**主成分分析**

> setwd("C:/Users/lordw/Desktop")

> student=read.table("student.txt",header=TRUE)

> student.pca = princomp(student, cor = TRUE)

> summary(student.pca, loadings = TRUE)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4

Standard deviation 1.8817805 0.55980636 0.28179594 0.25711844

Proportion of Variance 0.8852745 0.07834579 0.01985224 0.01652747

Cumulative Proportion 0.8852745 0.96362029 0.98347253 1.00000000

Loadings:

Comp.1 Comp.2 Comp.3 Comp.4

X1 0.497 0.543 0.450 0.506

X2 0.515 -0.210 0.462 -0.691

X3 0.481 -0.725 -0.175 0.461

X4 0.507 0.368 -0.744 -0.232

> student.pca$scores

Comp.1 Comp.2 Comp.3 Comp.4

[1,] -0.06990950 -0.23813701 0.35509248 -0.266120139

[2,] -1.59526340 -0.71847399 -0.32813232 -0.118056646

[3,] 2.84793151 0.38956679 0.09731731 -0.279482487

[4,] -0.75996988 0.80604335 0.04945722 -0.162949298

[5,] 2.73966777 0.01718087 -0.36012615 0.358653044

[6,] -2.10583168 0.32284393 -0.18600422 -0.036456084

[7,] 1.42105591 -0.06053165 -0.21093321 -0.044223092

[8,] 0.82583977 -0.78102576 0.27557798 0.057288572

[9,] 0.93464402 -0.58469242 0.08814136 0.181037746

[10,] -2.36463820 -0.36532199 -0.08840476 0.045520127

[11,] -2.83741916 0.34875841 -0.03310423 -0.031146930

[12,] 2.60851224 0.21278728 0.33398037 0.210157574

[13,] 2.44253342 -0.16769496 0.46918095 -0.162987830

[14,] -1.86630669 0.05021384 -0.37720280 -0.358821916

[15,] -2.81347421 -0.31790107 0.03291329 -0.222035112

[16,] -0.06392983 0.20718448 -0.04334340 0.703533624

[17,] 1.55561022 -1.70439674 0.33126406 0.007551879

[18,] -1.07392251 -0.06763418 -0.02283648 0.048606680

[19,] 2.52174212 0.97274301 -0.12164633 -0.390667991

[20,] 2.14072377 0.02217881 -0.37410972 0.129548960

[21,] 0.79624422 0.16307887 -0.12781270 -0.294140762

[22,] -0.28708321 -0.35744666 0.03962116 0.080991989

[23,] 0.25151075 1.25555188 0.55617325 0.109068939

[24,] -2.05706032 0.78894494 0.26552109 0.388088643

[25,] 3.08596855 -0.05775318 -0.62110421 -0.218939612

[26,] 0.16367555 0.04317932 -0.24481850 0.560248997

[27,] -1.37265053 0.02220972 0.23378320 -0.257399715

[28,] -2.16097778 0.13733233 -0.35589739 0.093123683

[29,] -2.40434827 -0.48613137 0.16154441 -0.007914021

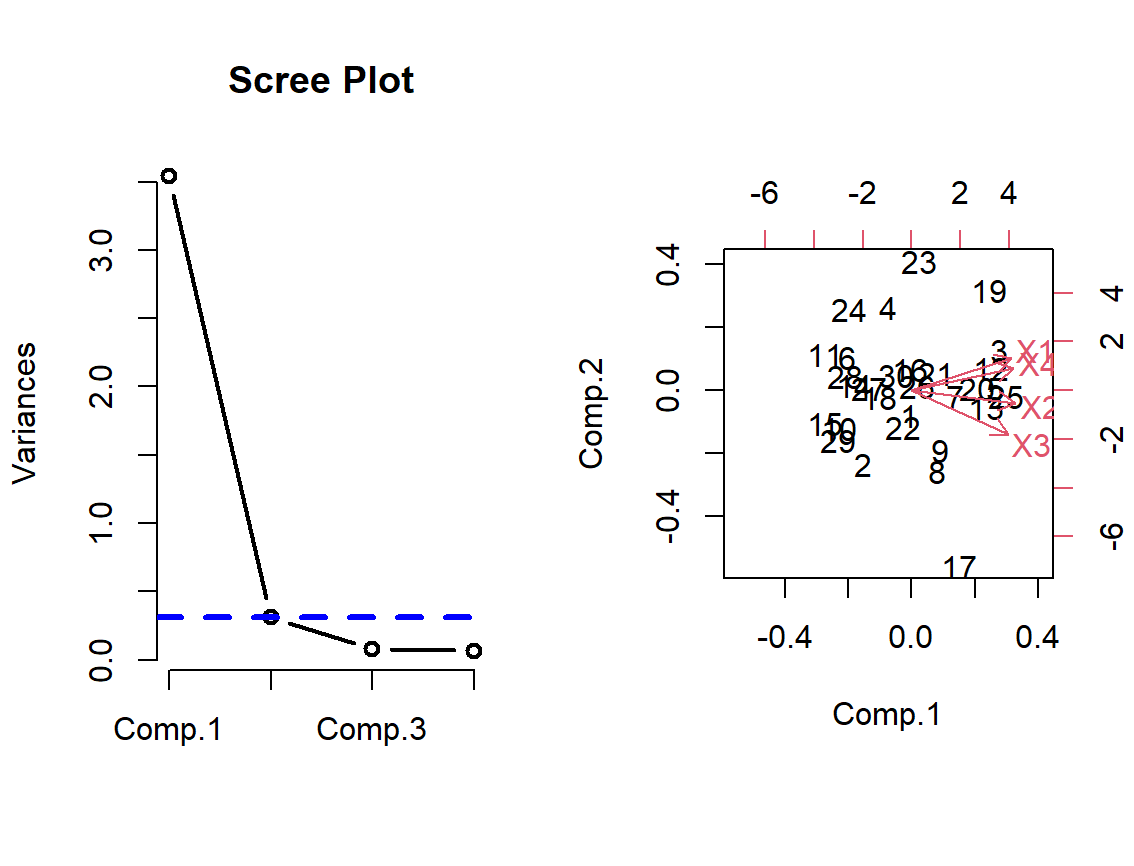
[30,] -0.50287468 0.14734317 0.20590831 -0.122078819

> par(mfrow = c(1, 2))

> screeplot(student.pca, type = "lines", main = "Scree Plot", lwd = 2)

> abline(h = 0.5598064^2, lty = 2, col = "blue", lwd = 3)

> biplot(student.pca)



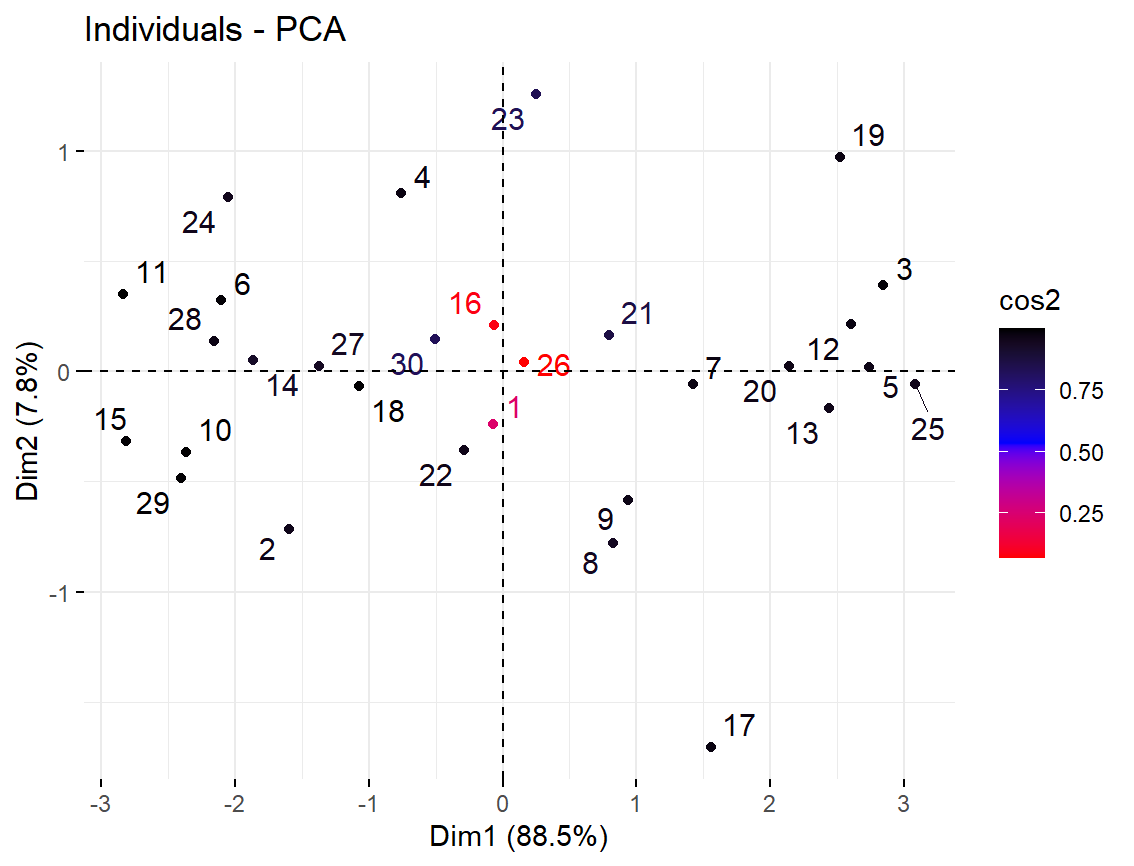
> install.packages("factoextra")

> library(factoextra)

> fviz\_pca\_ind(student.pca, col.ind = "cos2",

+ gradient.cols = c("red", "blue", "black"),

+ repel = TRUE)



**聚类分析**

> Consumer = read.csv("consumer2018.csv", header = TRUE, fileEncoding = "GBK")

> Cosumer = Consumer[, -c(2, 3)]

> data.mat = as.matrix(Cosumer[, 2:9]); rownames(data.mat)=Cosumer[, 1]

*#### 编写函数计算总离差平方和、类内离差平方和和R^2统计量*

> tss.cal = function(x) sum(scale(x, scale = F)^2)

> wss.cal = function(x, clst) sum(by(x, INDICES=clst,FUN = tss.cal))

> rsq.cal = function(x, clst) (tss.cal(x)-wss.cal(x, clst))/tss.cal(x)

> rsq = vector("numeric", 14)

> res = hclust(dist(scale(data.mat)), method = "ward.D")

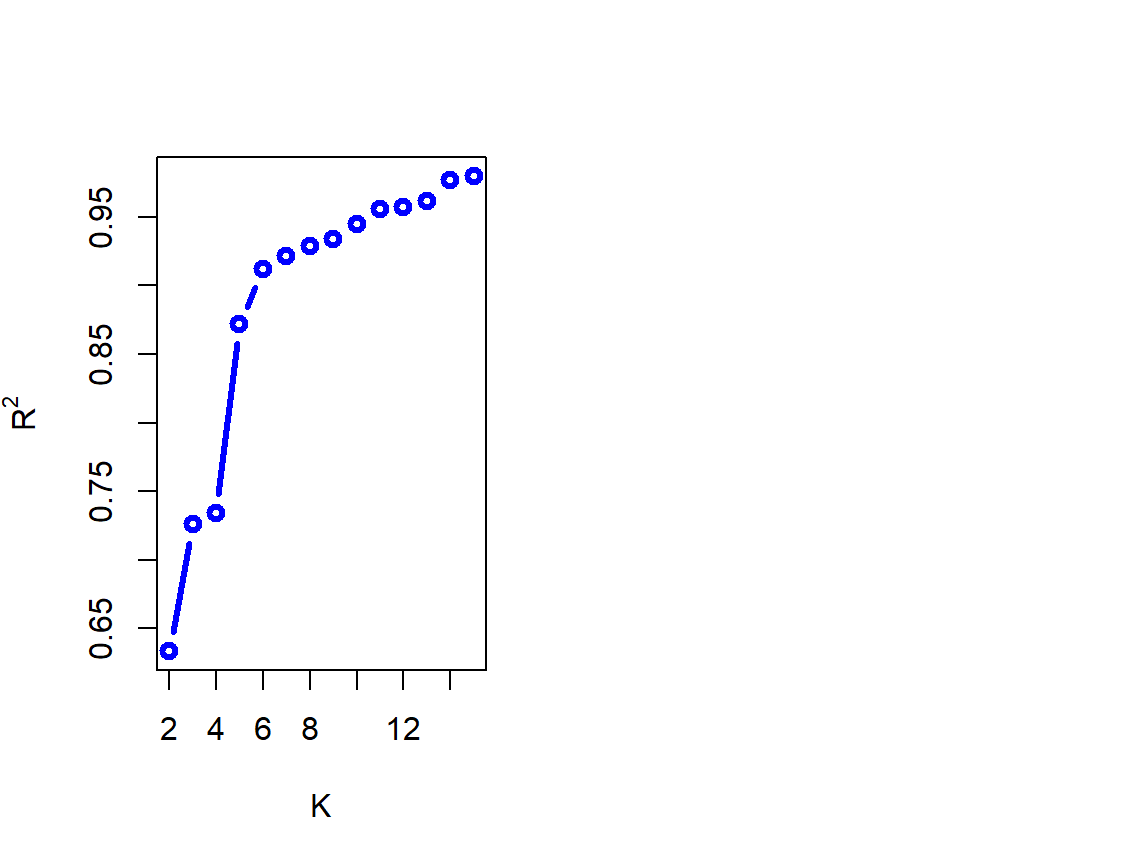
> for(ii in 2:15){

+ clst = cutree(res, ii)

+ rsq[ii-1] = rsq.cal(data.mat, clst)

+ }

> plot(2:15, rsq, type="b", xlab="K", ylab=expression(R^2), col = "blue", lwd = 3)



m = c("single","complete","median","average","centroid","ward.D")

> h = list(); reh = list(); opar = par(mfrow=c(3, 2))

> for (i in 1:length(m)){

+ h[[i]]=hclust(dist(scale(data.mat)), method = m[i])

+ plot(h[[i]], main = paste("Method: ", m[i]), hang = -1)

+ reh[[i]] = rect.hclust(h[[i]], k = 5, border = "red")

