



> jeto<-jet[order(jet[,6]),]

> jeto

FFD SPR RGF PLF SLF CAR

FH-1 82 1.468 3.30 0.166 0.10 no

FJ-1 89 1.605 3.64 0.154 0.10 no

F9F-2 107 2.054 4.72 0.275 1.10 no

F3D-1 122 1.294 3.75 0.150 0.90 no

XF10F-1 137 2.426 4.65 0.117 1.80 no

F9F-6 147 2.607 3.84 0.155 2.30 no

F4D-1 174 4.588 3.82 0.249 3.50 no

F11F-1 175 3.618 4.32 0.143 2.80 no

F3H-2 184 2.898 4.48 0.178 3.00 no

F-8A 189 0.455 4.99 0.008 2.64 no

F-4B 255 8.548 4.20 0.222 2.90 no

F-86A 101 2.168 4.87 0.177 2.90 yes

F-94A 115 2.467 4.11 0.298 1.00 yes

F-89A 127 2.183 3.97 0.000 2.40 yes

F-100A 166 4.567 4.92 0.138 3.20 yes

F-101A 177 5.855 4.53 0.172 2.50 yes

F-102A 187 3.880 5.39 0.101 3.00 yes

F-104B 194 8.088 4.50 0.251 2.70 yes

F-105B 197 6.502 5.20 0.366 2.90 yes

YF-107A 201 6.081 5.65 0.106 2.90 yes

F-106A 204 7.105 5.40 0.089 3.20 yes

F-111A 328 6.321 6.45 0.187 2.00 yes

> Mj<-jeto[,2:5]

> cov(Mj)

SPR RGF PLF SLF

SPR 5.60375912 0.790558442 0.068055061 1.284142165

RGF 0.79055844 0.566992208 -0.004315411 0.351729870

PLF 0.06805506 -0.004315411 0.007509160 -0.008956797

SLF 1.28414216 0.351729870 -0.008956797 1.006635498

Book says: "And given that the variables are on very different scales, we will standardise them to unit variance before clustering.”

> Mjb<-scale(Mj,center=FALSE,scale=TRUE)

> Mjb

SPR RGF PLF SLF

FH-1 0.31366825 0.6954522 0.8609797 0.0395801

FJ-1 0.34294111 0.7671049 0.7987402 0.0395801

F9F-2 0.43887915 0.9947074 1.4263217 0.4353811

F3D-1 0.27648959 0.7902866 0.7779937 0.3562209

XF10F-1 0.51836456 0.9799554 0.6068351 0.7124418

F9F-6 0.55703892 0.8092535 0.8039268 0.9103423

F4D-1 0.98032012 0.8050387 1.2914695 1.3853035

F11F-1 0.77305976 0.9104102 0.7416873 1.1082428

F3H-2 0.61921703 0.9441291 0.9232192 1.1874030

F-8A 0.09722006 1.0516081 0.0414930 1.0449146

F-4B 1.82645519 0.8851210 1.1514306 1.1478229

F-86A 0.46323758 1.0263189 0.9180325 1.1478229

F-94A 0.52712505 0.8661542 1.5456141 0.3958010

F-89A 0.46644264 0.8366501 0.0000000 0.9499224

F-100A 0.97583304 1.0368561 0.7157542 1.2665632

F-101A 1.25104061 0.9546663 0.8920994 0.9895025

F-102A 0.82904143 1.1359053 0.5238491 1.1874030

F-104B 1.72816677 0.9483440 1.3018427 1.0686627

F-105B 1.38928540 1.0958641 1.8983046 1.1478229

YF-107A 1.29933013 1.1906985 0.5497822 1.1478229

F-106A 1.51812870 1.1380128 0.4616096 1.2665632

F-111A 1.35061105 1.3592930 0.9698988 0.7916020

attr(,"scaled:scale")

SPR RGF PLF SLF

4.6801039 4.7451138 0.1928036 2.5265222

> head(Mj/Mjb)

SPR RGF PLF SLF

FH-1 4.680104 4.745114 0.1928036 2.526522

FJ-1 4.680104 4.745114 0.1928036 2.526522

F9F-2 4.680104 4.745114 0.1928036 2.526522

F3D-1 4.680104 4.745114 0.1928036 2.526522

XF10F-1 4.680104 4.745114 0.1928036 2.526522

F9F-6 4.680104 4.745114 0.1928036 2.526522

> cov(Mjb)

SPR RGF PLF SLF

SPR 0.25584002 0.03559851 0.07542056 0.10860117

RGF 0.03559851 0.02518163 -0.00471694 0.02933861

PLF 0.07542056 -0.00471694 0.20200447 -0.01838715

SLF 0.10860117 0.02933861 -0.01838715 0.15769794

Forget that!

> Mjs<-scale(Mj,center=TRUE,scale=TRUE)

> cov(Mjs)

SPR RGF PLF SLF

SPR 1.0000000 0.44351234 0.33176076 0.5406762

RGF 0.4435123 1.00000000 -0.06613601 0.4655696

PLF 0.3317608 -0.06613601 1.00000000 -0.1030198

SLF 0.5406762 0.46556957 -0.10301985 1.0000000

> DMjs<-as.matrix(DMjs)

> rownames(DMjs)<-1:22

> colnames(DMjs)<-1:22

> round(DMjs,1)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

1 0.0 0.5 2.5 1.0 2.6 2.4 3.8 3.2 3.3 3.9 4.3 3.5 2.1 3.1 4.0 3.4 4.2 4.2 4.9 4.7 4.9 5.0

2 0.5 0.0 2.2 0.8 2.2 2.2 3.8 3.0 3.2 3.6 4.2 3.3 2.0 2.9 3.7 3.2 3.9 4.1 4.7 4.3 4.6 4.7

3 2.5 2.2 0.0 2.0 2.0 2.2 2.9 2.4 2.3 3.5 3.4 2.1 0.9 3.6 2.8 2.4 3.0 3.0 2.9 3.4 3.8 3.2

4 1.0 0.8 2.0 0.0 1.6 1.5 3.2 2.3 2.4 2.9 3.8 2.5 1.8 2.3 3.1 2.7 3.3 3.7 4.3 3.8 4.1 4.3

5 2.6 2.2 2.0 1.6 0.0 1.3 2.7 1.2 1.4 1.8 3.1 1.3 2.3 1.7 1.7 1.7 1.7 3.0 3.6 2.3 2.6 3.0

6 2.4 2.2 2.2 1.5 1.3 0.0 1.8 0.9 1.1 2.5 2.7 1.5 2.1 1.8 1.9 1.7 2.3 2.7 3.5 2.9 3.0 3.8

7 3.8 3.8 2.9 3.2 2.7 1.8 0.0 1.6 1.5 3.7 1.9 2.0 2.7 3.2 2.0 1.7 2.8 1.9 2.5 3.1 3.0 3.9

8 3.2 3.0 2.4 2.3 1.2 0.9 1.6 0.0 0.6 2.2 2.3 1.0 2.6 1.9 1.0 1.1 1.5 2.3 3.1 2.1 2.2 3.2

9 3.3 3.2 2.3 2.4 1.4 1.1 1.5 0.6 0.0 2.3 2.5 0.6 2.5 2.3 1.0 1.3 1.6 2.4 2.8 2.2 2.4 3.2

10 3.9 3.6 3.5 2.9 1.8 2.5 3.7 2.2 2.3 0.0 4.4 2.1 4.0 1.6 2.4 3.0 1.9 4.3 4.9 2.8 3.1 3.8

11 4.3 4.2 3.4 3.8 3.1 2.7 1.9 2.3 2.5 4.4 0.0 2.9 3.3 3.8 2.2 1.4 2.9 0.6 2.3 2.6 2.3 3.3

12 3.5 3.3 2.1 2.5 1.3 1.5 2.0 1.0 0.6 2.1 2.9 0.0 2.6 2.4 1.2 1.7 1.3 2.7 2.9 2.1 2.4 2.9

13 2.1 2.0 0.9 1.8 2.3 2.1 2.7 2.6 2.5 4.0 3.3 2.6 0.0 3.7 3.2 2.6 3.5 3.0 3.0 3.9 4.2 3.9

14 3.1 2.9 3.6 2.3 1.7 1.8 3.2 1.9 2.3 1.6 3.8 2.4 3.7 0.0 2.4 2.6 2.4 3.9 4.9 3.1 3.1 4.3

15 4.0 3.7 2.8 3.1 1.7 1.9 2.0 1.0 1.0 2.4 2.2 1.2 3.2 2.4 0.0 1.1 0.8 2.1 2.8 1.3 1.4 2.5

16 3.4 3.2 2.4 2.7 1.7 1.7 1.7 1.1 1.3 3.0 1.4 1.7 2.6 2.6 1.1 0.0 1.7 1.3 2.5 1.7 1.7 2.6

17 4.2 3.9 3.0 3.3 1.7 2.3 2.8 1.5 1.6 1.9 2.9 1.3 3.5 2.4 0.8 1.7 0.0 2.8 3.3 1.0 1.4 2.2

18 4.2 4.1 3.0 3.7 3.0 2.7 1.9 2.3 2.4 4.3 0.6 2.7 3.0 3.9 2.1 1.3 2.8 0.0 1.8 2.4 2.3 2.9

19 4.9 4.7 2.9 4.3 3.6 3.5 2.5 3.1 2.8 4.9 2.3 2.9 3.0 4.9 2.8 2.5 3.3 1.8 0.0 3.1 3.2 2.8

20 4.7 4.3 3.4 3.8 2.3 2.9 3.1 2.1 2.2 2.8 2.6 2.1 3.9 3.1 1.3 1.7 1.0 2.4 3.1 0.0 0.7 1.7

21 4.9 4.6 3.8 4.1 2.6 3.0 3.0 2.2 2.4 3.1 2.3 2.4 4.2 3.1 1.4 1.7 1.4 2.3 3.2 0.7 0.0 2.2

22 5.0 4.7 3.2 4.3 3.0 3.8 3.9 3.2 3.2 3.8 3.3 2.9 3.9 4.3 2.5 2.6 2.2 2.9 2.8 1.7 2.2 0.0

> Mjsc<-hclust(as.dist(DMjs))

> plot(Mjsc)



> prcomp(Mjs)

Standard deviations:

[1] 1.4064330 1.0778969 0.7421941 0.5560867

Rotation:

PC1 PC2 PC3 PC4

SPR 0.6011860 -0.3052264 0.1840062 -0.7152300

RGF 0.5408329 0.2456147 -0.7910625 0.1462644

PLF 0.1064118 -0.8853327 -0.1215182 0.4359998

SLF 0.5785773 0.2503921 0.5706094 0.5262670

> cutree(Mjsc,2)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

1. 1 1 1 1 1 2 1 1 1 2 1 1 1 2 2 2 2 2 2 2 2

> plot(prcomp(Mjs)$x[,1:2],type="n")

> text(prcomp(Mjs)$x[,1:2],labels=1:22,col=as.numeric(jeto[,6])+1,cex=1.2)

> plot(prcomp(Mjs)$x[,1:2],type="n")

> text(prcomp(Mjs)$x[,1:2],labels=apply(cbind(1:22,cutree(Mjsc,2)),1,paste,collapse=":"),col=as.numeric(jeto[,6])+1)









\left\{\begin{matrix} n \\ k \end{matrix}\right\} = \frac{1}{k!}\sum_{j=0}^k (-1)^{k-j}{k \choose j} j^n.

> Stir2<-function (n,k)

{S<-0

for (j in 0:k)

{S<-S+(-1)^(k-j)\*choose(k,j)\*j^n}

S/factorial(k)

}

> Stir2(15,3)

[1] 2375101

> Stir2(20,4)

[1] 45232115901

> crime

Murder Rape Robbery Assault Burglary Theft Vehicle

ME 2.0 14.8 28 102 803 2347 164

NH 2.2 21.5 24 92 755 2208 228

VT 2.0 21.8 22 103 949 2697 181

MA 3.6 29.7 193 331 1071 2189 906

RI 3.5 21.4 119 192 1294 2568 705

CT 4.6 23.8 192 205 1198 2758 447

NY 10.7 30.5 514 431 1221 2924 637

NJ 5.2 33.2 269 265 1071 2822 776

PA 5.5 25.1 152 176 735 1654 354

OH 5.5 38.6 142 235 988 2574 376

IN 6.0 25.9 90 186 887 2333 328

IL 8.9 32.4 325 434 1180 2938 628

MI 11.3 67.4 301 424 1509 3378 800

WI 3.1 20.1 73 162 783 2802 254

MN 2.5 31.8 102 148 1004 2785 288

IA 1.8 12.5 42 179 956 2801 158

MO 9.2 29.2 170 370 1136 2500 439

ND 1.0 11.6 7 32 385 2049 120

SD 4.0 17.7 16 87 554 1939 99

NE 3.1 24.6 51 184 748 2677 168

KS 4.4 32.9 80 252 1188 3008 258

DE 4.9 56.9 124 241 1042 3090 272

MD 9.0 43.6 304 476 1296 2978 545

DC 31.0 52.4 754 668 1728 4131 975

VA 7.1 26.5 106 167 813 2522 219

WV 5.9 18.9 41 99 625 1358 169

NC 8.1 26.4 88 354 1225 2423 208

SC 8.6 41.3 99 525 1340 2846 277

GA 11.2 43.9 214 319 1453 2984 430

FL 11.7 52.7 367 605 2221 4373 598

KY 6.7 23.1 83 222 824 1740 193

TN 10.4 47.0 208 274 1325 2126 544

AL 10.1 28.4 112 408 1159 2304 267

MS 11.2 25.8 65 172 1076 1845 150

AR 8.1 28.9 80 278 1030 2305 195

LA 12.8 40.1 224 482 1461 3417 442

OK 8.1 36.4 107 285 1787 3142 649

TX 13.5 51.6 240 354 2049 3987 714

MT 2.9 17.3 20 118 783 3314 215

ID 3.2 20.0 21 178 1003 2800 181

WY 5.3 21.9 22 243 817 3078 169

CO 7.0 42.3 145 329 1792 4231 486

NM 11.5 46.9 130 538 1845 3712 343

AZ 9.3 43.0 169 437 1908 4337 419

UT 3.2 25.3 59 180 915 4074 223

NV 12.6 64.9 287 354 1604 3489 478

WA 5.0 53.4 135 244 1861 4267 315

OR 6.6 51.1 206 286 1967 4163 402

CA 11.3 44.9 343 521 1696 3384 762

AK 8.6 72.7 88 401 1162 3910 604

HI 4.8 31.0 106 103 1339 3759 328



> round(mahalanobis(crime,center=colMeans(crime),cov=cov(crime)),3)

ME NH VT MA RI CT NY NJ PA OH

2.180 1.839 1.865 16.347 10.726 3.097 16.881 7.215 4.052 1.788

IN IL MI WI MN IA MO ND SD NE

1.071 4.842 8.723 2.178 2.479 3.905 2.074 4.377 3.211 2.212

KS DE MD DC VA WV NC SC GA FL

0.940 10.358 5.317 36.360 2.141 5.744 4.310 10.389 1.982 10.854

KY TN AL MS AR LA OK TX MT ID

3.442 7.309 5.592 10.628 2.419 3.574 10.723 10.145 6.726 1.694

WY CO NM AZ UT NV WA OR CA AK

5.447 4.509 8.591 5.654 11.041 8.416 9.711 9.327 4.340 24.712

HI

6.545

> Mcr<-crime[(rownames(crime)!="DC"),]

> dim(Mcr)

[1] 50 7

> round(cov(Mcr),2)

Murder Rape Robbery Assault Burglary Theft Vehicle

Murder 11.93 32.44 248.02 379.44 946.42 766.49 342.08

Rape 32.44 209.76 858.75 1307.14 4326.61 6754.55 1673.21

Robbery 248.02 858.75 11889.56 10640.22 26207.26 28043.47 17453.14

Assault 379.44 1307.14 10640.22 19373.54 41188.58 49651.78 16935.34

Burglary 946.42 4326.61 26207.26 41188.58 175895.00 238148.74 49381.81

Theft 766.49 6754.55 28043.47 49651.78 238148.74 565276.56 52829.29

Vehicle 342.08 1673.21 17453.14 16935.34 49381.81 52829.29 43997.36

> Mcrs<-scale(Mcr,scale=TRUE,center=TRUE)

> round(cov(Mcrs),2)

Murder Rape Robbery Assault Burglary Theft Vehicle

Murder 1.00 0.65 0.66 0.79 0.65 0.30 0.47

Rape 0.65 1.00 0.54 0.65 0.71 0.62 0.55

Robbery 0.66 0.54 1.00 0.70 0.57 0.34 0.76

Assault 0.79 0.65 0.70 1.00 0.71 0.47 0.58

Burglary 0.65 0.71 0.57 0.71 1.00 0.76 0.56

Theft 0.30 0.62 0.34 0.47 0.76 1.00 0.33

Vehicle 0.47 0.55 0.76 0.58 0.56 0.33 1.00



> kmeans(Mcrs,centers=2)

K-means clustering with 2 clusters of sizes 28, 22

Cluster means:

Murder Rape Robbery Assault Burglary Theft

1 -0.5895495 -0.6248911 -0.6261833 -0.6723856 -0.6496108 -0.4701287

2 0.7503358 0.7953159 0.7969606 0.8557634 0.8267774 0.5983457

Vehicle

1 -0.6444851

2 0.8202538

Clustering vector:

ME NH VT MA RI CT NY NJ PA OH IN IL MI WI MN IA MO ND SD NE KS DE MD VA WV

1 1 1 2 1 1 2 2 1 1 1 2 2 1 1 1 2 1 1 1 1 1 2 1 1

NC SC GA FL KY TN AL MS AR LA OK TX MT ID WY CO NM AZ UT NV WA OR CA AK HI

1 2 2 2 1 2 1 1 1 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2 1

Available components:

[1] "cluster" "centers" "totss" "withinss"

[5] "tot.withinss" "betweenss" "size"

> prcomp(Mcrs)

Standard deviations:

[1] 2.1376534 0.9821717 0.7849079 0.5839701 0.4629549 0.4473740 0.3069276

Rotation:

PC1 PC2 PC3 PC4 PC5 PC6 PC7

Murder 0.3794812 -0.2532759 0.60240357 -0.04387921 -0.1992800 -0.32002112 -0.5337639

Rape 0.3917234 0.2129042 0.08903091 -0.82757875 0.2273617 0.04562582 0.2337444

Robbery 0.3811897 -0.4150948 -0.25857003 0.23354486 0.5286578 -0.47190706 0.2425152

Assault 0.4096666 -0.1372090 0.30642169 0.32964915 0.1893113 0.74117076 0.1600352

Burglary 0.4108901 0.3152629 0.01613341 0.24359841 -0.5966670 -0.24104906 0.5079981

Theft 0.3104366 0.6973264 -0.22227257 0.24780086 0.2909714 -0.03368514 -0.4698464

Vehicle 0.3524385 -0.3372512 -0.64712080 -0.17097024 -0.3906411 0.25344635 -0.3117965

> kmeans(Mcrs,centers=2)$cluster

ME NH VT MA RI CT NY NJ PA OH IN IL MI WI MN IA MO ND SD NE KS DE MD VA WV NC SC GA FL

1 1 1 2 1 1 2 2 1 1 1 2 2 1 1 1 2 1 1 1 1 1 2 1 1 1 2 2 2

KY TN AL MS AR LA OK TX MT ID WY CO NM AZ UT NV WA OR CA AK HI

1 2 1 1 1 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2 1

> plot(prMcrs$x[,1:2],type="n")

> text(prMcrs$x[,1:2],labels=rownames(Mcrs),col=kmeans(Mcrs,centers=2)$cluster)



> plot(prMcrs$x[,1:2],type="n")

> text(prMcrs$x[,1:2],labels=rownames(Mcrs),col=kmeans(Mcrs,centers=3)$cluster)



> plot(prMcrs$x[,1:2],type="n",xlim=c(-5,5),ylim=c(-5,5))

> text(prMcrs$x[,1:2],labels=rownames(Mcrs),col=kmeans(Mcrs,centers=4)$cluster)



> crime\_s2<-crime\_s

> crime\_s<-crime\_s[rownames(crime\_s)!="DC",]

> plot(prcomp(crime\_s2)$x[,1:2],type="n")

> text(prcomp(crime\_s2)$x[,1:2],labels=rownames(crime\_s2),col=kmeans(crime\_s2,centers=4)$cluster)



> plot(prcomp(crime\_s)$x[,1:2],type="n")

> text(prcomp(crime\_s)$x[,1:2],labels=rownames(crime\_s),col=kmeans(crime\_s,centers=4)$cluster)

