Analiza podataka i obrada informacija

Nositelj: izv. prof. dr. sc. Siniša Sovilj **Asistent**: mag. inf. Alesandro Žužić

Ustanova: Sveučilište Jurja Dobrile u Puli, Fakultet informatike u Puli



[2] Vizualizacija podataka

R nudi napredne mogućnosti za **grafičku vizualizaciju podataka**s bogatim skupom biblioteka i funkcija, omogućava učinkovito stvaranje vizualnih prikaza podataka, olakšavajući analizu i interpretaciju rezultata. Neke od osnovnih grafičkih prikaza koje ćemo proći su: osnovni X-Y prikaz: plot(), linijski prikaz: lines(), histogram: hist(), točkasti prikaz: dotchart(), stupčasti prikaz: barplot(), tortni prikaz: pie(), kutijasti prikaz: boxplot(), raspršeni graf: scatterplot()



Posljednje ažurirano: 19. ožujka 2025.

- Analiza podataka i obrada informacija
- [2] Vizualizacija podataka
 - X-Y prikaz plot
 - Osnovna sintaksa
 - Slova i margine
 - Abline
 - Lines
 - Ostali elementi

X-Y prikaz plot()

X-Y prikaz u R-u koristi se za vizualizaciju odnosa između dviju numeričkih varijabli. Ovaj tip grafa je koristan za ispitivanje korelacije ili uzorka među podacima, a osnovna funkcija za izradu ovih prikaza je plot(). X-Y prikaz daje jasnu sliku o tome kako se jedna varijabla mijenja u odnosu na drugu.

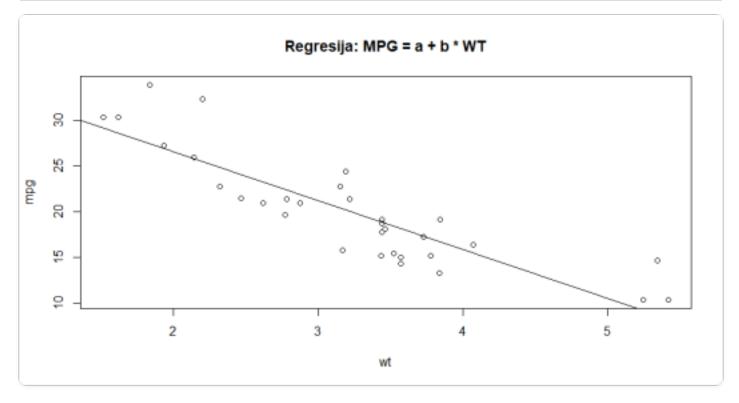
Osnovna sintaksa

U R-u, funkcija plot() koristi dva osnovna argumenta: **x** i **y**. Prvi argument predstavlja vrijednosti na *X-osi*, a drugi na *Y-osi*.

Na primjer, ako imamo skup podataka o težini vozila (wt) i potrošnji goriva (mpg) iz skupa podataka mtcars, možemo koristiti X-Y prikaz za vizualizaciju kako težina vozila utječe na njegovu potrošnju goriva.

Primjer.

```
attach(mtcars) # Učitavanje podataka
plot(wt, mpg) # Iscrtavanje podataka
abline(lm(mpg ~ wt)) # Dodavanje regresijskog pravca
title("Regresija: MPG = a + b * WT") # Naslov grafa
```



1. **attach(mtcars)**: Funkcija attach() omogućava lakši pristup varijablama unutar skupa podataka. Bez nje, morali bismo svaki put navoditi mtcars\$wt ili mtcars\$mpg. Korištenjem attach(mtcars), možemo direktno koristiti wt i mpg varijable.

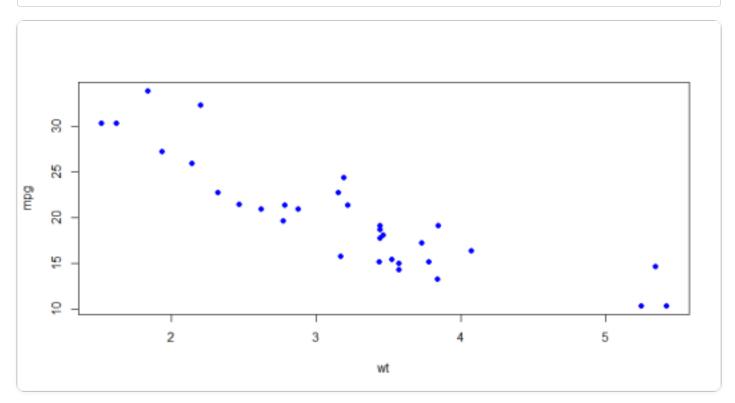
- 2. **plot(wt, mpg)**: Ova funkcija stvara osnovni X-Y graf, gdje je wt (težina vozila) *na X-osi*, a mpg (potrošnja goriva) na *Y-osi*.
- 3. **abline(Im(mpg ~ wt))**: Funkcija abline() dodaje pravac na graf. U ovom slučaju, koristi se linearni model (Im()) koji prikazuje regresijsku liniju između težine vozila i potrošnje goriva. Model je izražen kao (\text{MPG} = a + b \cdot \text{WT}), gdje su (a) i (b) koeficijenti koji se izračunavaju pomoću funkcije Im(). To omogućuje vizualno prikazivanje trenda ili povezanosti između tih dviju varijabli.
- 4. **title("Regresija: MPG = a + b * WT")**: Dodaje naslov grafu koji jasno označava o čemu se radi. U ovom slučaju, to je linearna regresija između potrošnje goriva i težine vozila.

Dodatne opcije i prilagodbe

Funkcija plot() omogućuje brojne prilagodbe kako bi graf bio jasniji i vizualno pregledniji:

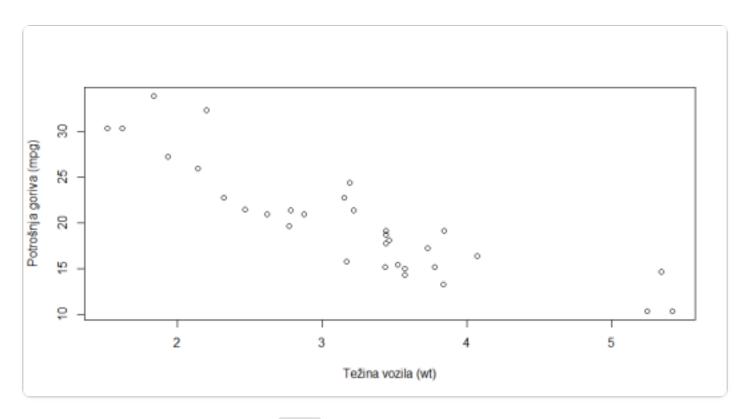
• Podešavanje boja i oznaka: Možemo koristiti argumente col (boja) i pch (simboli točaka)

```
plot(wt, mpg, col="blue", pch=16)
```



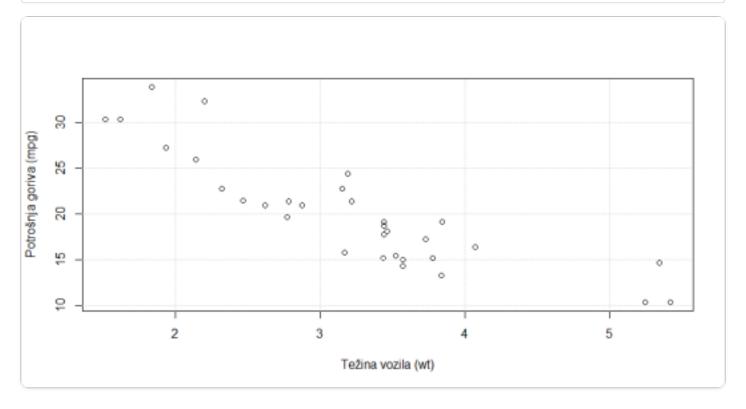
• Prilagodba osovina: Argumenti poput xlab i ylab omogućuju postavljanje oznaka za X i Y osi

```
plot(wt, mpg, xlab="Težina vozila (wt)", ylab="Potrošnja goriva (mpg)")
```



• **Dodavanje grid linija**: Funkcija grid() može se koristiti za dodavanje mreže na grafu, što može pomoći u preciznijem čitanju podataka

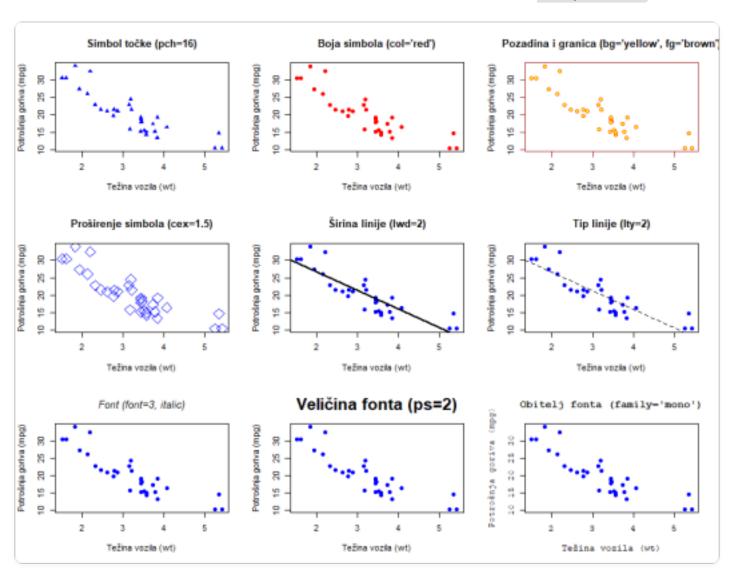
```
plot(wt, mpg)
grid()
```



Parametri grafova, linija i teksta:

Parametar	Opis	Primjer	Korištenje	
xlab, ylab	Naziv osi x i y	xlab="Težina", ylab="Potrošnja goriva"	plot(wt, mpg, pch=16, col="blue", xlab="Težina", ylab="Potrošnja goriva")	
pch = _ 1 ○ 6 ▽ 11☎ 16 ● 21 ◎ 2 △ 7 ⋈ 12 ⊞ 17 ♣ 22 Ⅲ 3 + 8 * 13 ⋈ 18 ● 23 ◆ 4 × 9 ◆ 14 ⋈ 19 ● 24 △ 5 ◇ 10 ⊕ 15 ■ 20 ● 25 ▽	Simbol za iscrtavanje točaka na grafu	pch=17 (trokutasti simbol)	plot(wt, mpg, pch=17, col="blue")	
col	Boja simbola ili linije na grafu	col="red" (crvena boja)	plot(wt, mpg, pch=16, col="red")	
bg, fg	Boje pozadine (ispune) i prednje strane (granice) objekta, radi samo za simbole 21-25	bg="yellow", fg="brown" (pozadina žuta, granica smeđa)	plot(wt, mpg, pch=21, col="orangered", bg="yellow", fg="brown")	
cex	Proširenje veličine simbola ili teksta	cex=2 (dvostruko veći simbol)	plot(wt, mpg, pch=5, col="blue", cex=2)	
lwd	Širina linije	lwd=2 (dvostruko šira linija)	plot(wt, mpg, pch=16, col="blue"); abline(lm(mpg ~ wt), lwd=2)	
Line style (lty) 0 "MANA"	Tip linije	lty=2 (crtasta linija)	plot(wt, mpg, pch=16, col="blue"); abline(lm(mpg ~ wt), lty=2)	
font (1=plain, 2=bold, 3=italic, 4=bold italic,	Tip fonta za tekst	font=3 (italic)	plot(wt, mpg, pch=16, col="blue", main="Font	

Parametar	Opis	Primjer	Korištenje
5=symbol)			(font=3, italic)", font.main=3)
ps	Veličina fonta u točkama	Veličina teksta = ps * cex	plot(wt, mpg, pch=16, col="blue", main="Veličina fonta (ps=2)", cex.main=2)
family	Obitelj fonta	family="mono"	plot(wt, mpg, pch=16, col="blue", main="Obitelj fonta (family='mono')", family="mono")



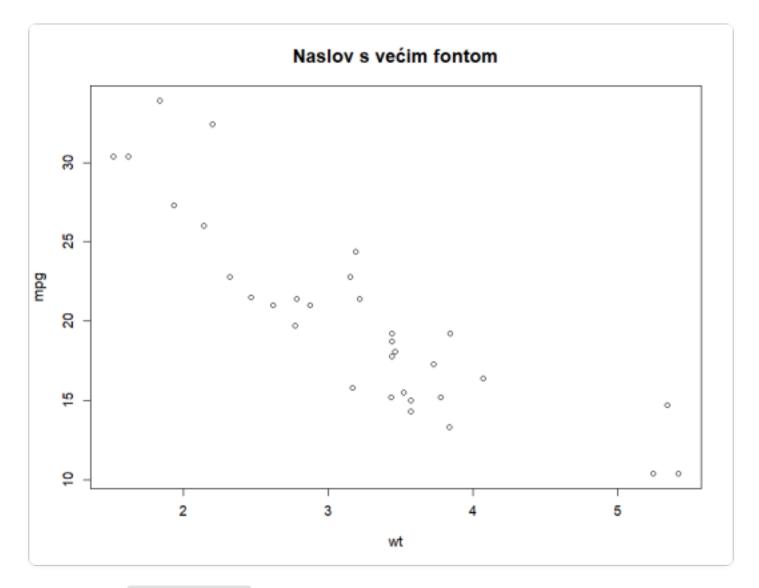
brown4	darkorange4	gray	gray57	hotpink3	lightsalmon4	navajowhite1	plum3	slategray3	antiquewhite
brown3	darkorange3	goldenrod4	gray56	hotpink2	lightsalmon3	navajowhite	plum2	slategray2	aliceblue
brown2	darkorange2	goldenrod3	gray55	hotpink1	lightsalmon2	moccasin	plum1	slategray1	white
brownf	darkorange1	goldenrod2	gray54	hotpink	lightsalmon1	mistyrose4	plum	slategray	yellowgreen
	darkorange	goldenrod1	gray53	honeydew4	lightsalmon	mistyrose3	pink4		yellow4
	darkolivegreen4	goldenrod	gray52	honeydew3	lightpink4	mistyrose2	pink3	slateblue3	yellow3
blue4	darkolivegreen3	gol64	gray51	honeydew2	lightpink3	mistyrose1	pink2	slateblue2	yellow2
blue3	darkolivegreen2	gold3	gray50	honeydew1	lightpink2	mistyrose	pink1	slateblue1	yellow1
blue2	darkolivegreen1	gold2	gray45	honeydew	lightpink1	mintoream	pink:	slateblue	yellow
blue1	darkolivegreen	gold1	gray48	greenyellow	lightpink	midnightblue	peru	skyblue4	whitesmoke
blus	darkmagenta	gold	gray47	green4	lightgrey	mediumvioletred	peachpuff4	skythue3	wheel4
olanchedalmond	darkkhaki	ghostwhite	gray46	green3		nediumturguoise	peachpuff3	skyblue2	wheat3
					lightgreen i				
black	darkgrey	gamsboro	gray45	green2		ediumspringgree	n peachpuft2	skyblue1	wheat2
bisque4	dangreen	torestgreen	gray44		htgoldenrodyelio		peachpuff1	skyblue	wheat1
bisque3	darkgray	floralwhite	gray43	green	lightgoidenrod4	nediumseagreen	peachpuff	sienna-i	wheat
bisque2	darkgoidenrod4	firelatick4	gray42	gray100	lightgoiderrod3	mediumpurple4	papayawhip	Cannels	violetred4
bisque1	darkgoldenrod3			gray99	lightgoiderrod2	mediumpurple3		sienna2	
bisque	darkgoldenrod2	firebrick2		gray98	lightgoiderrod1	mediumpurple2	palevioletred3	sienna1	violetred2
beige	darkgoldenrod1	firebrick1		gray97	lightgoldenrod	mediumpurplef	palevioletred2	sierma	violetred1
azure4	darkgoldenrod	firebrick	gray38	gray96	lightcyan4	mediumpurple	palevioletred1	seashel4	violetred
azure3	darkoyan	dodgerblue4	gray37	gray95	lightcyan3	mediumorchid4	palevioletred	seashel3	violet
azure2	darkblue	dodgerblue3	gray36	gray94	lightcyan2	mediumorchid3	paleturquoise4	seashelt2	turquoise4
azure1	cyan4	dodgerblue2	gray35	gray93	lightcyan1	mediumorchid2	paleturguoise3	seashell1	turquoise3
azure	cyan3	dodgerblue1	gray34	gray92	lightoyan	mediumorchid1	paleturquoise2	seashell	turquoise2
aquamarine4	cyan2	dodgerblue	gray33	gray91	lightcoral	mediumorchid	paleturquoise1	seagreené	turquoise1
aguamarine3					lightblue4		paleturquoise		turquoise
	cyan1	dingrey	gray32	gray90		mediumblue		seagreen3	
aquamarine2	cyan	dirigray	gray31	gray89	lightblue3 m	ediumaquamarin		seagreen2	Iomato4
aquamarine 1	comsik4	deepskyblue4	gray30	grayee	lightblue2	marcon4	palegreen3	seagreen1	Iomato3
aquatrarine	comsik3	deepskyblus3	gray29	gray87	lightblue1	maroon3	palegreen2	seagreen	tomato2
antiquewhite4	oomsik2	deepskyblus2		gray86	lightblue	maroon2	palegreen1	sandybrown	tomato1
Estidosupitres	comsikt	deepskyblue1		gray85	lemonchiffon4	marcon1	palegreen	salmon4	tomato
antiquewhite2	comsilk	deepskyblue		gray84	lemonchiffon3		palegolderrod	salmon3	thistle4
antiquewhite1	comflowerblue			gray83	lemonchiffon2		orchid4	salmon2	thistle3
antiquewhite	coral4		gray24	gray82	lemonchiffon1	magenta3	orchid3	salmont	thistie2
aliceblue	coral3	deeppink2	gray23	gray81	lemonchiffon	magenta2	orchid2	salmon	thistle1
white	corat2	deeppink1	gray22	gray80	lawngreen	magenta1	orchid1	saddlebrown	thistie
bisque3	coral1	deeppink	gray21	gray79	lavenderblush4	maganta	orchid	royalblue4	tand
bisque2	coral	darkviolet	gray20	gray78	lavenderblush3	linen		royalblue3	tan3
							orangered4		
bisque1	chocolate4	darkturquoise	gray19	gray77	lavenderblush2	limegreen	orangered3	royalblue2	tan2
bisque	chocolate3	darkslategrey	gray18	gray76	lavenderblush1	lightyelow4	orangered2	royalblue1	tan1
beige	chocolate2	darkslategray4	gray17	gray75	lavenderblush	lightyellow3	orangered1	royaldiue	tan
azure4	chocolate1	darkslategray3	gray16	gray74	lavender	lightyellow2	orangered	rosybrown4	steelblue4
azure3	chocolate	darkslategray2		gray73	khaki4	lightyellow1	orange4	rosybrown3	steelblue3
azure2	chartreuse4	darkslategray1		gray72	khaki3	lightyellow	orange3	rosybrown2	steelblue2
azure1	chartreuse3			gray71	khaki2	lightsteelblue4	orange2	rosybrown1	steelblue1
azure	chartreuse2			gray70	khaki1	lightsteelblue3	orange1	rosybrown	steeblue
aguamarine4	chartreuse1	darkseagreen4	gray11	gray69	Mhaki	lightsteelblue2	orange	red4	springgreen
aquamarine3	chartreuse	darkseagreen3	gray10	gray68	livory4	lightsteelblue1	olivedrab4		springgreen
aquamarine2	cadetblue4	darkseagreen2	gray9	gray67	ivory3	lightsteelblue	olivedrab3	red2	springgreen
aguamarine1	cadetblue3	darkseagreen1	gray6	gray66	lvory2	lightslategrey	olivedrab2	red1	springgreen
aquamarine	cadetblue2	darkseagreen	gray7	gray65	ivory1	lightslategray	olivedrab1		springgreen
								ramina	
antiques/hite4	cedetblue1	darksalmon	gray6	gray64	Ivary	lightslateblue	olivedrab	purple4	snow4
antiquewhite3	cadefblue	darkred	gray5	gray63	indianred4	lightskyblue4	oldlace	purple3	snow3
artiquewhite2	burlywood4	darkorchid4	gray4	gray62	indianred3	lightskyblue3	navyblue	purple2	snow2
antiquewhite1	burlywood3	darkorchid3	gray3	gray61	indianned2	lightskyblue2	navy	purple1	snow1
antiquewhite	burlywood2	darkorchid2	gray2	gray60	indianred1	lightskyblue1	navajowhite4	purple	snow
aliceblue	burlywood1	darkorchid1	gray1	gray59	indianred	lightskyblue	navajowhite3	powderblue	slategrey
	burlywood	darkorchid	gray0	gray58	hotpink4	lightseagreen	navajowhite2	plumé	sistegray4

Slova i margine

U R-u, slova i margine na grafu mogu se prilagoditi pomoću nekoliko parametara. Parametri koji kontroliraju veličinu fonta, poziciju teksta, kao i margine između grafova, omogućuju precizno oblikovanje vizualnih prikaza.

Parametar	Opis	Primjer	
cex	Veličina teksta na grafu.	plot(wt, mpg, cex=1.5)	
cex.axis	Veličina teksta na osovinama (X i Y os).	plot(wt, mpg, cex.axis=1.2)	
cex.lab	Veličina teksta na oznakama osovina (X i Y).	plot(wt, mpg, cex.lab=1.3)	
cex.main	Veličina teksta naslova grafa.	plot(wt, mpg, cex.main=2)	
mar	Parametar koji kontrolira margine oko grafa. Definira veličinu prostora između osovina i grafičkog prikaza.	par(mar=c(5, 4, 4, 2) + 0.1); plot(wt, mpg)	

```
par(mar=c(5, 4, 4, 2) + 0.1) # Margine: (donja, lijeva, gornja, desna)
plot(wt, mpg, cex.main=1.5, cex.lab=1.2, cex.axis=1.1, main="Naslov s većim fontom")
```



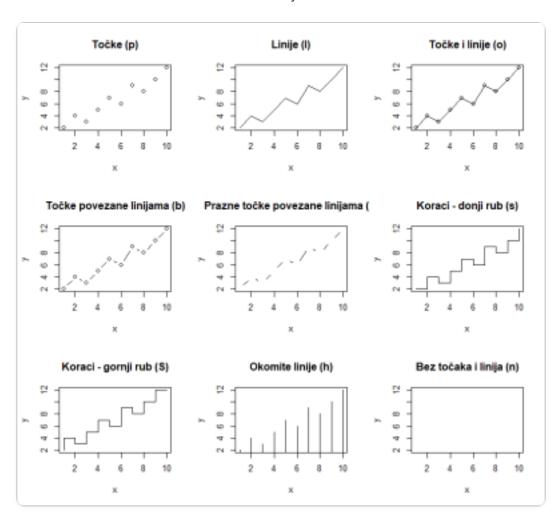
Linijski prikaz lines(x, y, type=)

Funkcija **lines(x, y, type=)** u R-u koristi se za dodavanje linija na postojeći graf izrađen pomoću plot(). Ova funkcija je korisna kada želimo prikazati više serija podataka na istom grafu ili naglasiti određene trendove.

- **x** vrijednosti na x-osi
- **y** vrijednosti na y-osi
- **type=** određuje stil crte, npr. "I" za linije, "b" za točke i linije itd.

```
x <- 1:10
y <- c(2, 4, 3, 5, 7, 6, 9, 8, 10, 12)
plot(x, y, type = "b", main = "Točke povezane linijama (b)")</pre>
```

Grafički parametar	Opis
р	Točke (points)
1	Linije (lines)
0	Preklopljene točke i linije (overplotted)
b	Točke povezane linijama
С	Prazne točke povezane linijama
S	Koraci (stair steps)
S	Koraci s drugačijim stilom (stair steps)
h	Okomite linije (histogram-like)
n	Bez točaka i linija



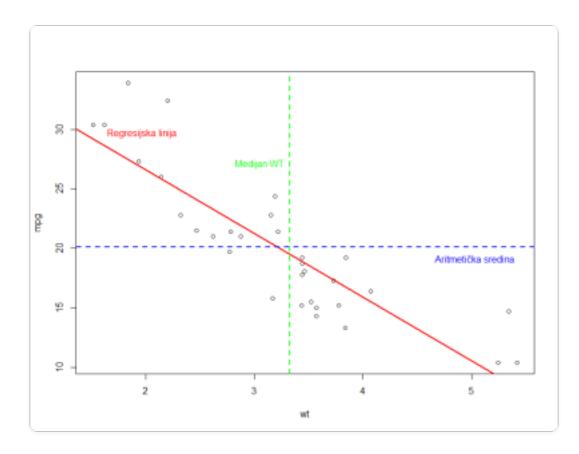
Abline

Funkcija abline() koristi se za dodavanje linija na grafove. Možemo dodati horizontalne, vertikalne linije, kao i linije temeljene na regresijskim modelima.

Funkcija	Opis	Argumenti	Primjer
h	Dodavanje horizontalne linije na	y (y-koordinata),	abline(h = 5, col = "red",
	određenu y-koordinatu	col, lty, lwd	lty = 2, lwd = 2)
v	Dodavanje vertikalne linije na	x (x-koordinata),	abline(v = 3, col = "blue",
	određenu x-koordinatu	col, lty, lwd	lty = 1, lwd = 1)
lm()	Dodavanje linije temeljene na	Model (npr. lm()),	abline(lm(mpg ~ wt), col
	linearnom modelu (regresija)	col, lty, lwd	= "green")

```
plot(wt, mpg)
abline(lm(mpg ~ wt), col="red", lwd=2) # linija linearne regresije
abline(h=mean(mpg), col="blue", lwd=2, lty=2) # linija aritmetičke sredine
abline(v=median(wt), col="green", lwd=2, lty=2) # linija medijana wt

text(x=min(wt)*1.3, y=max(mpg)*0.9, labels="Regresijska linija", col="red", pos=1)
text(x=max(wt)*0.85, y=mean(mpg)*0.95, labels="Aritmetička sredina", col="blue", pos=4)
text(x=median(wt), y=max(mpg)*0.8, labels="Medijan WT", col="green", pos=2)
```



Lines

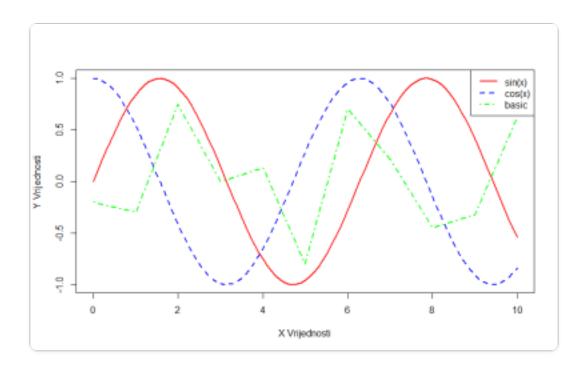
Funkcija lines() koristi se za dodavanje linija na postojeći graf u R-u. Za razliku od plot(), koji kreira novi graf, lines() samo nadodaje linije na postojeći prikaz. Može se koristiti za povezivanje točaka u skupu podataka, crtanje funkcija ili dodavanje trendova.

```
x <- seq(0, 10, by=0.1)
x2 <- seq(0, 10, by=1)

y1 <- sin(x)
y2 <- cos(x)
y3 <- c(-0.2, -0.3, 0.75, 0, 0.13, -0.8, 0.7, 0.21, -0.45, -0.32, 0.625)

plot(x, y1, type="1", col="red", lwd=2, ylim=c(-1,1), xlab="X Vrijednosti", ylab="Y Vrijednosti")
lines(x, y2, col="blue", lwd=2, lty=2)
lines(x2, y3, col="green", lwd=2, lty=4)

legend("topright", legend=c("sin(x)", "cos(x)", "rigid"), col=c("red", "blue", "green"), lty=c(1,2,4), lwd=2)</pre>
```



Ostali elementi

U R-u, razni elementi mogu se dodavati na grafove kako bi poboljšali čitljivost i interpretaciju podataka. To uključuje naslove, oznake osi, legende i grafičke oblike.

Funkcija	Opis	Argumenti	Primjer
title()	Dodaje naslov, podnaslov ili oznake osi	main, sub, xlab, ylab, col.main, col.sub itd.	title(main="Naslov", sub="Podnaslov", col.main="red")
mtext()	Dodaje tekst izvan granica grafičkog prikaza	text, side (1=bottom, 2=left, 3=top, 4=right), line, col, at	mtext("Dodatni tekst", side=3, line=2, col="blue")
legend()	Dodaje legendu unutar ili izvan grafičkog prikaza	x, y, legend, col, lty, pch, bty	legend("topright", legend=c("Linija A", "Linija B"), col=c("red", "blue"), lty=1:2, cex=0.8, bty="n")
text()	Dodaje prilagođeni tekst unutar grafičkog prostora	x, y, labels, col, pos, cex	text(5, 0.5, "Ovdje je oznaka", col="green", pos=4)
grid()	Dodaje mrežu za lakše očitavanje vrijednosti	col, Ity, Iwd	grid(col="gray", lty=2, lwd=0.5)
rect()	Ističe određeni dio grafa bojom u	xleft, ybottom, xright, ytop, col, border	rect(2, -1, 4, 1, col=rgb(1,0,0,0.2), border=NA)

Funkcija	Opis	Argumenti	Primjer
	pozadini		
segments()	Dodaje referentne crte	x0, y0, x1, y1, col, lty, lwd	segments(2, -1, 2, 1, col="black", lty=2, lwd=2)
polygon()	Iscrtava područje između određenih točaka	x, y, col, border	polygon(c(2,3,3,2), c(-1,-1,1,1), col=rgb(0,0,1,0.3), border=NA)
points()	Dodaje specifične oblike na pojedine točke	x, y, pch, col, cex	points(5, 0, pch=19, col="red", cex=2)
arrows()	Dodaje strelice za označavanje smjera	x0, y0, x1, y1, col, lty, length	arrows(2,0,3,0.5, col="red", lwd=2, length=0.1)
par(mfrow)	Omogućuje prikaz više grafova u istom prozoru	mfrow=c(n,m)	par(mfrow=c(2,2)); plot(x,y); hist(x); boxplot(y); barplot(y)

Primjer 1:

```
x <- seq(0, 10, by=0.1)
y <- sin(x)
z <- cos(x)

par(mfrow=c(1, 2))

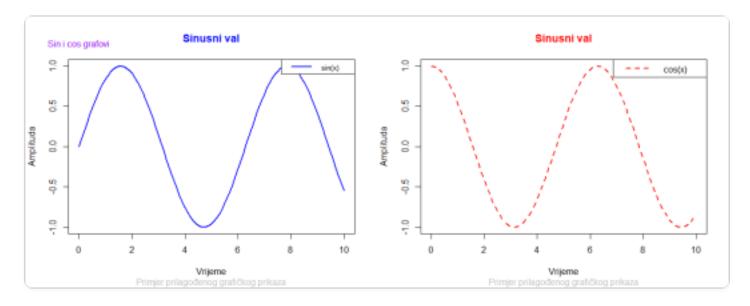
plot(x, y, type="l", col="blue", lwd=2, xlab="Vrijeme", ylab="Amplituda")

title(main="Sinusni val", sub="Primjer prilagođenog grafičkog prikaza",
col.main="blue", col.sub="gray")
legend("topright", legend=c("sin(x)"), col="blue", lty=1, cex=0.8, bty="n")

mtext("Sin i col grafovi", line=1, col="purple", at=c(0,0))

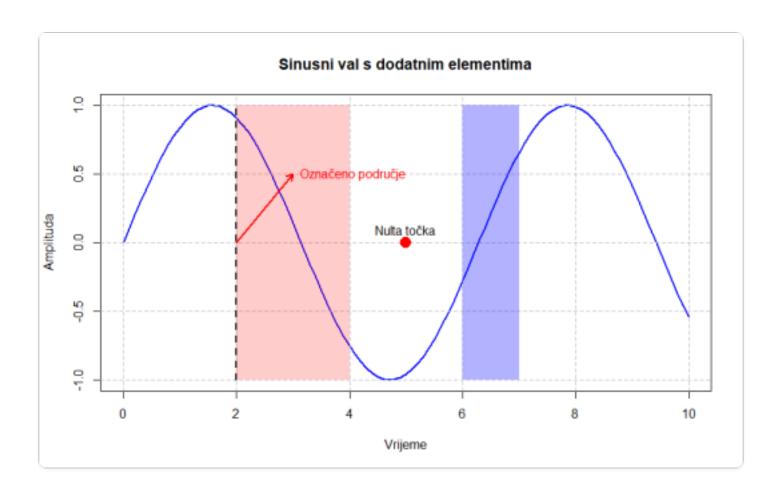
plot(x, z, type="l", col="red", lwd=2, lty=2, xlab="Vrijeme", ylab="Amplituda")

title(main="Sinusni val", sub="Primjer prilagođenog grafičkog prikaza", col.main="red",
col.sub="gray")
legend("topright", legend=c("cos(x)"), col="red", lty=2, cex=0.8, bty="n")</pre>
```



Primjer 2:

```
x \leftarrow seq(0, 10, by=0.1)
y \leftarrow sin(x)
plot(x, y, type="l", col="blue", lwd=2, xlab="Vrijeme", ylab="Amplituda")
title(main="Sinusni val s dodatnim elementima")
# Dodavanje mreže
grid(col="gray", lty=2, lwd=0.5)
# Dodavanje pravokutnika
rect(2, -1, 4, 1, col=rgb(1,0,0,0.2), border=NA)
# Dodavanje referentne linije
segments(2, -1, 2, 1, col="black", lty=2, lwd=2)
# Dodavanje poligona
polygon(c(6,7,7,6), c(-1,-1,1,1), col=rgb(0,0,1,0.3), border=NA)
# Dodavanje oznaka
text(5, 0, "Nulta točka", col="black", pos=3)
text(3, 0.5, "Označeno područje", col="red", pos=4)
# Dodavanje strelice
arrows(2, 0, 3, 0.5, col="red", lwd=2, length=0.1)
# Dodavanje točke s posebnim simbolom
points(5, 0, pch=19, col="red", cex=2)
```



Vježba

Istražujemo podatke o automobilima i želimo vizualizirati odnos između težine vozila (wt) i potrošnje goriva (mpg) koristeći skup podataka mtcars. Cilj je dodati korisne vizualne elemente kako biste istaknuli važne vrijednosti i trendove.

- 1. Nacrtajte raspršeni dijagram (plot()) za wt (x-os) i mpg (y-os)
- 2. Dodajte regresijsku liniju koristeći abline(lm())
- 3. Dodajte horizontalnu liniju na prosječnu vrijednost mpg i vertikalnu liniju na medijan wt
- 4. Dodajte naslov i oznake osi koristeći title()
- 5. Dodajte legendu koja označava regresijsku liniju, prosječnu vrijednost i medijan
- 6. Koristite text() za dodavanje oznaka uz sve tri linije grafa
- 7. Dodajte mrežu (grid()) radi boljeg pregleda podataka
- 8. Ručno označite područje između prvog i trećeg kvartila pomoću rect().
- 9. Dodajte strelicu (arrows()) koja upućuje na vozilo s najvećom potrošnjom (mpg)

