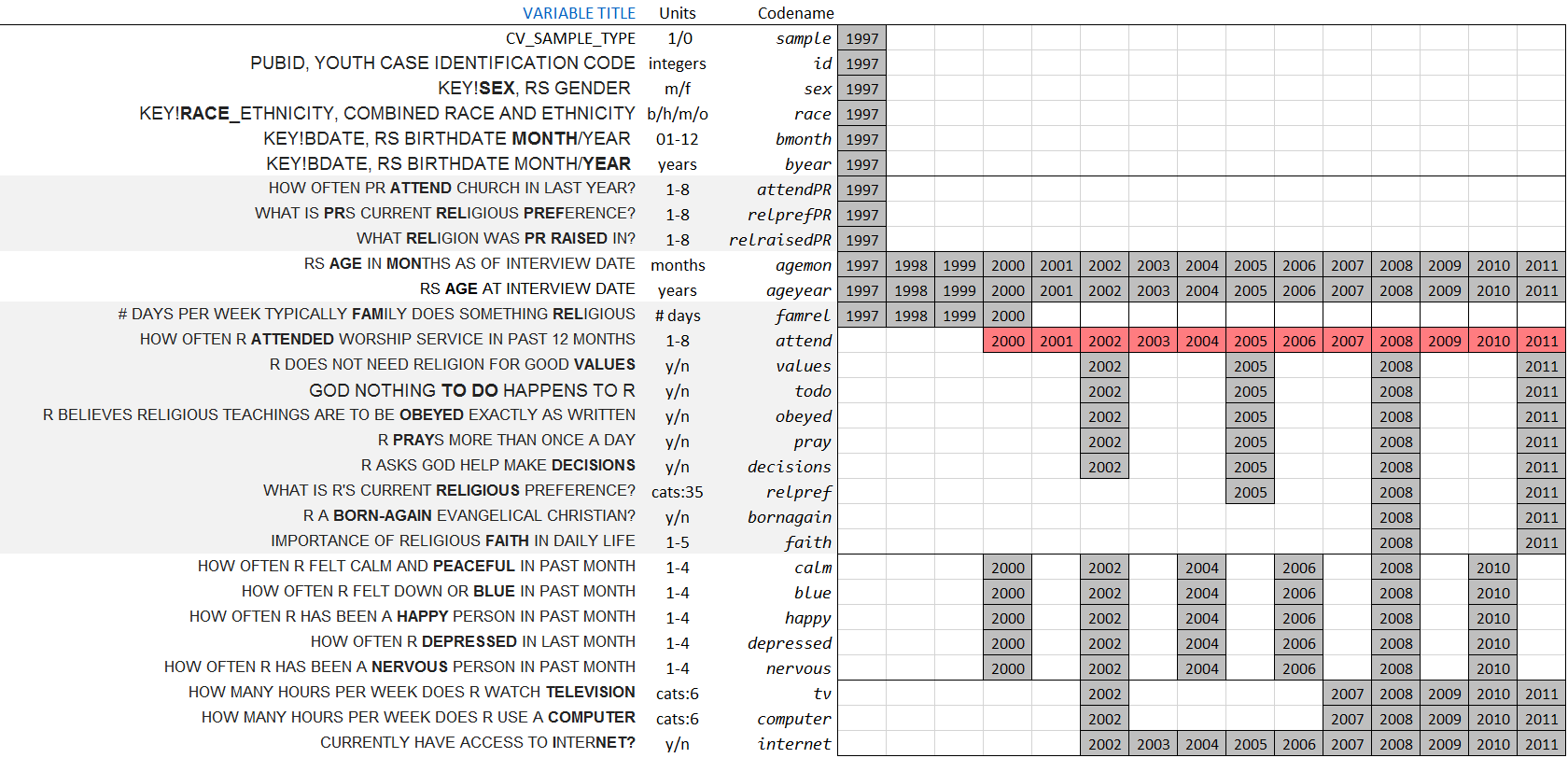
Metrics

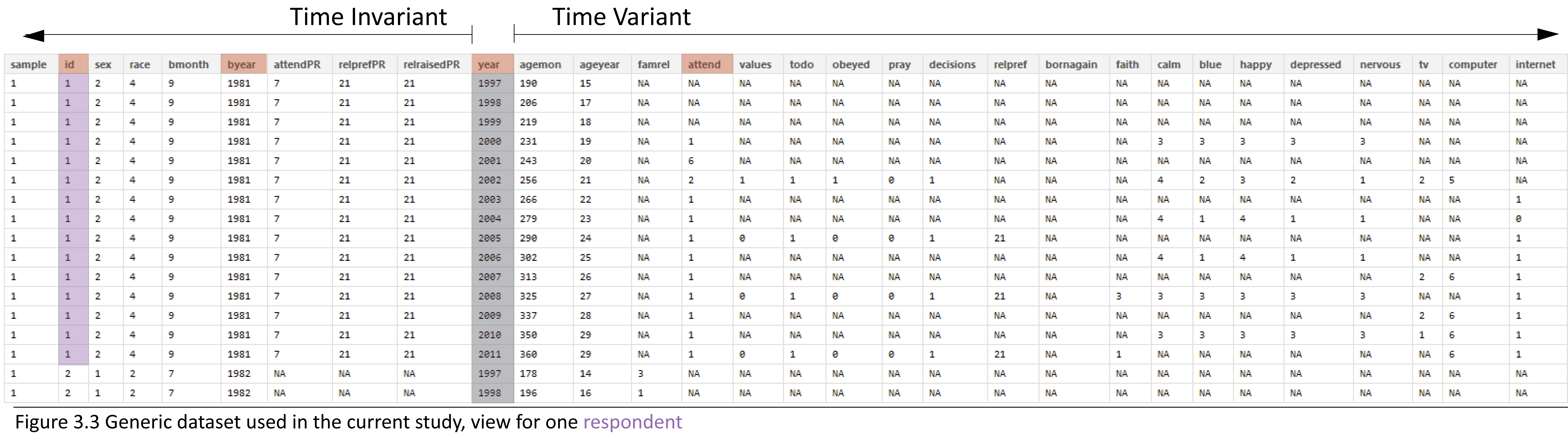
===

Labeling factors and exploring scales.

## Data preliminaries

Initial point of departure - the [databox][databox] of the selected variables, described in the Methods chapter.  
 This [databox][databoxStatcanvas] corresponds to the dataset **dsL** produced by [Derive\_dsL\_from\_Extract][derive] report, given in the Appendix.

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



Note that the variable **year** serves as a natural devided between time invariant (TIvars) and time variant (TVvars) variables. All modeling operations beging with subsetting this dataset. For the grammer rules of operations with relevant data see [Data Manipulation Guide][manipulate].

## Labeling Factor Levels

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

ds<- dsL %>%  
 dplyr::select(  
 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)   
str(ds)

'data.frame': 134745 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][labels] sourced at the end of [Derive\_dsL\_from\_Extract][derive] matches numeric values with response labels from the questionnaire and adds copies of the variables, saved as labeled factors, to **dsL**. For estimations routines such as lme4 or graphing functions such as ggplot2, 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': 134745 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 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 ...  
 $ sampleF : Ord.factor w/ 2 levels "Cross-Sectional"<..: 1 1 1 1 1 1 1 1 1 1 ...  
 $ idF : Factor w/ 8983 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"<..: 4 4 4 4 4 4 4 4 4 4 ...  
 $ 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 NA NA ...  
 $ attendF : Ord.factor w/ 8 levels "Never"<"Once or Twice"<..: NA NA NA 1 6 2 1 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 NA ...  
 $ faithF : Ord.factor w/ 5 levels "Exrtemely"<"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 ...

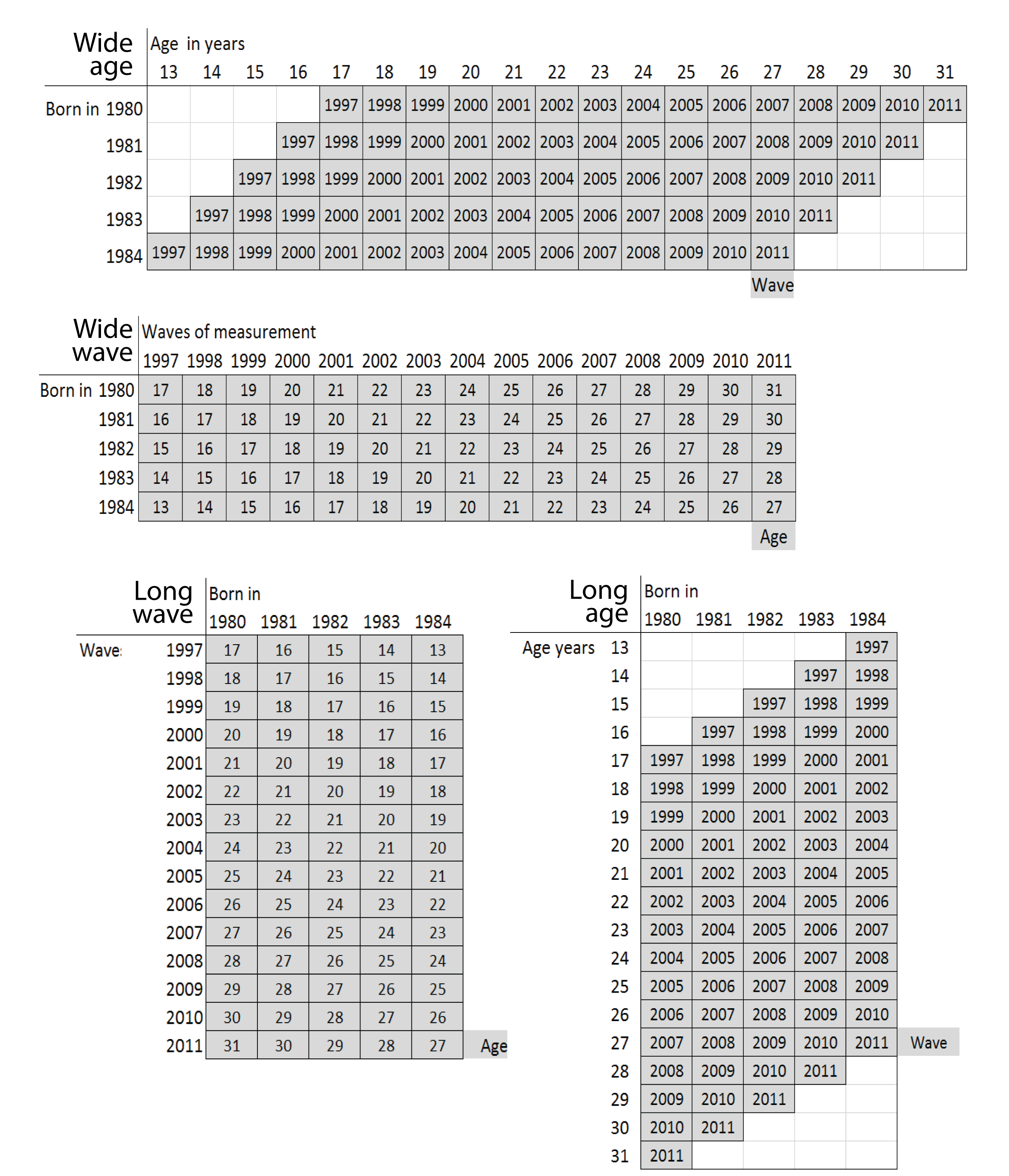
This give a certain flexibity in assembling needed dataset quickly and have access to factor labels. One can alternate between the raw metric and labeled factor by adding "F" suffix to the end of the variable name:

ds<- dsL %>%  
 dplyr::filter(id==25) %>%  
 dplyr::select(id,byear,year, attend,attendF)  
ds

id byear year attend attendF  
1 25 1983 1997 NA <NA>  
2 25 1983 1998 NA <NA>  
3 25 1983 1999 NA <NA>  
4 25 1983 2000 5 About twice/month  
5 25 1983 2001 7 Several times/week  
6 25 1983 2002 7 Several times/week  
7 25 1983 2003 2 Once or Twice  
8 25 1983 2004 7 Several times/week  
9 25 1983 2005 5 About twice/month  
10 25 1983 2006 7 Several times/week  
11 25 1983 2007 5 About twice/month  
12 25 1983 2008 7 Several times/week  
13 25 1983 2009 7 Several times/week  
14 25 1983 2010 7 Several times/week  
15 25 1983 2011 7 Several times/week

Having quick access to factor labels will be especially useful during graph production.

## 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.  


NSLY97 contains static (**bmonth**, **byear**) and dynamic (**agemon**, **ageyear**) indicators of age :

ds<- dsL %>%   
 dplyr::filter(id==25, year %in% c(1997:2011)) %>%   
 dplyr::select(id,byear,bmonthF,year,agemon,ageyear)  
print(ds)

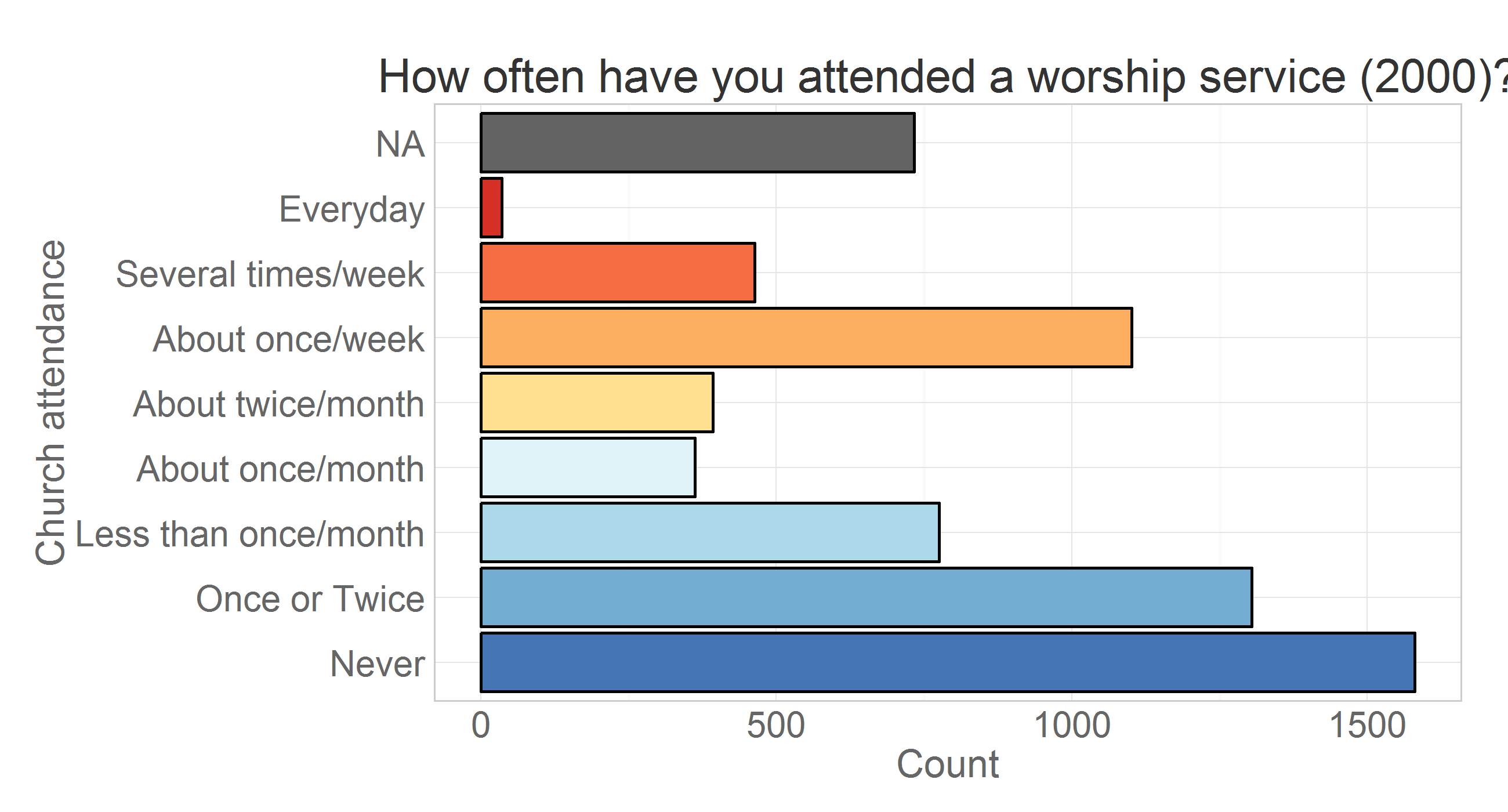
id byear bmonthF year agemon ageyear  
1 25 1983 Mar 1997 167 13  
2 25 1983 Mar 1998 188 15  
3 25 1983 Mar 1999 201 16  
4 25 1983 Mar 2000 214 17  
5 25 1983 Mar 2001 226 18  
6 25 1983 Mar 2002 236 19  
7 25 1983 Mar 2003 254 21  
8 25 1983 Mar 2004 261 21  
9 25 1983 Mar 2005 272 22  
10 25 1983 Mar 2006 284 23  
11 25 1983 Mar 2007 295 24  
12 25 1983 Mar 2008 307 25  
13 25 1983 Mar 2009 319 26  
14 25 1983 Mar 2010 332 27  
15 25 1983 Mar 2011 342 28

Variable **year** is used as cohort indicator. Variable **year** enumerates NLSY97 rounds, recording the calendaric year during which it took place. When transforming the metric of time, and using biological age instead of **year** as the temporal dimension, the value of age at the time of the interview will be computed as **age** = **agemon**/12

ds<- dsL %>%   
 dplyr::filter(id==25, year %in% c(1997:2011)) %>%   
 dplyr::select(id,bmonthF,byear,year, agemon,ageyear) %>%  
 dplyr::mutate (age = agemon/12)  
print(ds)

id bmonthF byear year agemon ageyear age  
1 25 Mar 1983 1997 167 13 13.92  
2 25 Mar 1983 1998 188 15 15.67  
3 25 Mar 1983 1999 201 16 16.75  
4 25 Mar 1983 2000 214 17 17.83  
5 25 Mar 1983 2001 226 18 18.83  
6 25 Mar 1983 2002 236 19 19.67  
7 25 Mar 1983 2003 254 21 21.17  
8 25 Mar 1983 2004 261 21 21.75  
9 25 Mar 1983 2005 272 22 22.67  
10 25 Mar 1983 2006 284 23 23.67  
11 25 Mar 1983 2007 295 24 24.58  
12 25 Mar 1983 2008 307 25 25.58  
13 25 Mar 1983 2009 319 26 26.58  
14 25 Mar 1983 2010 332 27 27.67  
15 25 Mar 1983 2011 342 28 28.50

## Attendance

NLSY97 asked to report church attendance (**attend**) for the past 12 months preceding the interview date. The response card offered a choice of 7 categories ordered by magnitude. 

## Read more

in ./Models/Descriptives:

* [Metrics][metrics] - how values of items are labeled
* [Descriptives][descriptives] - basic stats of various items (**Continue**)
* [Attendance][attend] - focus on church attendence over time
* [Databox][databox]

See also

* [Deriving Data from NLYS97 extract][derive]
* [Data Manipulation Guide][manipulate]