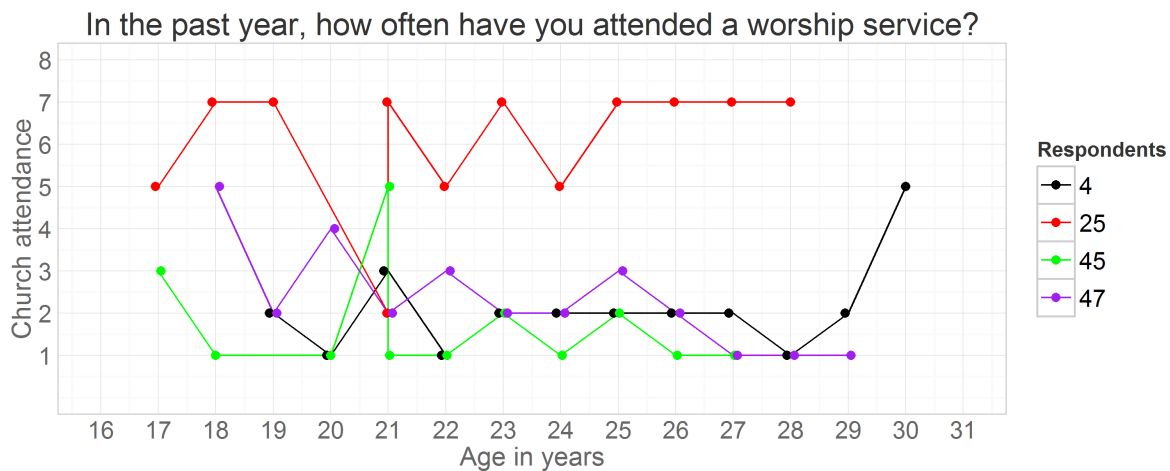


Persons **id** = 35 and **id** = 25 are peers, in 2000 they were both 17. Respondent **id** = 47 is a year older, in 2000 he was 18. The oldest is **id** = 4, who by the last round of measurement in 2011 is 30 years of age. Perhaps, his increased church attendance could be explained by starting a family of his own?

Note that for person **id** = 25 the age was recorded as 21 years for both 2003 and 2004. However, when you examine age in months (**agemon**) you can see this is rounding issue that disappears once a more precise scale is used. To avoid this potentially confusing peculiarity, age in years will be either calculated as (**age** = **year** - **byear**)

```
ds<- dsL %>% dplyr::filter(id %in% c(4,25,45,47),year %in% c(2000:2011)) %>%
  dplyr::select(idF,year,attend,agemon,ageyear) %>%
  mutate(time=year-2000, age=ageyear)
head(ds,12)
```

	idF	year	attend	agemon	ageyear	time	age
1	4	2000	2	238	19	0	19
2	4	2001	1	251	20	1	20
3	4	2002	3	262	21	2	21
4	4	2003	1	276	22	3	22
5	4	2004	2	287	23	4	23
6	4	2005	2	297	24	5	24
7	4	2006	2	309	25	6	25
8	4	2007	2	320	26	7	26
9	4	2008	2	336	27	8	27
10	4	2009	1	344	28	9	28
11	4	2010	2	357	29	10	29
12	4	2011	5	368	30	11	30



or as (`age = agemon/12`).

```
ds<- dsL %>% dplyr::filter(id %in% c(4,25,45,47),year %in% c(2000:2011)) %>%
  dplyr::select(idF,year,attend,agemon,ageyear,byear) %>%
  mutate(time=year-2000, age=agemon/12)
head(ds,12)
```

	idF	year	attend	agemon	ageyear	byear	time	age
1	4	2000	2	238	19	1981	0	19.83
2	4	2001	1	251	20	1981	1	20.92
3	4	2002	3	262	21	1981	2	21.83
4	4	2003	1	276	22	1981	3	23.00
5	4	2004	2	287	23	1981	4	23.92
6	4	2005	2	297	24	1981	5	24.75
7	4	2006	2	309	25	1981	6	25.75
8	4	2007	2	320	26	1981	7	26.67
9	4	2008	2	336	27	1981	8	28.00
10	4	2009	1	344	28	1981	9	28.67
11	4	2010	2	357	29	1981	10	29.75
12	4	2011	5	368	30	1981	11	30.67

