Data Analysis Using R: Chapter10

罗智超 (ROKIA.ORG)

1 通过本章你将学会

- Visualizing bivariate and multivariate relationships
- Working with scatter and line plots
- Understanding correlograms
- Using mosaic and association plots

2 Scatter plots

- plot(x, y)

```
attach(mtcars)
plot(wt, mpg,
     main="Basic Scatterplot of MPG vs. Weight",
     xlab="Car Weight (lbs/1000)",
     ylab="Miles Per Gallon ", pch=19)
abline(lm(mpg ~ wt), col="red", lwd=2, lty=1)
lines(lowess(wt, mpg), col="blue", lwd=2, lty=2)
#lowess(),This function performs the computations
#for the LOWESS smoother which uses
#locally-weighted polynomial regression
```

- library(car),scatterplot()

```
library(car)
scatterplot(mpg ~ wt | cyl, data=mtcars, lwd=2,
    main="Scatter Plot of MPG vs. Weight by # Cylinders",
    xlab="Weight of Car (lbs/1000)",
    ylab="Miles Per Gallon", id.method="identify",
    legend.plot=TRUE, labels=row.names(mtcars),
    boxplots="xy")
```

3 Scatter plot matrices

- pairs()
- library(car):scatterplotMatrix()
- library(gclus):cpairs()
- library(GGally):ggpairs()

4

```
pairs(~ mpg + disp + drat + wt, data=mtcars,
      main="Basic Scatterplot Matrix")

library(car)
scatterplotMatrix(~ mpg + disp + drat + wt,
                  data=mtcars, spread=FALSE,
                  lty.smooth=2,
                  main="Scatterplot Matrix via car package")

scatterplotMatrix(~ mpg + disp + drat + wt | cyl,
                  data=mtcars, spread=FALSE,
                  main="Scatterplot Matrix via car package",
```

```
                  diagonal="histogram")

cor(mtcars[c("mpg", "wt", "disp", "drat")])
```

5

```
library(gclus)
mydata <- mtcars[c(1,3,5,6)]
mydata.corr <- abs(cor(mydata))
mycolors <- dmat.color(mydata.corr)
myorder <- order.single(mydata.corr)
cpairs(mydata,
    myorder,
    panel.colors=mycolors,
    gap=.5,
    main="Variables Ordered and Colored by Correlation"
)
```

6 High-density scatter plots

- smoothScatter()
- library(hexbin):hexbin()
- library(IDPmisc):iplot()

7

```
# high density scatterplots

set.seed(1234)
```

```
n <- 10000
c1 <- matrix(rnorm(n, mean=0, sd=.5), ncol=2)
c2 <- matrix(rnorm(n, mean=3, sd=2), ncol=2)
mydata <- rbind(c1, c2)
mydata <- as.data.frame(mydata)
names(mydata) <- c("x", "y")

with(mydata,
    plot(x, y, pch=19,
         main="Scatter Plot with 10000 Observations"))

with(mydata,
    smoothScatter(x, y,
                  main="Scatterplot colored by Smoothed Densities"))
```

8

```
library(hexbin)
with(mydata, {
    bin <- hexbin(x, y, xbins=50)
    plot(bin, main="Hexagonal Binning with 10,000 Observations")
    })
#Unable to install hexbin package in R3.1
```

9

```
library(IDPmisc)
with(mydata,
    iplot(x, y, main="Image Scatter Plot with Color Indicating Density"))
```

10 3D scatter plots

- library(scatterplot3d):scatterplot3d()
- library(rgl):plot3d(x, y, z)
- library(Rcmdr):scatter3d()

11

```
dev.off()
library(scatterplot3d)
attach(mtcars)
scatterplot3d(wt, disp, mpg,
     main="Basic 3D Scatterplot")

scatterplot3d(wt, disp, mpg,
    pch=16,
    highlight.3d=TRUE,
    type="h",
    main="3D Scatterplot with Vertical Lines")

s3d <-scatterplot3d(wt, disp, mpg,
    pch=16,
    highlight.3d=TRUE,
    type="h",
    main="3D Scatter Plot with Verical Lines and Regression Plane")
fit <- lm(mpg ~ wt+disp)
s3d$plane3d(fit)
detach(mtcars)
```

12

```
library(rgl)
attach(mtcars)
plot3d(wt, disp, mpg, col="red", size=5)

# alternative
rgl.open()
library(RcmdrMisc)
attach(mtcars)
scatter3d(wt, disp, mpg)
```

13 Bubble plots

- symbols(x, y, circle=radius)

14

```
attach(mtcars)
r <- sqrt(disp/pi)
symbols(wt, mpg, r, inches=0.30,
        fg="white", bg="lightblue",
        main="Bubble Plot with point size proportional to displacement",
        ylab="Miles Per Gallon",
        xlab="Weight of Car (lbs/1000)")
text(wt, mpg, rownames(mtcars), cex=0.6)
detach(mtcars)
```

15 Line charts

- plot(x, y, type=)

- lines(x, y, type=)

16

```
dev.off()
t1 <- subset(Orange, Tree==1)
plot(t1$age, t1$circumference,
    xlab="Age (days)",
    ylab="Circumference (mm)",
    main="Orange Tree 1 Growth")
plot(t1$age, t1$circumference,
    xlab="Age (days)",
    ylab="Circumference (mm)",
    main="Orange Tree 1 Growth",
    type="b")
```

17

Type	What is plottedp Points
p	Point Only
l	Lines only
o	Over-plotted points (that is, lines overlaid on top of points)
b, c	Points (empty if c) joined by lines
s,S	Stair steps
h	Histogram-line vertical lines
n	Doesn??t produce any points or lines (used to set up the axes for later commands)

18

```
# type= options in the plot() and lines() functions

x <- c(1:5)
y <- c(1:5)
par(mfrow=c(2,4))
types <- c("p", "l", "o", "b", "c", "s", "S", "h")
for (i in types){
    plottitle <- paste("type=", i)
    plot(x,y,type=i, col="red", lwd=2, cex=1, main=plottitle)
}
```

19 Correlograms

- cor()
- library(corrgram):corrgram()

20

```
options(digits=2)
cor(mtcars)
```

21

```
library(corrgram)
corrgram(mtcars, order=TRUE, lower.panel=panel.shade,
    upper.panel=panel.pie, text.panel=panel.txt,
    main="Correlogram of mtcar intercorrelations")
```

```
corrgram(mtcars, order=TRUE, lower.panel=panel.ellipse,
    upper.panel=panel.pts, text.panel=panel.txt,
    diag.panel=panel.minmax,
    main="Correlogram of mtcar data using scatterplots and ellipses")

corrgram(mtcars, lower.panel=panel.shade,
    upper.panel=NULL, text.panel=panel.txt,
    main="Car Mileage Data (unsorted)")

col.corrgram <- function(ncol){
    colorRampPalette(c("darkgoldenrod4", "burlywood1",
                                    "darkkhaki", "darkgreen"))(ncol)}
corrgram(mtcars, order=TRUE, lower.panel=panel.shade,
    upper.panel=panel.pie, text.panel=panel.txt,
    main="A Corrgram (or Horse) of a Different Color")
```

22 Mosaic plots

- When plot with more than two categorical variables.
- The frequencies in amultidimensional contingency table are represented by nested rectangular regions that are proportional to their cell frequency.
- mosaicplot()
- library(vcd):mosaic()

23

```
ftable(Titanic)
library(vcd)
mosaic(Titanic, shade=TRUE, legend=TRUE)
# Michael Friendly http://www.datavis.ca
```