



Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

# Programiranje v R-ju

## 5. Prikazi podatkov

Vladimir Batagelj

Univerza v Ljubljani, FMF, Matematika

Finančna matematika  
Ljubljana, februar 2009  
3. december 2012



# Kazalo

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

- 1 Prikazi podatkov
- 2 Porazdelitve v R-ju
- 3 Modeli
- 4 R in funkcije
- 5 ggplot2
- 6 Zemljevidi



# Prikazi podatkov

## Prikazi podatkov

V. Batagelj

Prikazi podatkov

Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Za prikaze podatkov imamo v R-ju na voljo nekaj **knjižnic**. Osnovna knjižnica je `graphics`. Pozna tudi S-ovo knjižnico `lattice`. V razvoju je nova knjižnica `grid` (Paul Murrell), ki je precej zmogljivejša od osnovne in je osnova večine novejših knjižnic za prikaze. Trirazsežne interaktivne prikaze omogočata knjižnici `rgl` in `rggobi` (vmesnik za program GGobi). Leta 2005 je Hadley Wickham začel razvijati knjižnico `ggplot2`, ki temelji na knjigi Leland Wilkinson: *The Grammar of Graphics* (2005).

Različne posebne slikovne prikaze najdemo tudi v drugih knjižnicah. Na primer: drevesa razvrstitev in prikaze na zemljevidih.



# Prikazi podatkov

## Prikazi podatkov

V. Batagelj

Prikazi podatkov

Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Posamezni prikaz ustvarimo s funkcijo `plot` ali posebnimi funkcijami `hist`, `barplot`, `pie`, `dotchart`, `pairs`, `qqplot`, `curve`, `persp`, ... Prikaz lahko dopolnimo s funkcijami `points`, `lines`, `abline`, `polygon`, `arrows`, `rect`, `text`, ...

Izvedbo prikaza lahko nadziramo z več parametri (funkcija `par`) `col`, `bg`, `fg`, `bty`, `lty`, `lwd`, `pch`, `cex`, ... Na posamezni sliki lahko združimo več prikazov – parametra `mfrow` in `mfcol`.

Sliko lahko pripravimo v različnih *oblikah* (devices) `windows` (zaslon), `postscript`, `pdf`, `pictex`, `png`, `svg`, ... Z zahtevo

```
> pdf("Slika.pdf",width=11.7,height=8.3,paper="a4r")
```

povemo, da bi radi sliko ustvarili na datoteki `Slika.pdf` oblike pdf velikosti A4. Izdelavo slike zaključimo z zahtevo

```
> dev.off()
```



# Prikazi podatkov

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

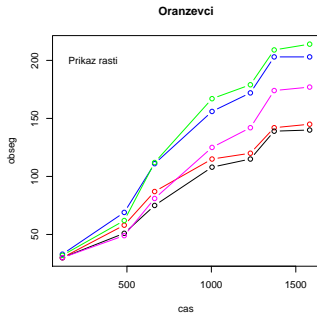
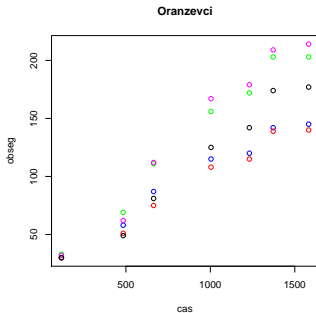
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> help(plot)
> (c <- Orange[29:35,2])
[1] 118 484 664 1004 1231 1372 1582
> b <- c("red", "blue", "black", "green", "magenta")
> plot(Orange[,2], Orange[,3], col=b[Orange[,1]], xlab="cas", ylab="obseg",
      main="Oranzevci")
> plot(Orange[,2], Orange[,3], xlab="cas", ylab="obseg", main="Oranzevci",
      type="n")
> for(k in 1:5){points(c, Orange[(7*k-6):(7*k),3], col=b[k], type="b")}
> text(300,200,"Prikaz rasti")
```



# Prikazi – znaki

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

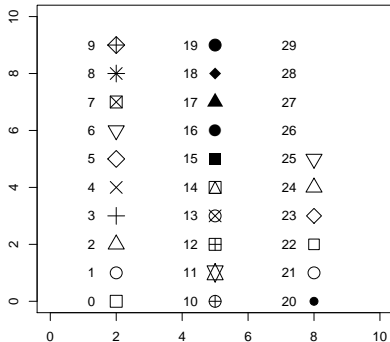
Modeli

R in funkcije

ggplot2

Zemljevidi

Znaki



```
> plot(0:10,0:10,type="n",main="Znaki",xlab="",ylab="")
> k <- -1
> for(i in c(2,5,8)){for(j in 0:9){
  k <- k+1;text(i-0.75,j,k);points(i,j,pch=k,cex=2)}}}
```



# Prikazi – barve

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



Spectral (divergent)

```
> colors()
[1] "white"
[655] "yellow3"
> library(RColorBrewer)
> display.brewer.pal(11, 'Spectral')
> help(rgb); help(palette); help(RColorBrewer)
```



# Prikaz – imenska in številka spremenljivka

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

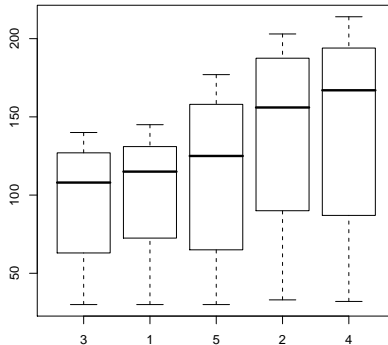
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> plot(Orange$Tree, Orange$circumference)
```





# Prikazi – imenska in številka spremenljivka

## Prikazi podatkov

V. Batagelj

Prikazi podatkov

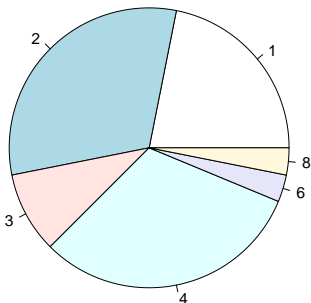
Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

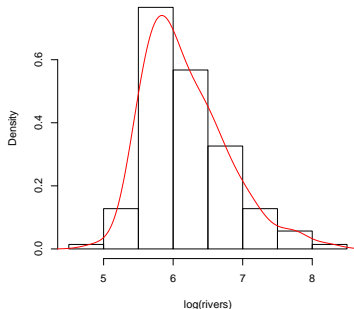
Zemljevidi



```
> table(mtcars$carb)
 1  2  3  4  6  7  8 
10 10  1 10  1  1  1
```

```
> barplot(table(mtcars$carb))
> pie(table(mtcars$carb))
```

Histogram of log(rivers)



```
> dotchart(table(mtcars$carb))
> stripchart(mtcars$carb,method="stack",pch=16)
> hist(log(rivers),prob=TRUE)
> lines(density(log(rivers)),col="red")
```



# Razni prikazi

## Prikazi podatkov

V. Batagelj

Prikazi podatkov

Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

```
> attach(faithful)
> hist(waiting)
> summary(waiting)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  43.0   58.0   76.0   70.9   82.0   96.0
> bins <- seq(42,109,by=10)
> bins
[1] 42 52 62 72 82 92 102
> freqs <- table(cut(waiting,bins))
> y <- c(0,freqs,0)
> x <- seq(37,107,by=10)
> plot(x,y,type="l")
> rug(waiting)
> hist(waiting,breaks="Scott",prob=TRUE,ylab="",main="Faithful")
> lines(density(waiting),col="blue",lwd=2)
> boxplot(rivers)
> plot(rev(rivers[order(rivers)]))
> boxplot(rivers)
> f <- fivenum(rivers)
> f
[1] 135 310 425 680 3710
> text(rep(1.3,5),f,labels=c("min","1/4","1/2","3/4","max"))
```



# Prikazi – odvisnosti med spremenljivkami

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

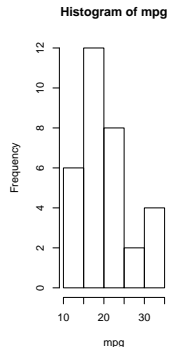
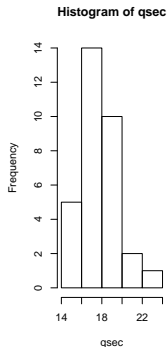
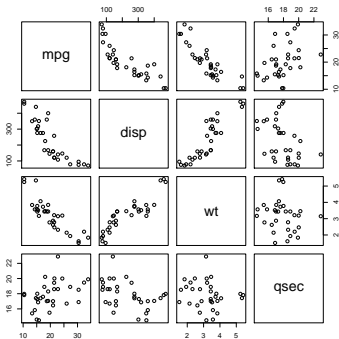
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> attach(mtcars)
> pairs(mtcars[,c(1,3,6,7)])
> par(mfrow=c(1,2))
> hist(qsec,breaks="scott")
> hist(mpg,breaks="scott")
> par(mfrow=c(1,1))
```



# Prikaz porazdelitve s stopničasto funkcijo

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

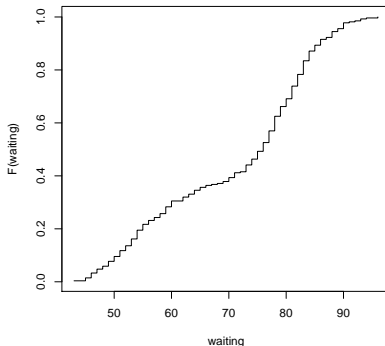
Modeli

R in funkcije

ggplot2

Zemljevidi

Porazdelitev waiting



```
> attach(faithful)
> n <- length(waiting)
> plot(sort(waiting), (1:n)/n, type="s", main="Porazdelitev waiting",
      xlab="waiting", ylab="F(waiting)")
> plot(ecdf(waiting)) # empirical cumulative distribution function
```



# Porazdelitve v R-ju

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Večina standardnih porazdelitev je že vgrajenih v R kot ustrezne funkcije. Za porazdelitev *dist* so: *ddist* – gostota porazdelitve  $g(x)$ , *pdist* – porazdelitvena funkcija  $F(x) = \int_{-\infty}^x g(t)dt$ , *qdist* – obratna – kvantilna funkcija  $q = F^{-1}(p)$ , *rdist* – zaporedje slučajnih števil porazdeljeno po *dist*.

Porazdelitve *dist* (za podrobnosti `help`): `unif`, `beta`, `binom`, `cauchy`, `exp`, `chisq`, `f`, `gamma`, `geom`, `hyper`, `lnorm`, `logis`, `nbinom`, `norm`, `pois`, `signrank`, `t`, `weibull`, `wilcox`.

Funkcija `sample` omogoča slučajno vzorčenje (`replace=TRUE` – z vračanjem) iz dane množice.

Z večanjem števila členov v povprečju gre porazdelitev povprečja slučajne spremenljivke (s končno varianco) proti normalni porazdelitvi.



# Prikazi – centralni limitni izrek

## Prikazi podatkov

### V. Batagelj

Prikazi podatkov

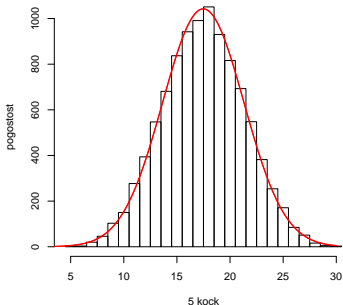
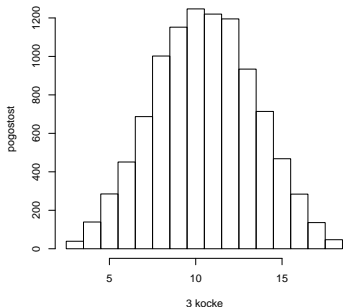
Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> a <- sample(1:6,replace=TRUE,10000); b <- sample(1:6,replace=TRUE,10000)
> c <- sample(1:6,replace=TRUE,10000); s <- a+b+c
> hist(s,breaks=2.5:18.5,xlab="3 kocke",ylab="pogostost",main="")
> d <- sample(1:6,replace=TRUE,10000); e <- sample(1:6,replace=TRUE,10000)
> s <- s+d+e; x <- seq(1,30,0.1)
> hist(s,breaks=4.5:30.5,xlab="5 kock",ylab="pogostost",main="")
> lines(x,dnorm(x,mean(s),sd(s))*10000,lwd=2,col="red")
```



# Prikazi – odvisnosti med porazdelitvami

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

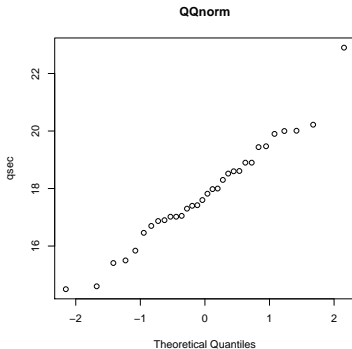
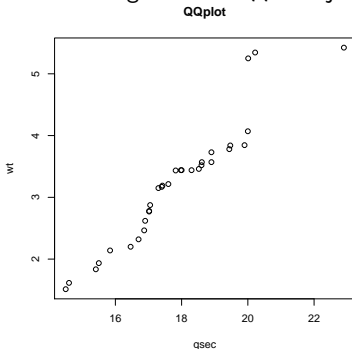
Modeli

R in funkcije

ggplot2

Zemljevidi

QQplot sestavljajo točke  $(x, y)$  nad definicijskima intervaloma porazdelitev  $F_1$  in  $F_2$ , za katere velja  $F_1(x) = F_2(y)$ . Če sta porazdelitvi enaki, ležijo točke na diagonali. Pri QQnorm je porazdelitev  $F_1$  normalna.



```
> attach(mtcars)
> qqplot(qsec,wt,main="QQplot")
> qqnorm(qsec,ylab="qsec",main="QQnorm")
```



# Prikazi – binomska porazdelitev

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

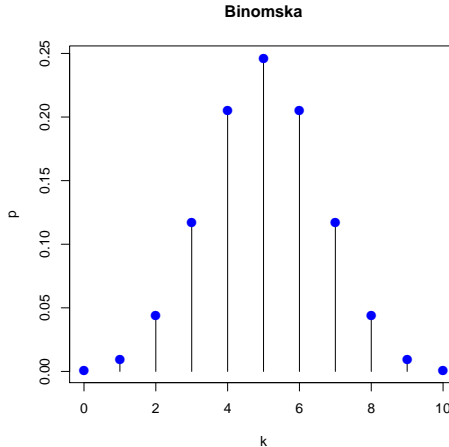
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> h <- dbinom(0:10,size=10,prob=0.5)
> plot(0:10,h,type="h",main="Binomska",xlab="k",ylab="p")
> points(0:10,h,pch=16,cex=1.5,col="blue")
```



V. Batagelj

Prikazi podatkov





# Prikazi – modeli

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Z izrazom oblike  $odvisna \sim f(neodvisne)$  v R-ju opišemo *model* – obliko zveze med odvisno in neodvisnimi spremenljivkami. Obstaja več funkcij, ki iz danih podatkov tak model določijo. Npr. `lm`, `gam`, `loess`, `lowess`, ... Vrednost modelne funkcije v izbranih točkah dobimo s funkcijo `predict`.



# Prikazi – regresijska premica

## Prikazi podatkov

### V. Batagelj

Prikazi podatkov

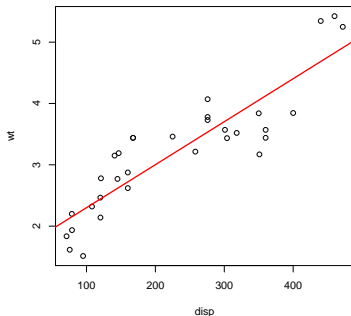
Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> attach(mtcars)
> res <- lm(wt ~ disp)
> res[[1]]
(Intercept)      disp
1.599814597 0.007010325
```

```
> plot(wt ~ disp)
> abline(res,col="red",lwd=2)
> predict(res,list(dis=c(410,200)))
      1      2
4.474048 3.001880
```



# Prikazi – potenčni zakon (Zipf, Lotka, Pareto)

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

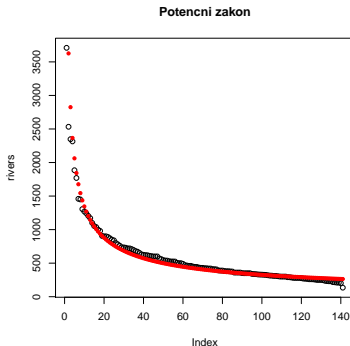
R in funkcije

ggplot2

Zemljevidi

Če vrednosti neke spremenljivke uredimo v padajočem vrstnem redu in jih prikažemo logaritemski lestvici na obeh oseh ( $x$  – zaporedna številka,  $y$  – vrednost spremenljivke), dobimo pogosto za sliko (skoraj) premico. Torej količini povezuje **potenčni zakon**  $y \sim cx^k$ .

```
> plot(rev(sort(rivers)))
> plot(rev(sort(rivers)), log="xy")
> x <- log(1:length(rivers))
> y <- log(rev(sort(rivers)))
> plot(y ~ x)
> rp <- lm(y ~ x)
> rp[[1]]
(Intercept)          x
 8.6233680    -0.6160568
> abline(rp, col="red", lwd=2)
> plot(rev(sort(rivers)), ylab="rivers",
      main="Potencni zakon")
> pot <- function(x){exp(8.623)*x^(-0.616)}
> x <- 1:length(rivers)
> y <- pot(x)
> points(x,y,pch=20,col="red")
```





# Prikazi – neparametrično glajenje / Boston

## Prikazi podatkov

V. Batagelj

Prikazi podatkov

Porazdelitve R-ju

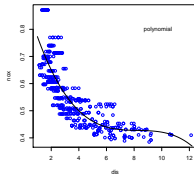
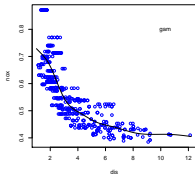
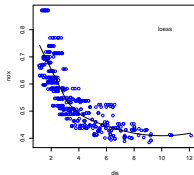
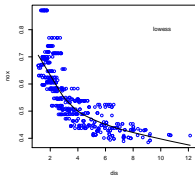
Modeli

R in funkcije

ggplot2

Zemljevidi

```
> library(MASS); attach(Boston)
> pairs(Boston)
> plot(dis,nox); s <- order(dis)
> plot(dis,nox,col="blue")
> lines(dis[s],nox[s])
> par(mfrow=c(2,2),cex=0.5)
> plot(dis,nox,col="blue")
> text(11,0.8,"lowess",pos=2)
> lines(lowess(dis,nox))
> plot(dis,nox,col="blue")
> text(11,0.8,"loess",pos=2)
> model <- loess(nox ~ dis)
> x <- seq(1,12.2,0.05)
> y <- predict(model,data.frame(dis=x))
> lines(x,y)
> plot(dis,nox,col="blue")
> text(11,0.8,"gam",pos=2)
> library(mgcv)
> model <- gam(nox ~ s(dis))
> y <- predict(model,list(dis=x))
> lines(x,y)
> plot(dis,nox,col="blue")
> text(11,0.8,"polynomial",pos=2)
> model <- lm(nox ~ dis+I(dis^2)+I(dis^3))
> y <- predict(model,list(dis=x))
> lines(x,y)
> par(mfrow=c(1,1),cex=1)
```





# Funkcije in R

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

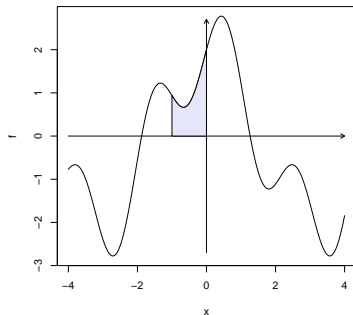
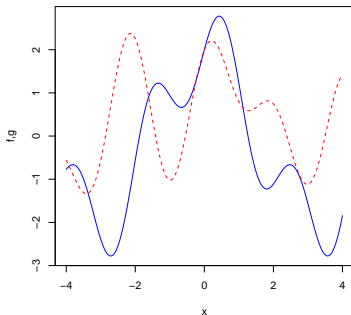
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
f <- function(x){sin(3*x)+2*cos(x)}
g <- function(x){sin(2*x)+cos(x/2)+cos(3*x)}
x <- seq(-4,4,0.05); yf <- f(x); yg <- g(x)
plot(c(x,x),c(yf,yg),xlab="x",ylab="f,g",type="n")
lines(x,yf,lty=1,col="blue"); lines(x,yg,lty=2,col="red")
plot(x,yf,xlab="x",ylab="f",type="l")
arrows(-4,0,4,0,length=0.07); arrows(0,-2.7,0,2.7,length=0.07)
polygon(c(x[(-1<=x)&(x<=0)],0,-1),c(yf[(-1<=x)&(x<=0)],0,0),col="lavender")
```



# Funkcije in R

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

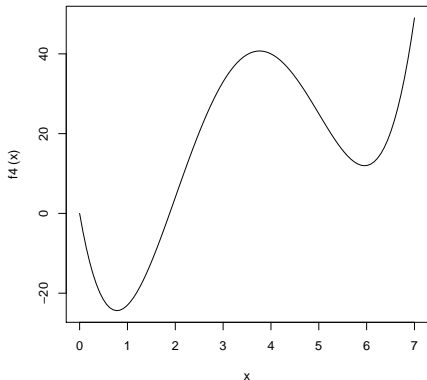
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> f4 <- function(x){x^4 - 14*x^3 + 60*x^2 - 70*x }
> curve(f4,0,7)
> nicla <- uniroot(f4,lower=1,upper=3)
> unlist(nicla)
      root      f.root      iter      estim.prec
1.882058e+00 -2.013040e-06 5.000000e+00 6.103516e-05
> nicla$root
[1] 1.882058
```



# Odvajanje in integriranje

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Za simbolno odvajanje nerazčlenjenih izrazov ('zaščitenih' s funkcijo `expression`) je na voljo funkcija `D`. Žal R dobljenih izrazov-odvodov ne poenostavlja. Za uporabe pri optimizaciji obstaja še funkcija `deriv`. Funkcija `integrate` ( $f, a, b$ ) izračuna vrednost določenega integrala funkcije  $f(x)$  na intervalu  $(a, b)$ . Meji sta lahko tudi v neskončnosti (`Inf`).

```
> D(expression(2*x^3+7*x-1), "x")
2 * (3 * x^2) + 7
> D(expression(sin(cos(x+3*y))), 'y')
-(cos(cos(x + 3 * y)) * (sin(x + 3 * y) * 3))
> D(D(expression(sin(x^2)), "x"), "x")
cos(x^2) * 2 - sin(x^2) * (2 * x) * (2 * x)
> D(D(expression(3*x^2*y+5*x*y^3), "x"), "y")
3 * (2 * x) + 5 * (3 * y^2)
> odvod <- function(f,h){function(x){(f(x+h)-f(x-h))
> f <- function(x){sin(x)+cos(2*x)}
> g <- odvod(f,0.001)
> x <- seq(-4,4,0.1)
> plot(x,g(x),type='l',col='blue')
> lines(c(-4,4),c(0,0))
> lines(x,f(x),type='l',col='red')
> integrate(f,lower=-4,upper=4)
0.9893582 with absolute error < 4.9e-06
> integrate(dnorm,-Inf,Inf)
1 with absolute error < 9.4e-05
```



# Funkcije in R

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

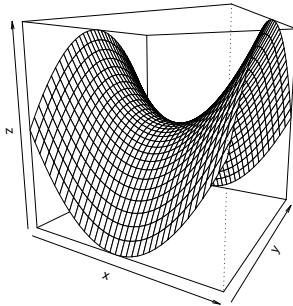
Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



```
> y <- x <- seq(-1,1,length=35)
> z <- outer(x,y,function(x,y){x*x-y*y})
> persp(x,y,z,theta=30,phi=10)
```





# Optimizacija v R

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

`optimize` – enorazsežna optimizacija; `optim`, `constrOptim` – večrazsežna optimizacija (z linearnimi omejitvami); `nlm` – večrazsežna optimizacija; `lm` – linearni modeli; `glm` – posplošeni linearni modeli; `nls` – metoda najmanjših kvadratov (primeri **NIST**, za standardne modele so pripravljene funkcije `SS` (self-starting)), `lsfit` – metoda najmanjših kvadratov za model  $y = Xb + e$ . `lp` – iz paketa `lpSolve`, linearna optimizacija. Funkcija `lp` predpostavlja, da so vse spremenljivke nenegativne – če je  $z$  neomejena, uporabimo zamenjavo  $z = u - v$ , kjer sta  $u, v \geq 0$ . Maksimizacijo zahtevamo s parametrom `direction="max"`. V primeru je rešena naloga

$$P(x, y, z) = 12x - 10y - 30z \text{ in } \Phi = \{(x, y, z) : -3x + 2y + 8z \leq 17, -x + y + 3z \leq 9, -2x + y + 8z \leq 16, x \geq 0, y \geq 0, z \geq 0\}.$$



# Optimizacija – primeri

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

```
f4 <- function(x){x^4 - 14*x^3 + 60*x^2 - 70*x}
curve(f4,0,7)
m4 <- optimize(f4,interval=c(5,7),tol=0.000001)
lines(c(m4$min,m4$min,-100),c(-100,m4$obj,m4$obj),col="red")

fr <- function(x) { ## Rosenbrock Banana function
  x1 <- x[1]; x2 <- x[2];
  100 * (x2 - x1^2)^2 + (1 - x1)^2
}
gr <- function(x) { ## Gradient of 'fr'
  x1 <- x[1]; x2 <- x[2]
  c(-400*x1*(x2 - x1^2) - 2*(1 - x1),200*(x2 - x1^2))
}
m <- optim(c(-1.2,1),fr,control=list(trace=TRUE))
mg <- optim(c(-1.2,1),fr,gr=gr,method="BFGS",
  control=list(trace=TRUE))

> library(lpSolve)
> s <- lp(objective.in=c(12,-10,-30),
+ const.mat=matrix(c(-3,2,8,-1,1,3,-2,1,8),nrow=3),
+ const.rhs=c(17,9,16), const.dir=c("<=","<=","<="))
> s
Success: the objective function is -60
> s$solution
[1] 0 0 2
```

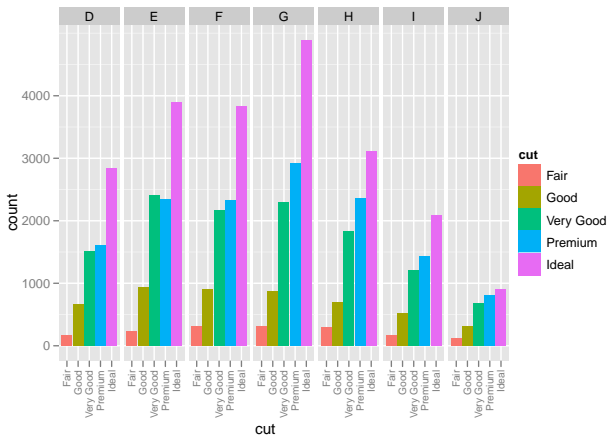


# ggplot2

Prikazi  
podatkov

V. Batagelj

```
> library(ggplot2)
> qplot(cut, data = diamonds, geom = "bar", fill = cut) + facet_grid(. ~ color) +
+ opts(axis.text.x = theme_text(angle = 90, hjust = 1, size = 8, colour = "grey50"))
> ggsave("cut.pdf", width = 7, height = 5)
```



Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

V. Batagelj

Prikazi podatkov





# Zemljevidi

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Osnovna knjižnica za delo z zemljevidi je `sp`. Veliko koristnih funkcij najdemo še v knjižnicah `maps`, `mapproj`, `maptools` in `rgdal`. Nekaj zemljevidov najdemo v `mapdata`, še več pa na spletu – npr. na <http://www.maplibrary.org/stacks/gadm/> oziroma <http://gadm.org/>, <http://www.cdc.gov/epiinfo/shape.htm>. Za barve se naslonimo na knjižnico `RColorBrewer`.

Opisi zemljevidov se nahajajo v datotekah `shape/ESRI`. Posamezni opis zemljevida sestavlja več datotek – vsaj tri: `dbf` (podatki), `shp` (obrisi), `shx` (index). Za slovenske občine so pripravljene podatki na [Slovenija](#); veliko različnih podatkov dobimo na [SURSu](#).



# Slovenija: delež tujcev po občinah

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

```
> setwd("C:/Users/....")
> library(maptools)
> pop <- read.table("population.csv", sep = ";", as.is = TRUE,
+ col.names = c("commune", "slovenians", "others"))
> pop$total <- pop$slovenians+pop$others
> pop$proportion <- pop$others/pop$total
> pop[order(pop$proportion), ]
      commune slovenians others total proportion
27          Dobje          954     0    954 0.000000000
191         Velika Polana    1478     0   1478 0.000000000
161 Sveti Andraž v Slov. goricah 1208     2   1210 0.001652893
164         Sveti Tomaž     2106     4   2110 0.001895735
15          Cankova        1903     4   1907 0.002097535
.....
68          Koper/Capodistria 48171  3744 51915 0.072117885
26          Divača          3578    279  3857 0.072336012
117         Osilnica         369     33    402 0.082089552
150         Sežana         11720  1108 12828 0.086373558
87          Loška dolina    3590    340   3930 0.086513995
56          Izola/Isola    14482  1553 16035 0.096850639
> SIcom <- readShapeSpatial("OB/OB.shp")
> comOrder <- rank(SIcom$OB_UIME)
> SIcom$proportion <- pop$proportion[comOrder]
> spplot(SIcom,"proportion")
```



# Slovenija: delež tujcev po občinah

Prikazi podatkov

V. Batagelj

Prikazi podatkov

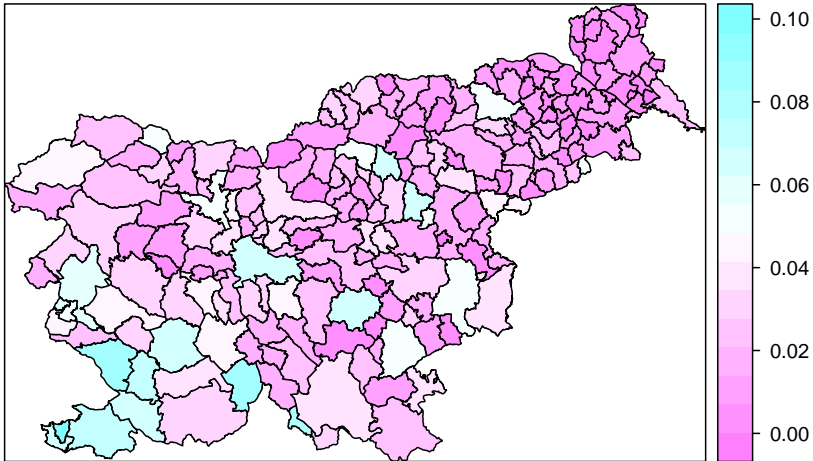
Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi



V. Batagelj

Prikazi podatkov



# Romunija: nezaposlenost maj 2010

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Podatki:

[http://www.insse.ro/cms/files/arhiva\\_buletine2010/bsl\\_5.pdf](http://www.insse.ro/cms/files/arhiva_buletine2010/bsl_5.pdf) / stran  
128

Datoteka shape: <http://www.cdc.gov/epiinfo/europe.htm>  
Okrožje Ilfov (okolica Bukarešte) ni v datoteki shape.

```
> setwd("C:/Users/....")
> library(maptools)
> gpclibPermit()
> ro <- read.csv2("unemployedRO.txt",skip=3,colClasses = "character")
> ROcou <- readShapeSpatial("epi/ro.shp")
> names(ROcou)
> cbind(ro[[1]],levels(ROcou$ADMIN_NAME))
> un <- sapply(ro[,-1],as.numeric)
> rownames(un) <- ro[[1]]
> rate <- un[1:41,"Rate.T"]
> brks <- quantile(rate, seq(0,1,1/7))
> library(RColorBrewer)
> pal <- brewer.pal(8,'Greens')
> plot(ROcou,col=pal[findInterval(rate,brks,all.inside=TRUE)])
> title("Unemployment May 2010")
> text(coordinates(ROcou),labels=as.character(ROcou$ADMIN_NAME),cex=0.4)
> legend("topright",
+ legend=c("2.5-6.2","6.2-7.8","7.8-8.2","8.2-8.7","8.7-9.8","9.8-11.2","11.2-14.1"),
+ fill=pal,cex=0.75)
```



# Romunija: nezaposlenost maj 2010

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

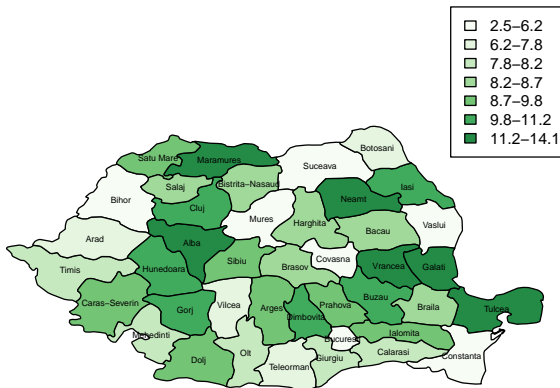
Modeli

R in funkcije

ggplot2

Zemljevidi

## Unemployment May 2010







# Okrožja (counties) ZDA

Prikazi  
podatkov

V. Batagelj

Prikazi  
podatkov

Porazdelitve v  
R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Datoteke shape: **Gadm**.

```
> USout <- readShapeSpatial("USA/USA_adm0.shp")
> USsta <- readShapeSpatial("USA/USA_adm1.shp")
> UScou <- readShapeSpatial("USA/USA_adm2.shp")
> pdf("UScounties.pdf",width=11.7,height=8.3,paper="a4r")
> plot(UScou,xlim=c(-124,-67),ylim=c(23,48),col="wheat",bg="skyblue",border="red",lwd=0.05)
> plot(USsta,xlim=c(-124,-67),ylim=c(23,48),lwd=0.2,border="blue",add=TRUE)
> plot(USout,xlim=c(-124,-67),ylim=c(23,48),lwd=0.2,add=TRUE)
> title("Central US"); dev.off()
```

Dobljena slika `UScounties.pdf` je zelo obsežna (okrog 70 MB).

Zato je shranjena na posebni **datoteki**.

Na spletišču **Gadm** lahko dobimo zemljevide tudi v obliki `Rdata`. V R jih vnesemo takole:

```
> load("USA/USA_adm0.RData"); USout <- gadm
> load("USA/USA_adm1.RData"); USsta <- gadm
> load("USA/USA_adm2.RData"); UScou <- gadm
```



# Razvrščanje v skupine – perunike

Prikazi podatkov

V. Batagelj

Prikazi podatkov

Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

Za določanje skupin v podatkih se uporabljajo metode razvrščanja v skupine. Pri metodah združevanja lahko izid slikovno predstavimo z drevesom razvrščanja – dendrogramom.

```
> help(iris)
> attach(iris)
> z <- function(x){(x-mean(x))/sd(x)}
> d <- cbind(z(Sepal.Length),z(Sepal.Width),z(Petal.Length),z(Petal.Width))
> iris
      Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
1              5.1           3.5           1.4           0.2     setosa
2              4.9           3.0           1.4           0.2     setosa
...
150             5.9           3.0           5.1           1.8  virginica
> d
      [,1]      [,2]      [,3]      [,4]
[1,] -0.89767388  1.01560199 -1.33575163 -1.3110521482
[2,] -1.13920048 -0.13153881 -1.33575163 -1.3110521482
...
[150,]  0.06843254 -0.13153881  0.76021149  0.7880306775
> t <-hclust(dist(d))
> plot(t,main="iris",cex=0.2)
> library(cluster)
> r <- agnes(d,method="ward")
> plot(r,which.plots=2,main="iris",cex=0.2)
```



# Drevo razvrstitve – perunike

Prikazi podatkov

V. Batagelj

Prikazi podatkov

Porazdelitve v R-ju

Modeli

R in funkcije

ggplot2

Zemljevidi

