

Uporaba R

2.del

A. Blejec

21. oktober 2009

Tabela s podatki

Table 4.1 Concentration of key metabolites in five patients. The urea concentration has not been recorded for patient 3

Patient number	Concentration of compound in whole blood (g/m^3)			
	Bilirubin	Cholesterol	Glucose	Urea
1	3.0	1150	750	310
2	4.5	1650	2200	200
3	5.0	2150	260	–
4	14.0	1200	650	270
5	3.5	2000	700	320

Priprava podatkov

Profil prvega pacienta

```
> profile <- c(3, 1150, 750, 310)
```

```
> profile
```

```
[1] 3 1150 750 310
```

Sprememba enote mere

$g/m^3 \rightarrow g/cm^3$

```
> profile * 1e-06
```

```
[1] 0.000003 0.001150 0.000750 0.000310
```

Podatkovni okvir:

```
data.frame
```

```
> bilirubin <- c(3, 4.5, 5, 14, 3.5)
> cholesterol <- c(1150, 1650, 2150, 1200, 2000)
> results <- data.frame(bilirubin, cholesterol)
```

```
> results
```

	bilirubin	cholesterol
1	3.0	1150
2	4.5	1650
3	5.0	2150
4	14.0	1200
5	3.5	2000

Urejanje podatkov

S funkcijo `edit()` lahko uredimo podatke

```
> results <- edit(results)
```

Dodajmo še dva stolpca z imenoma `glucose` in `urea`

```
> results[, "glucose"] <- c(750, 800, 260, 650, 700)
> results[, "urea"] <- c(310, 200, NA, 270, 350)
> results
```

	<code>bilirubin</code>	<code>cholesterol</code>	<code>glucose</code>	<code>urea</code>
1	3.0	1150	750	310
2	4.5	1650	800	200
3	5.0	2150	260	NA
4	14.0	1200	650	270
5	3.5	2000	700	350

Osnovna statistika podatkov: summary

```
> summary(results)
```

bilirubin	cholesterol	glucose	urea
Min. : 3.0	Min. :1150	Min. :260	Min. :200
1st Qu.: 3.5	1st Qu.:1200	1st Qu.:650	1st Qu.:252
Median : 4.5	Median :1650	Median :700	Median :290
Mean : 6.0	Mean :1630	Mean :632	Mean :282
3rd Qu.: 5.0	3rd Qu.:2000	3rd Qu.:750	3rd Qu.:320
Max. :14.0	Max. :2150	Max. :800	Max. :350
			NA's : 1

Izbira stolpcev in vrstic

```
> results[4, ]
  bilirubin cholesterol glucose urea
4         14         1200      650  270

> results[4, 1]
[1] 14

> results[, 2]
[1] 1150 1650 2150 1200 2000

> results[1:3, -2]
  bilirubin glucose urea
1         3.0      750  310
2         4.5      800  200
3         5.0      260   NA

> results[c(2, 4), 1:2]
  bilirubin cholesterol
2         4.5         1650
4        14.0         1200
```

Organizacija podatkov

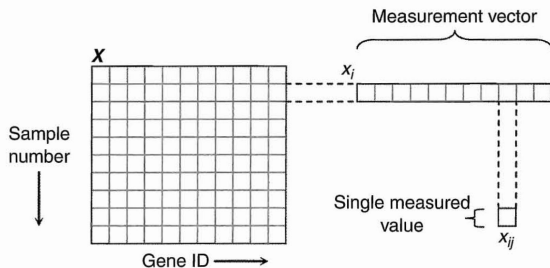
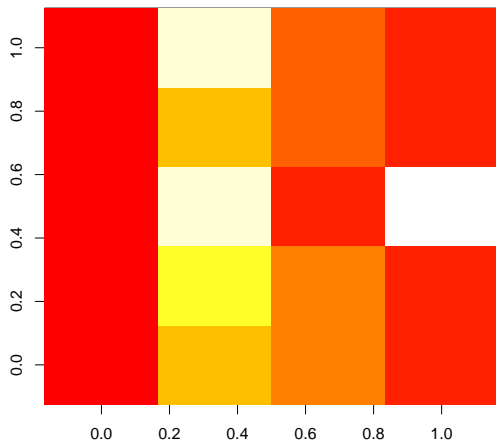


Fig. 4.2 Organization of gene experimental data into a data matrix. For data from a gene expression microarray, the row vector x_i would be the gene expression profile over all genes for a specific sample. Each column of the matrix captures the variation of an individual gene over all samples.

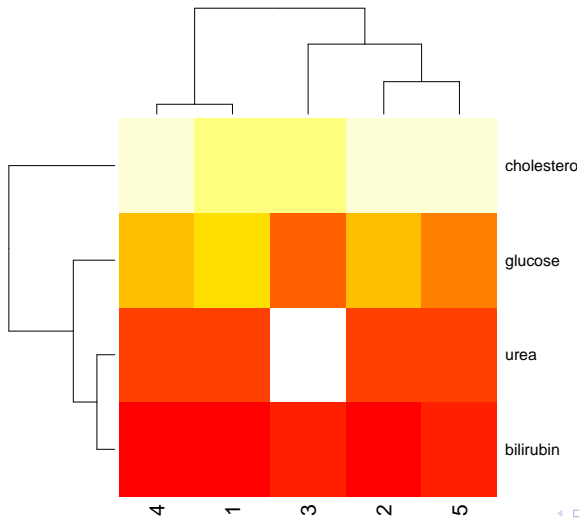
image

```
> image(t(as.matrix(results)))
```



heatmap

```
> heatmap(t(as.matrix(results)), scale = "column", cexRo
```



Programske ukazi, ki jih uporabljamo so lahko kjerkoli ...

```
> source("http://www.bixsolutions.net/profiles.r")
```

Funkcija apply

Maximum v stolpcih

```
> apply(results, 2, max, na.rm = TRUE)
```

bilirubin	cholesterol	glucose	urea
14	2150	800	350

Funkcije

n-ti koren

```
> koren <- function(x, n = 2) {  
+   y <- x^(1/n)  
+   return(y)  
+ }  
> koren(2)  
  
[1] 1.414214  
  
> koren(1:4)  
  
[1] 1.000000 1.414214 1.732051 2.000000  
  
> koren(8, 3)  
  
[1] 2
```

Primer priprave funkcije

Funkcija za skaliranje po stolpcih

```
> rangescale <- function(X, na.rm = TRUE) {  
+   Xmax <- apply(X, 2, max, na.rm = na.rm)  
+   Xscaled <- scale(X, scale = Xmax, center = FALSE)  
+   return(Xscaled)  
+ }
```

heatmap

```
> scaledResults <- rangescale(results)
> round(scaledResults, 2)
```

```
      bilirubin cholesterol glucose urea
[1,]      0.21         0.53    0.94 0.89
[2,]      0.32         0.77    1.00 0.57
[3,]      0.36         1.00    0.32  NA
[4,]      1.00         0.56    0.81 0.77
[5,]      0.25         0.93    0.88 1.00
```

```
attr(,"scaled:scale")
```

```
      bilirubin cholesterol      glucose      urea
           14           2150           800           350
```

Slika skaliranih podatkov

```
> heatmap(t(as.matrix(scaledResults)), scale = "column",  
+         cexRow = 1.2)
```

