

R in datumske spremenljivke

A. Blejec

andrej.blejec@nib.si

9. februar 2008

1 Datum kot spremenljivka razreda Date

1.1 Pretvarjanje časovnih podatkov

Datumi, ki so organizirani kot `class: Date`, R obravnava kot število dni od 1. Januarja 1970:¹

```
> str(as.Date("1970-1-1"))
Class 'Date' num 0
> str(as.Date("1971-1-1"))
Class 'Date' num 365
```

Ker je vrednosti pripisan `Class: Date`, ta števila izpisuje formatirano kot datume v obliki "yyyy-mm-dd". Obliko izpisa lahko s `format` spremenimo v bolj domač "dd.mm.yyyy", "dd.mm.yy" in še kakšen format.

```
> x <- 364
> x
[1] 364
> class(x) <- "Date"
> str(x)
Class 'Date' num 364
> x
[1] "1970-12-31"
> format(x, "%d.%m.%Y")
[1] "31.12.1970"
> format(x, "%d.%b.%Y")
[1] "31.dec.1970"
> format(x, "%d %B %Y")
[1] "31 december 1970"
```

Formatna določila (npr: `'%d%b%Y'`) lahko uporabljamo za pretvorbo datumskih zapisov v obliki znakovnih nizov v datume razreda `Date`:

```
> x <- c("1jan1960", "2jan1960", "31mar1960", "30jul1960")
> (z <- as.Date(x, "%d%b%Y"))
[1] "1960-01-01" "1960-01-02" "1960-03-31" "1960-07-30"
> format(z, "%x")
[1] "1.1.1960" "2.1.1960" "31.3.1960" "30.7.1960"
```

S `'%x'` pretvorimo datum v standardno obliko glede na trenutni *locale* (za Slovenijo `d.m.llll`). Podrobnosti v zvezi s formatiranjem si oglejte v `help('strptime')`.

¹Namig: R programske stavke in izpise lahko skopirate na clipboard in jih prenesete v R z uporabo izbire `'Edit/Paste commands only'`

1.2 Število dni med dvema datumoma

Pri časovnih podatkih je smiselna računsko operacija odštevanje, ki ugotovi število dni med dvema datumoma:

```
> z
[1] "1960-01-01" "1960-01-02" "1960-03-31" "1960-07-30"
> diff(z)
Time differences in days
[1] 1 89 121
> Sys.Date() - as.Date("24.7.1953", "%d.%m.%Y")
Time difference of 19923 days
> as.Date("24.7.1953", "%d.%m.%Y") + 20000
[1] "2008-04-26"
```

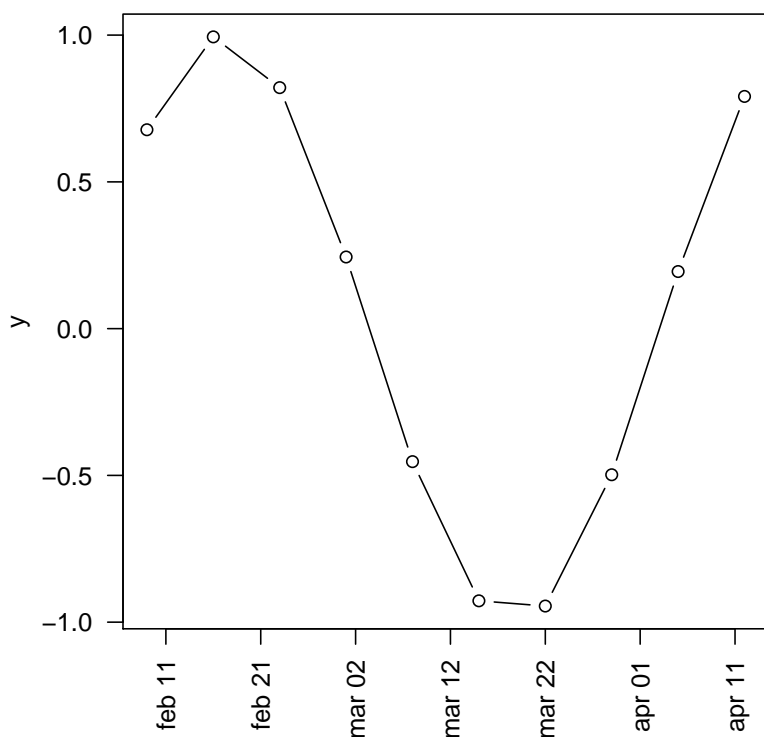
1.3 Risanje časovnih podatkov

Pripravimo nekaj podatkov za risanje: za naslednjih 10 tednov (**tenweeks**) bi radi narisali gibanje neke vrednosti shranjene v **y**. Pripravimo tudi primerno formatirano obliko datuma (**datumi** "dd. mm. yy"). Ker bomo datume na oseh izpisali pokončno (**las=2**), moramo določiti dovolj prostora pri prvi in tretji strani grafa (**mar = 6 ali 8**). Datumi na oseh so v skrajšanem formatu brez oznake leta. Da bo bolj zagatno, imamo v drugem tednu še en dodaten podatek

```
> today <- Sys.Date()
> (tenweeks <- seq(today, len = 10, by = "1 week"))

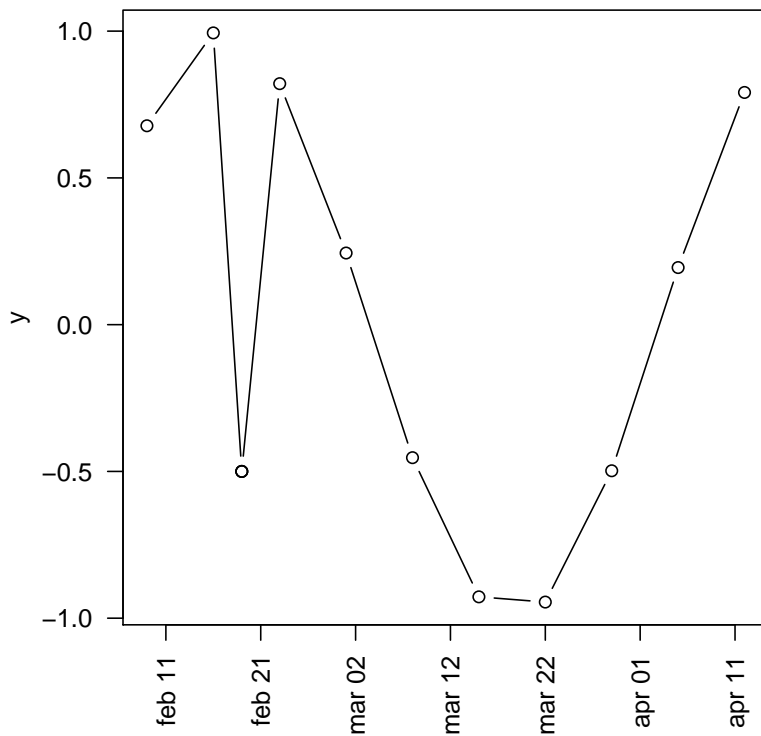
[1] "2008-02-09" "2008-02-16" "2008-02-23" "2008-03-01" "2008-03-08"
[6] "2008-03-15" "2008-03-22" "2008-03-29" "2008-04-05" "2008-04-12"

> y <- sin(as.numeric(tenweeks))
> datumi <- format(tenweeks, "%d. %m. %y")
> par(mar = c(6, 4, 4, 5))
> plot(tenweeks, y, type = "b", las = 2)
```



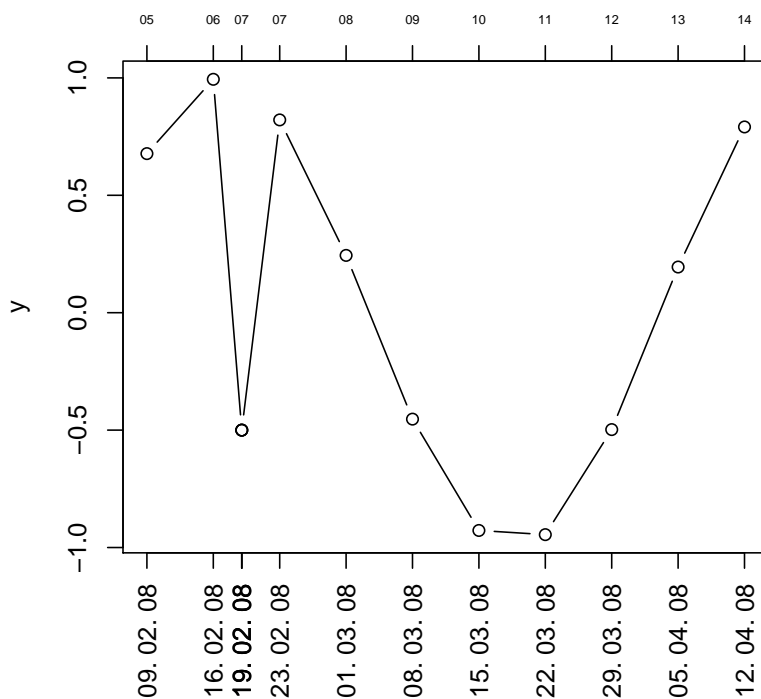
Da bo bolj zagatno, imejmo v drugem tednu še en dodaten podatek:

```
> (tenweeks <- sort(c(tenweeks, today + 10)))  
[1] "2008-02-09" "2008-02-16" "2008-02-19" "2008-02-23" "2008-03-01"  
[6] "2008-03-08" "2008-03-15" "2008-03-22" "2008-03-29" "2008-04-05"  
[11] "2008-04-12"  
  
> y <- c(y[1:2], -0.5, y[-(1:2)])  
> par(mar = c(6, 4, 4, 5))  
> plot(tenweeks, y, type = "b", las = 2)
```



Pri risanju lahko onemogočimo avtomatski izris **x**-osi (**xaxt='n'**) in s funkcijo **axis** dorišemo osi na prvo in (v našem primeru) tretjo stranico. Prvi ukaz nariše os s prednastavljenim formatom, drugemu (**axis.Date**) pa obliko izpisa določimo ob klicu in je zato morda priročnejši. V našem primeru s formatom **%U** na zgornjo os (3) izpišemo za vsak datum številko tedna v letu.

```
> par(mar = c(6, 4, 8, 5))
> plot(tenweeks, y, type = "b", xaxt = "n")
> datumi <- format(tenweeks, "%d. %m. %y")
> axis(1, at = tenweeks, labels = datumi, las = 2)
> axis.Date(3, at = tenweeks, format = "%U", cex.axis = 0.5)
```

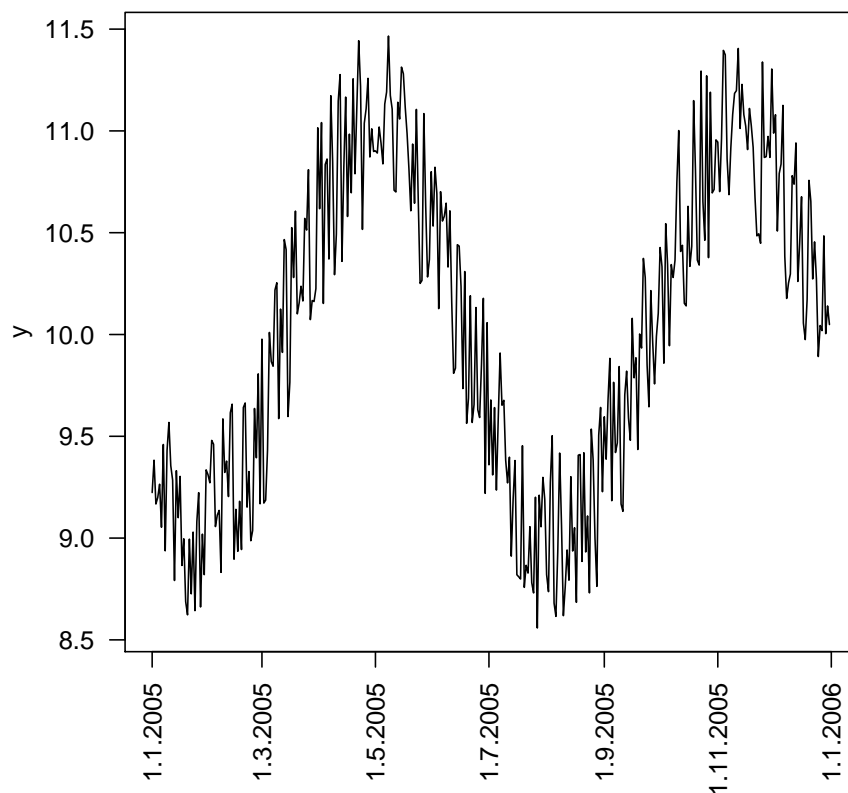


1.4 Podatki za eno leto

```
> start <- as.Date("2005-01-01")
> end <- as.Date("2005-12-31")
> dnevi <- seq(start, end, by = "1 day")
> head(format(dnevi, "%x"))

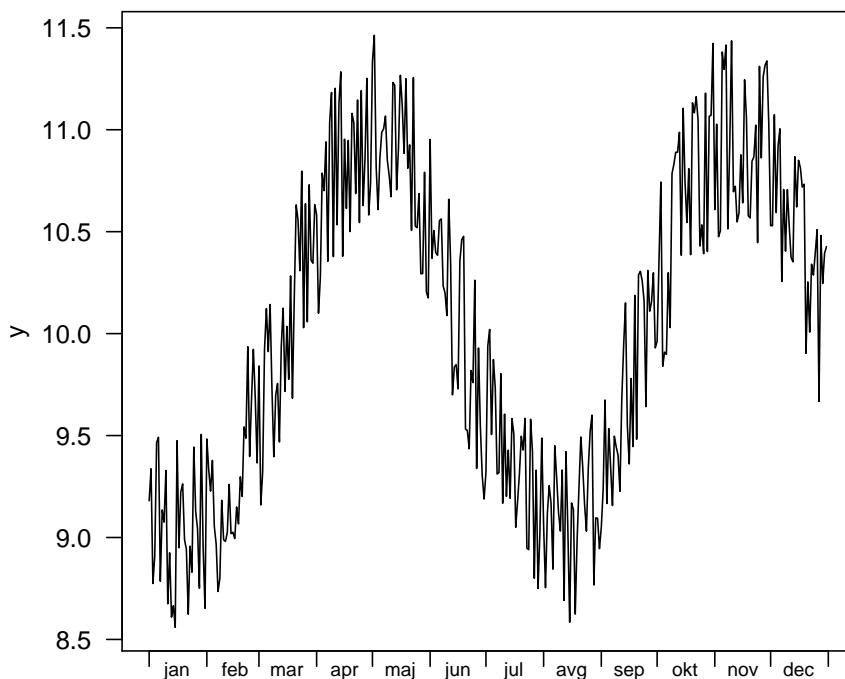
[1] "1.1.2005" "2.1.2005" "3.1.2005" "4.1.2005" "5.1.2005" "6.1.2005"

> y <- 10 + sin(as.numeric(dnevi)/31) + runif(length(dnevi), -0.5,
+      0.5)
> plot(dnevi, y, las = 2, type = "l", format = "%x")
```



Funkciji `axis` in `axis.Date` omogočata pripravo posebnih osi, npr. skalo razdelimo na mesečne intervale in pod sredino mesečnega intervala izpišemo kratko ime meseca:

```
> start <- as.Date("2005-01-01")
> end <- as.Date("2005-12-31")
> dnevi <- seq(start, end, by = "1 day")
> head(format(dnevi, "%x"))
[1] "1.1.2005" "2.1.2005" "3.1.2005" "4.1.2005" "5.1.2005" "6.1.2005"
> y <- 10 + sin(as.numeric(dnevi)/31) + runif(length(dnevi), -0.5,
+      0.5)
> plot(dnevi, y, las = 2, type = "l", xaxt = "n")
> mStart <- seq(start, end + 1, by = "1 month")
> axis(1, at = mStart, labels = rep("", length(mStart)), tick = 1)
> axis.Date(1, at = mStart + 15, format = "%b", tick = 0, cex.axis = 0.75,
+      line = -1)
```



2 Datum v POSIX obliki

Časovni podatki so lahko pripravljene tudi v POSIX obliki, ki vsebuje ne le datum ampak tudi dnevni čas in podatek o časovnem pasu.

```
> as.POSIXct(Sys.Date())
[1] "2008-02-09 01:00:00 Central Europe Standard Time"
```

Manipulacija s tako zapisanimi datumi je podobna kot je opisano zgoraj.

3 Povezava z Excel

Excel šteje datume kot dneve od 1. 1. 1900 (ta dan je dan 1), kar lahko preverimo, če v neko celico vpišemo vrednost 1, potem pa spremenimo format v datumskega. Pri branju z `read.xls` iz knjižnice `xlsReadWrite` se datumske vrednosti v R preberejo kot števila in jih je potrebno pretvoriti v datume.

Razlika v šteju dni je 25569. Razliko začetkov dobimo kot razliko dneva 0 za R (1. 1. 1970) in dneva 0 za Excel (31. 12. 1899):

```
> library("xlsReadWrite")

> deltaDate <- as.numeric((as.Date("1970-1-1") - as.Date("1899-12-31") +
+ 1))
> deltaDate
[1] 25569
```

V datoteki 'datumi.xls' imamo tri datume:

```
datum
1.1.1901
1.1.1970
24.7.1953
```

K datoteko preberemo, se v R prepisejo številske vrednosti, ki po pretvorbi v `class Date` kažejo nenavadne datume. Če odštejemo razliko v dneh med začetkoma štetja datumov 25569, dobimo prave datume, ki jih lahko poljubno formatiramo:

```
> (dateXls <- read.xls("../data/datumi.xls"))
  datum
1  367
2 25569
3 19564
> class(dateXls$datum) <- "Date"
> dateXls
  datum
1 1971-01-03
2 2040-01-03
3 2023-07-26
> dateXls$datum <- dateXls$datum - deltaDate
> dateXls
  datum
1 1901-01-01
2 1970-01-01
3 1953-07-24
> format(dateXls, "%x")
  datum
1  1.1.1901
2  1.1.1970
3 24.7.1953
```

Opomba: Excel je za datume do 1.3.1900 nezanesljiv, jih premakne za en dan nazaj.