Using parametric splines and Fourier series for seasonal effects

2024-03-15

- Using parametric splines for piece-wise polynomial curves
- and Fourier series for seasonal effects

Data set simulates data from Statistics Canada NPHS from 1994 to 2011. Participants were surveyed every 2 years for up to 7 occasions.

Some participants happened to give birth during the study but since data was collected every two years there was little data on individual longitudinal sleep patterns before and after birth.

However, using mixed models with a parametric model for sleep behaviour before and after birth, it's possible to 'stitch' trajectories together to get a picture of individual predicted sleep trajectories.

```
library(spida2)
library(nlme)
```

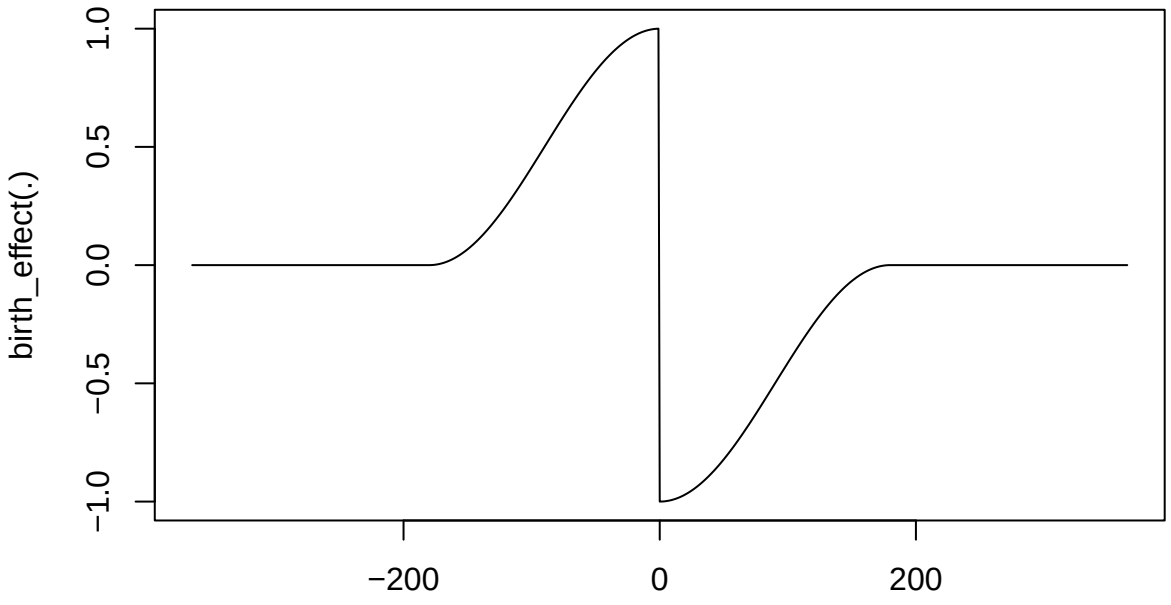
```
    Attaching package: 'nlme'

    The following object is masked from 'package:spida2':

        getData
```

```
library(latticeExtra)
```

```
    Loading required package: lattice
```

Hypothetical perinatal 'birth effect' on maternal sleep relative to days before and after birth

```
birth_effect <- function(d, plus = 1, minus = 1) {
  ifelse(d < -180, 0,
         ifelse(d < 0, plus * (.5 -.5* cos(pi*(d+180)/180)),
                ifelse(d < 180, - minus * (.5 -.5* cos(pi*(d+180)/180)), 0)
         )
  )
}
# test
seq(-365,365) %>% plot(., birth_effect(.), type = 'l')
```



.

Generate a data set

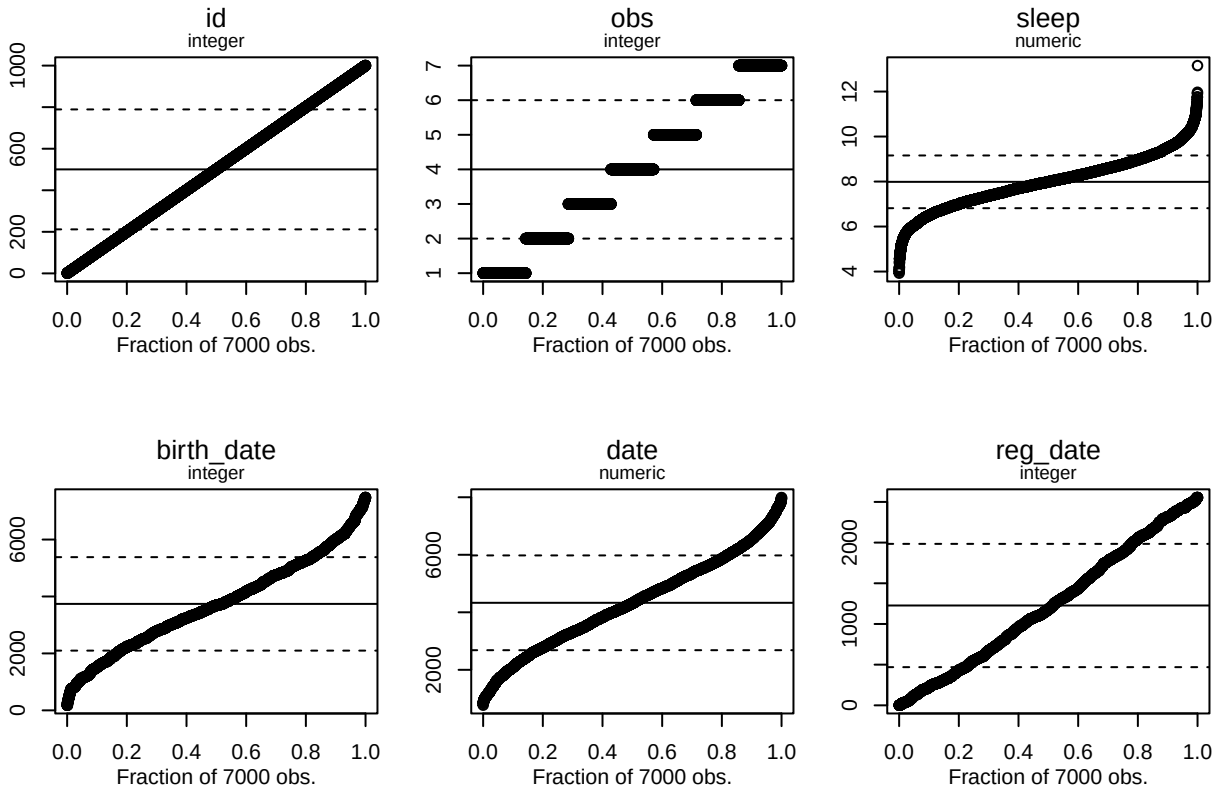
Note that many women in the NPHS gave birth more than once. Here there is only one birth recorded per person.

```
# sample(100000, 1)
{
  set.seed(4728)
  Nid <- 1000    # number of subjects
  Nobs <- 7      # observartions per subject

  expand.grid(id = 1:Nid, obs = 1:Nobs) %>%  # basic skeleton for data set
    within(
      {

        # date id registered
        reg_date <- sample(Nobs * 365, Nid, replace = TRUE)[id]  # generating one value per id

        # dates id observed (approx every 2 years)
        date <- reg_date + obs*2*365 + sample(365, length(id), replace = TRUE)  # generating one value

        birth_date <- reg_date + sample(365*14, Nid, replace = TRUE)[id]       # date giving birth

        ..plus <- runif(Nid)[id]       # extra sleep pre birth
        ..minus <- runif(Nid)[id]      # less sleep after birth
        ..birth_effect <- birth_effect( date - birth_date, ..plus, ..minus)

        ..seasonal <- .5 * cos(2*pi*(date-30)/365)
        ..sd_between <- 1
        ..sd_within <- .5

        sleep <- 8 + ..sd_between * rnorm(Nid)[id] + ..sd_within * rnorm(id) +
          ..birth_effect + ..seasonal
```

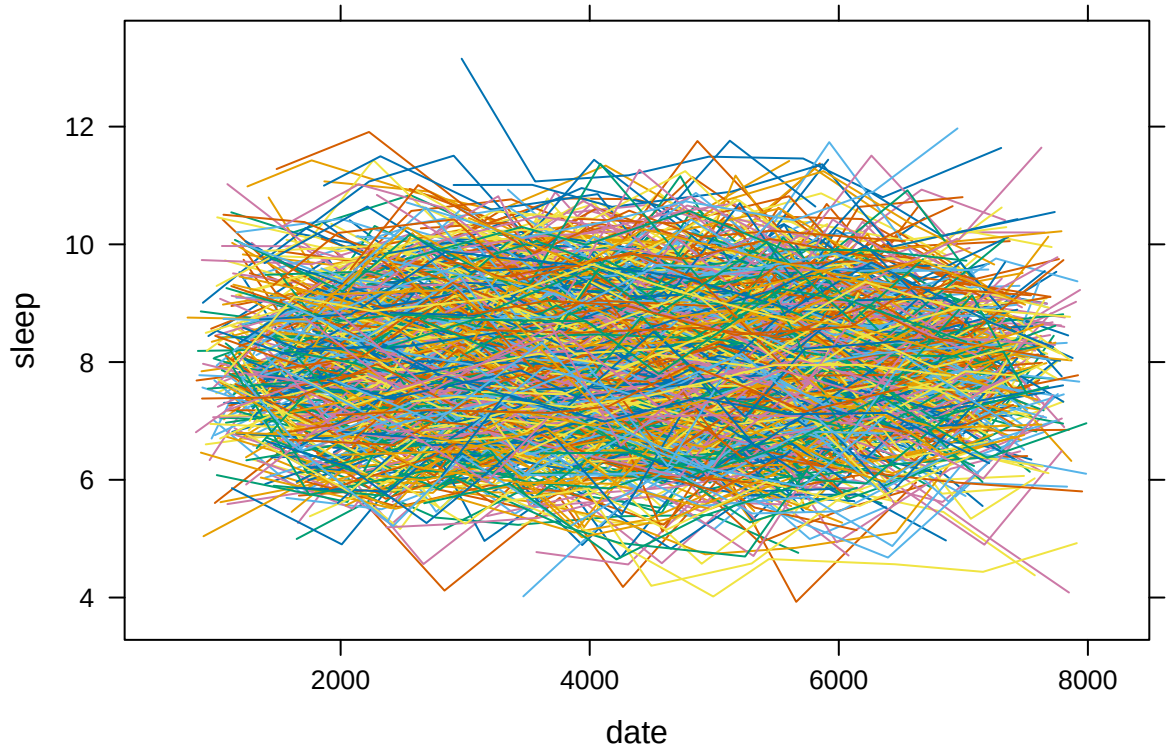
```
        ..plus <- ..minus <- ..birth_effect <- ..seasonal <- ..sd_between <- ..sd_within <- NULL

      }
    ) %>%
    sortdf(~id/date)-> dd
}
head(dd)
```

```
     id obs    sleep birth_date date reg_date
1     1   1 7.088871       5298 1197      288
1001  1   2 7.336316       5298 1891      288
2001  1   3 7.909879       5298 2655      288
3001  1   4 7.489981       5298 3441      288
4001  1   5 7.087731       5298 4084      288
5001  1   6 6.920117       5298 4893      288
```

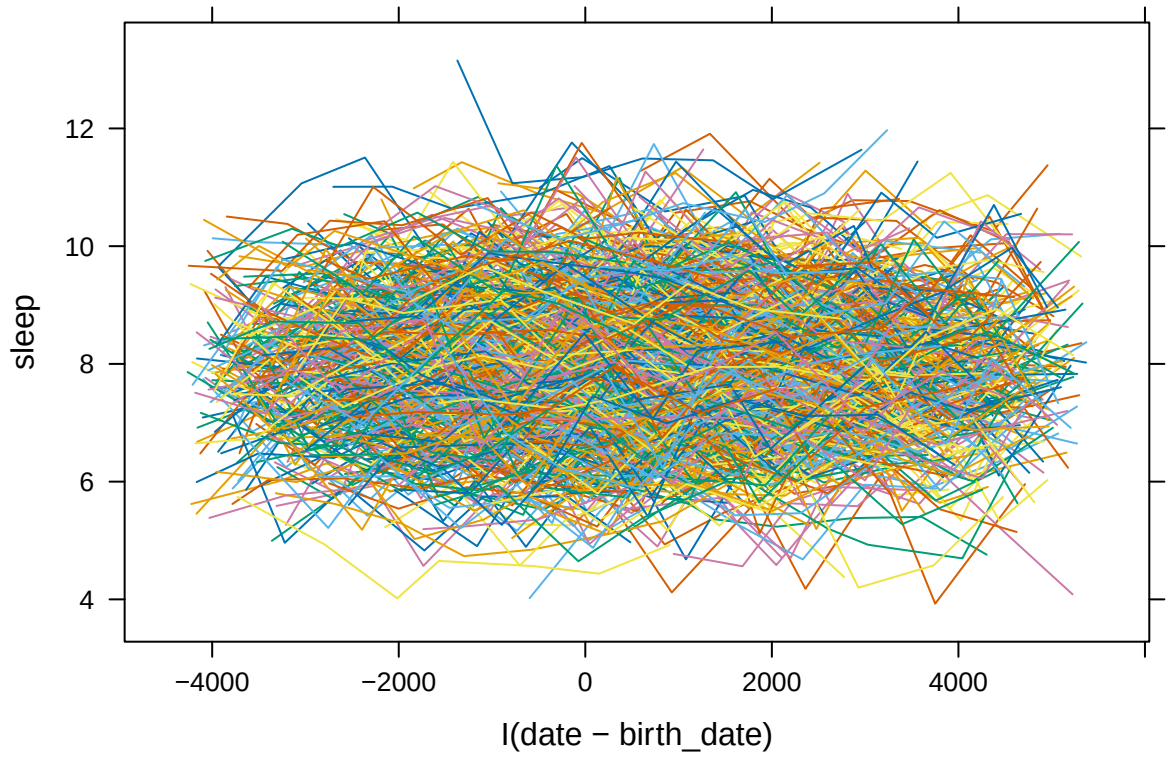
```
xqplot(dd)
```




```
xyplot(sleep ~ date, dd, groups = id, type = 'l')
```



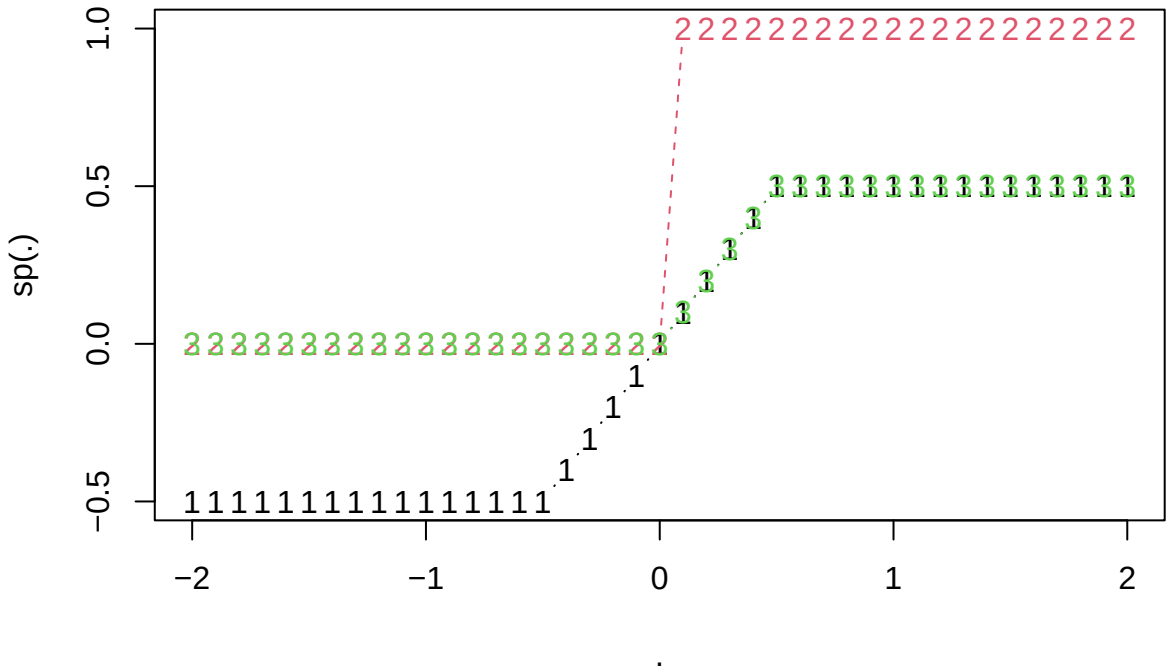

```
xyplot(sleep ~ I(date-birth_date), dd, groups = id, type = 'l')
```




```
# Note: one observation every two years on each person

fit <- lme(sleep ~ 1, dd, random = ~1 |id)
summary(fit)
```

```
Linear mixed-effects model fit by REML
  Data: dd
       AIC      BIC    logLik
  16048.53 16069.09 -8021.267

Random effects:
 Formula: ~1 | id
        (Intercept)  Residual
StdDev:   0.9974714 0.6156731

Fixed effects:  sleep ~ 1
               Value  Std.Error   DF  t-value p-value
(Intercept) 7.987087 0.03238981 6000 246.5926       0

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.85320384 -0.64160961  0.00633589  0.63204103  3.68567783

Number of Observations: 7000
Number of Groups: 1000
```

```
# define a parametric spline using years as unit to avoid large numbers

sp <- function(y) {
  gsp(y, knots = c(-.5,0,.5), degree = c(0,1,1,0), c(0, -1, 0))
}

seq(-2,2,.1) %>% matplot(., sp(.), type ='b')
```




```
sp(seq(-2,2,.1))
```

```
          D1(0) C(0).0 C(0).1
f(-2)    -0.5      0    0.0
f(-1.9)  -0.5      0    0.0
```

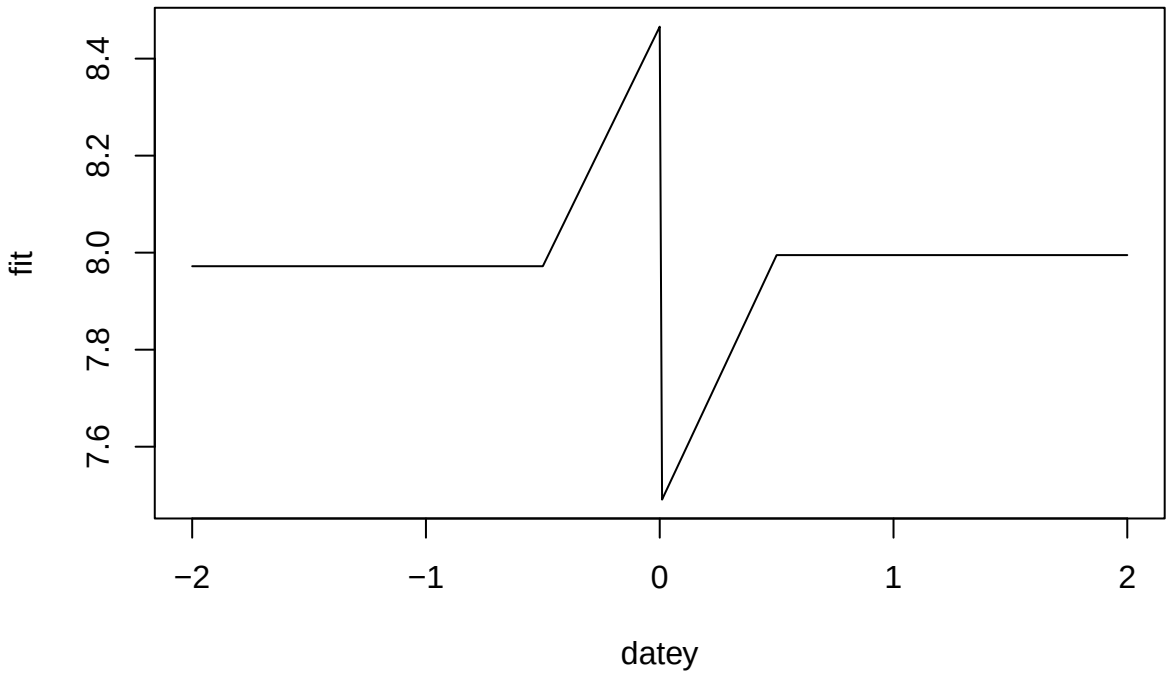
```
f(-1.8)  -0.5      0    0.0
f(-1.7)  -0.5      0    0.0
f(-1.6)  -0.5      0    0.0
f(-1.5)  -0.5      0    0.0
f(-1.4)  -0.5      0    0.0
f(-1.3)  -0.5      0    0.0
f(-1.2)  -0.5      0    0.0
f(-1.1)  -0.5      0    0.0
f(-1)    -0.5      0    0.0
f(-0.9)  -0.5      0    0.0
f(-0.8)  -0.5      0    0.0
f(-0.7)  -0.5      0    0.0
f(-0.6)  -0.5      0    0.0
f(-0.5)  -0.5      0    0.0
f(-0.4)  -0.4      0    0.0
f(-0.3)  -0.3      0    0.0
f(-0.2)  -0.2      0    0.0
f(-0.1)  -0.1      0    0.0
f(0)      0.0      0    0.0
f(0.1)    0.1      1    0.1
f(0.2)    0.2      1    0.2
f(0.3)    0.3      1    0.3
f(0.4)    0.4      1    0.4
f(0.5)    0.5      1    0.5
f(0.6)    0.5      1    0.5
f(0.7)    0.5      1    0.5
f(0.8)    0.5      1    0.5
f(0.9)    0.5      1    0.5
f(1)      0.5      1    0.5
f(1.1)    0.5      1    0.5
f(1.2)    0.5      1    0.5
f(1.3)    0.5      1    0.5
f(1.4)    0.5      1    0.5
f(1.5)    0.5      1    0.5
f(1.6)    0.5      1    0.5
f(1.7)    0.5      1    0.5
f(1.8)    0.5      1    0.5
f(1.9)    0.5      1    0.5
f(2)      0.5      1    0.5
attr(,"spline.attr")
attr(,"spline.attr")$knots
[1] -0.5  0.0  0.5

attr(,"spline.attr")$degree
[1] 0 1 1 0

attr(,"spline.attr")$smoothness
[1]  0 -1  0

attr(,"spline.attr")$lin
NULL

attr(,"spline.attr")$intercept
[1] 0
```

```
    attr(,"spline.attr")$signif
    [1] 3

    attr(,"class")
    [1] "gsp"
```

```
# use years as time units

dd <- within(dd,
           {
             datey <- date /365
             birthy <- birth_date / 365
           })

fit <- lme(sleep ~ sp(datey - birthy) , dd, random = ~ 1 | id)
summary(fit)
```

```
    Linear mixed-effects model fit by REML
      Data: dd
           AIC      BIC    logLik
      15985.45 16026.57 -7986.724

    Random effects:
     Formula: ~1 | id
            (Intercept)  Residual
    StdDev:   0.9970469 0.6118624

    Fixed effects:  sleep ~ sp(datey - birthy)
                                Value  Std.Error   DF   t-value p-value
    (Intercept)              8.465813 0.08105925 5997 104.43982  0.0000
    sp(datey - birthy)D1(0)   0.987491 0.15450552 5997   6.39130  0.0000
    sp(datey - birthy)C(0).0 -0.985077 0.11077282 5997  -8.89277  0.0000
    sp(datey - birthy)C(0).1  0.040963 0.22128064 5997   0.18512  0.8531
     Correlation:
                             (Intr) s(-b)D s(-b)C(0).0
    sp(datey - birthy)D1(0)   0.907
    sp(datey - birthy)C(0).0 -0.637 -0.682
    sp(datey - birthy)C(0).1 -0.617 -0.677 -0.063

    Standardized Within-Group Residuals:
             Min           Q1          Med           Q3          Max
    -3.877379606 -0.646902798  0.002500854  0.631484893  3.660662143

    Number of Observations: 7000
    Number of Groups: 1000
```

```
pred <- data.frame(datey = seq(-2,2,.01), birthy = 0)
pred$fit <- predict(fit, newdata = pred, level = 0)
with(pred, plot(datey, fit, type = 'l'))
```



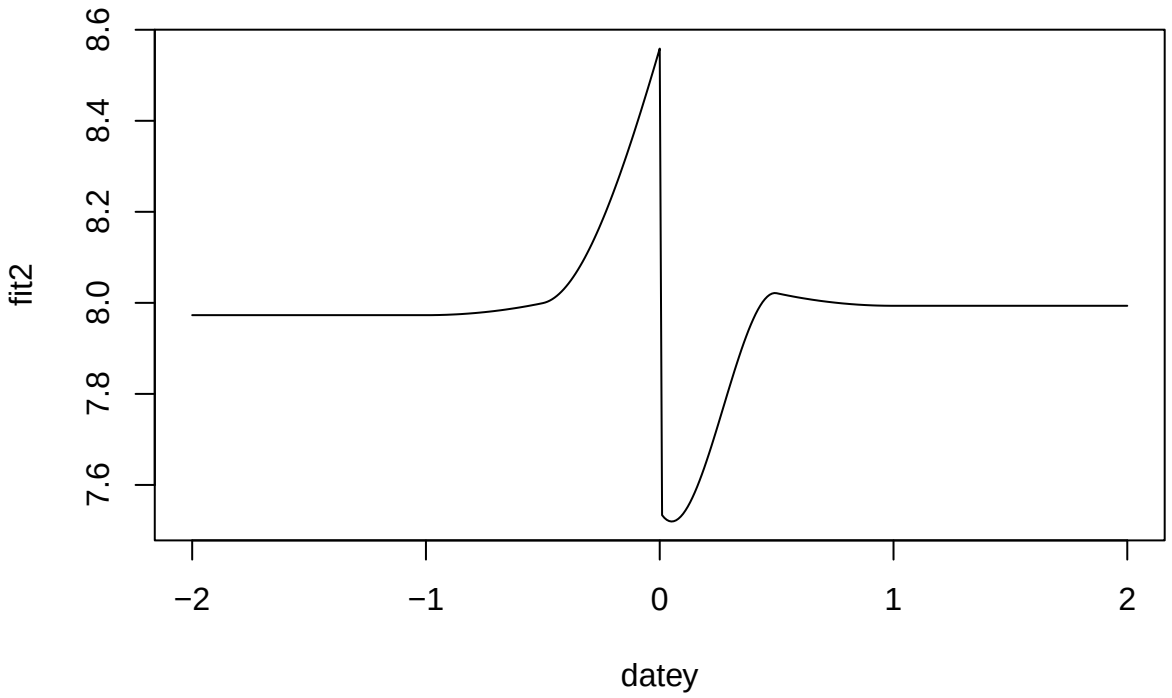
Try a different spline

```
sp2 <- function(y) gsp(y, c(-1,-.5, 0, .5, 1), c(0,2,3,3,2,0), c(1,1,-1,1,1))

fit2 <- lme(sleep ~ sp2(datey - birthy) , dd, random = ~ 1 | id)
summary(fit2)
```

```
Linear mixed-effects model fit by REML
  Data: dd
      AIC      BIC    logLik
  15967.9 16036.42 -7973.948

Random effects:
 Formula: ~1 | id
        (Intercept)  Residual
StdDev:   0.9971542 0.6118803

Fixed effects:  sleep ~ sp2(datey - birthy)
                              Value Std.Error   DF  t-value p-value
(Intercept)                8.55861   0.15489 5993 55.25753  0.0000
sp2(datey - birthy)D1(0)   1.75659   2.02286 5993  0.86837  0.3852
sp2(datey - birthy)D2(0)   1.05499  14.57452 5993  0.07239  0.9423
sp2(datey - birthy)D3(0)  -8.98711  43.99968 5993 -0.20425  0.8382
sp2(datey - birthy)C(0).0 -1.01700   0.21748 5993 -4.67626  0.0000
sp2(datey - birthy)C(0).1 -2.64154   2.94118 5993 -0.89812  0.3692
sp2(datey - birthy)C(0).2 17.97831  21.12807 5993  0.85092  0.3948
sp2(datey - birthy)C(0).3 -60.95171  63.63423 5993 -0.95784  0.3382
 Correlation:
                           (Intr) s2(-b)D1 s2(-b)D2 s2(-b)D3 s2(-b)C(0).0
sp2(datey - birthy)D1(0)    0.840
sp2(datey - birthy)D2(0)    0.726  0.975
sp2(datey - birthy)D3(0)    0.663  0.945    0.994
sp2(datey - birthy)C(0).0 -0.685 -0.599   -0.518   -0.474
```

```
sp2(datey - birthy)C(0).1 -0.577 -0.688   -0.671   -0.650   -0.031
sp2(datey - birthy)C(0).2 -0.502 -0.672   -0.689   -0.685    0.742
sp2(datey - birthy)C(0).3 -0.457 -0.654   -0.688   -0.692   -0.024
                          s2(-b)C(0).1 s2(-b)C(0).2
sp2(datey - birthy)D1(0)
sp2(datey - birthy)D2(0)
sp2(datey - birthy)D3(0)
sp2(datey - birthy)C(0).0
sp2(datey - birthy)C(0).1
sp2(datey - birthy)C(0).2 -0.050
sp2(datey - birthy)C(0).3  0.946       -0.046

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-3.877143934 -0.645090676  0.001593104  0.629206506  3.666263366

Number of Observations: 7000
Number of Groups: 1000
```

```
pred$fit2 <- predict(fit2, newdata = pred, level = 0)
with(pred, plot(datey, fit2, type = 'l'))
```



fit and fit2 have different FE models so we must refit with ML to do LRT test

```
anova(update(fit, method = "ML"), update(fit2, method = "ML"))
```

```
                           Model df      AIC      BIC    logLik   Test
update(fit, method = "ML")     1  6 15970.56 16011.68 -7979.281
update(fit2, method = "ML")    2 10 15975.12 16043.65 -7977.558 1 vs 2
                            L.Ratio p-value
update(fit, method = "ML")
update(fit2, method = "ML") 3.446274  0.4861
```

Adding seasonal effects with sin/cos pair harmonics

```
Sin <- function(x) cbind(sin(x), cos(x))
#
fit3 <- lme(sleep ~ sp2(datey - birthy) + Sin(2*pi*datey) , dd, random = ~ 1 | id)
```

```
    Error in lme.formula(sleep ~ sp2(datey - birthy) + Sin(2 * pi * datey), : nlminb problem, convergen
      message = false convergence (8)
```

```
summary(fit3)
```

```
    Error in eval(expr, envir, enclos): object 'fit3' not found
```

```
preds <- data.frame(datey = seq(0,2,.01))
preds$birthy <- preds$datey - 2

preds$fit3 <- predict(fit3, newdata = preds, level = 0)
```

```
    Error in eval(expr, envir, enclos): object 'fit3' not found
```

```
with(preds, plot(datey, fit3, type = 'l'))
```

```
    Error in eval(substitute(expr), data, enclos = parent.frame()): object 'fit3' not found
```

Fitting higher harmonics

```
fit4 <- lme(sleep ~ sp2(datey - birthy) + Sin(pi * 2 * datey) + Sin(2 * pi * 2 * datey) , dd, random =

summary(fit4)
```

```
    Linear mixed-effects model fit by REML
      Data: dd
           AIC      BIC    logLik
      13673.58 13769.51 -6822.792

    Random effects:
     Formula: ~1 | id
            (Intercept)  Residual
    StdDev:   0.9939207 0.5048021

    Fixed effects:  sleep ~ sp2(datey - birthy) + Sin(pi * 2 * datey) + Sin(2 * pi *      2 * datey)
                                  Value Std.Error   DF  t-value p-value
    (Intercept)                8.497029   0.12918 5989 65.77654  0.0000
    sp2(datey - birthy)D1(0)   1.034671   1.67192 5989  0.61885  0.5360
    sp2(datey - birthy)D2(0)  -0.875249  12.04574 5989 -0.07266  0.9421
    sp2(datey - birthy)D3(0)  -9.083108  36.36442 5989 -0.24978  0.8028
    sp2(datey - birthy)C(0).0 -1.060313   0.17974 5989 -5.89907  0.0000
    sp2(datey - birthy)C(0).1 -0.083085   2.43197 5989 -0.03416  0.9727
    sp2(datey - birthy)C(0).2  7.990061  17.47028 5989  0.45735  0.6474
    sp2(datey - birthy)C(0).3 -28.218451  52.61140 5989 -0.53636  0.5917
    Sin(pi * 2 * datey)1       0.254883   0.00925 5989 27.54042  0.0000
    Sin(pi * 2 * datey)2       0.421422   0.00920 5989 45.79406  0.0000
    Sin(2 * pi * 2 * datey)1  -0.010800   0.00931 5989 -1.15948  0.2463
    Sin(2 * pi * 2 * datey)2  -0.001778   0.00913 5989 -0.19475  0.8456
     Correlation:
                              (Intr) s2(-b)D1 s2(-b)D2 s2(-b)D3 s2(-b)C(0).0
    sp2(datey - birthy)D1(0)   0.832
    sp2(datey - birthy)D2(0)   0.719  0.975
    sp2(datey - birthy)D3(0)   0.657  0.945    0.994
```

```
    sp2(datey - birthy)C(0).0 -0.679 -0.599   -0.518   -0.474
    sp2(datey - birthy)C(0).1 -0.571 -0.688   -0.671   -0.650   -0.032
    sp2(datey - birthy)C(0).2 -0.497 -0.672   -0.689   -0.685    0.742
    sp2(datey - birthy)C(0).3 -0.453 -0.653   -0.687   -0.691   -0.025
    Sin(pi * 2 * datey)1      -0.005 -0.006   -0.003   -0.001    0.015
    Sin(pi * 2 * datey)2      -0.007 -0.005   -0.001    0.001   -0.015
    Sin(2 * pi * 2 * datey)1   0.019  0.029    0.030    0.029   -0.007
    Sin(2 * pi * 2 * datey)2   0.011  0.015    0.013    0.012   -0.022
                              s2(-b)C(0).1 s2(-b)C(0).2 s2(-b)C(0).3 S(*2*d)1
    sp2(datey - birthy)D1(0)
    sp2(datey - birthy)D2(0)
    sp2(datey - birthy)D3(0)
    sp2(datey - birthy)C(0).0
    sp2(datey - birthy)C(0).1
    sp2(datey - birthy)C(0).2 -0.051
    sp2(datey - birthy)C(0).3  0.946       -0.047
    Sin(pi * 2 * datey)1      -0.013        0.021       -0.019
    Sin(pi * 2 * datey)2       0.031       -0.026        0.025        0.007
    Sin(2 * pi * 2 * datey)1  -0.029       -0.014       -0.025       -0.003
    Sin(2 * pi * 2 * datey)2   0.005       -0.026        0.011        0.013
                              S(*2*d)2 S(2*p*2*d)1
    sp2(datey - birthy)D1(0)
    sp2(datey - birthy)D2(0)
    sp2(datey - birthy)D3(0)
    sp2(datey - birthy)C(0).0
    sp2(datey - birthy)C(0).1
    sp2(datey - birthy)C(0).2
    sp2(datey - birthy)C(0).3
    Sin(pi * 2 * datey)1
    Sin(pi * 2 * datey)2
    Sin(2 * pi * 2 * datey)1  -0.015
    Sin(2 * pi * 2 * datey)2   0.018   -0.001

    Standardized Within-Group Residuals:
             Min           Q1          Med           Q3          Max
    -3.721992681 -0.629511457 -0.005996216  0.641210512  3.536125756

    Number of Observations: 7000
    Number of Groups: 1000
```

```
wald(fit4, 'Sin\\(2')
```

```
            numDF denDF   F-value p-value
    Sin\\(2     2  5989 0.6914484 0.50089
                            Estimate  Std.Error DF   t-value   p-value Lower 0.95
    Sin(2 * pi * 2 * datey)1 -0.010800 0.009314  5989 -1.159480 0.24631 -0.02906
    Sin(2 * pi * 2 * datey)2 -0.001778 0.009127  5989 -0.194755 0.84559 -0.01967
                            Upper 0.95
    Sin(2 * pi * 2 * datey)1 0.007460
    Sin(2 * pi * 2 * datey)2 0.016115
```

```
wald(fit4, 'Sin')
```

```
        numDF denDF  F-value p-value
    Sin     4  5989 709.7704 <.00001
```

```
                          Estimate  Std.Error DF   t-value   p-value Lower 0.95
    Sin(pi * 2 * datey)1      0.254883 0.009255  5989 27.540420 <.00001  0.236740
    Sin(pi * 2 * datey)2      0.421422 0.009203  5989 45.794064 <.00001  0.403382
    Sin(2 * pi * 2 * datey)1 -0.010800 0.009314  5989 -1.159480 0.24631 -0.029060
    Sin(2 * pi * 2 * datey)2 -0.001778 0.009127  5989 -0.194755 0.84559 -0.019670
                          Upper 0.95
    Sin(pi * 2 * datey)1     0.273025
    Sin(pi * 2 * datey)2     0.439463
    Sin(2 * pi * 2 * datey)1 0.007460
    Sin(2 * pi * 2 * datey)2 0.016115
```

```
preds <- data.frame(datey = seq(0,2,.01))
preds$birthy <- preds$datey - 2

preds$fit3 <- predict(fit3, newdata = preds, level = 0)
```

```
    Error in eval(expr, envir, enclos): object 'fit3' not found
```

```
with(preds, plot(datey, fit3, type = 'l'))
```

```
    Error in eval(substitute(expr), data, enclos = parent.frame()): object 'fit3' not found
```