Q\_@.R

Nadim

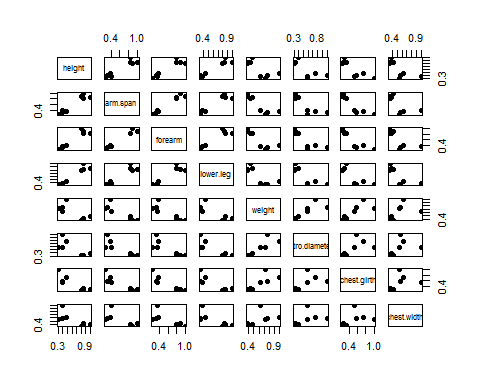
Wed Jan 30 22:54:15 2019

#######Question 2  
library('datasets')  
data\_<-Harman23.cor  
dat\_<-Harman23.cor$cov  
pairs(dat\_,pch=19)  
library(GGally)

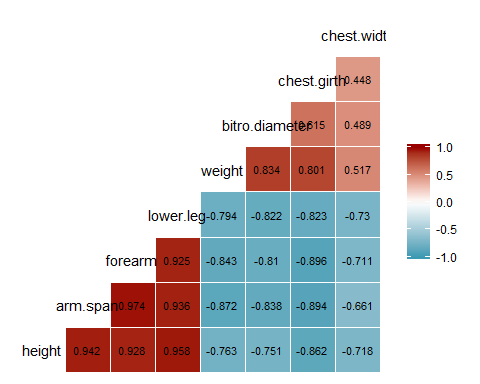
## Warning: package 'GGally' was built under R version 3.5.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.2



ggcorr(dat\_, low = "#3B9AB2", mid = "#FFFFFF", high = "#990000",label = T, label\_color = "black",label\_size = 3, label\_round = 3)



##Since we already have the covariance matrix we need not scale  
  
eig.out<-eigen(dat\_)  
str(eig.out)

## List of 2  
## $ values : num [1:8] 4.673 1.771 0.481 0.421 0.233 ...  
## $ vectors: num [1:8, 1:8] -0.398 -0.389 -0.376 -0.388 -0.351 ...  
## - attr(\*, "class")= chr "eigen"

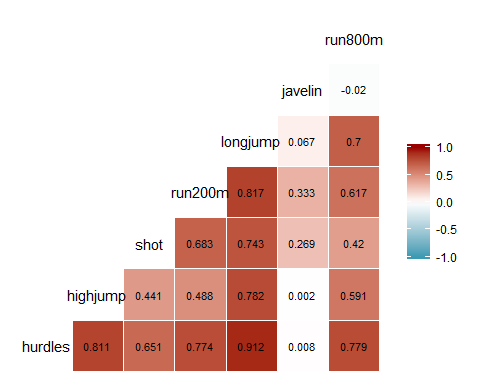
#retain only the first two PCs?  
lam <- eig.out$values  
  
tab <- rbind(lam,  
 lam/sum(lam), # proportion of variance explained  
 cumsum(lam)/sum(lam)) # cumulative proportion of var explained  
rownames(tab) <- c("Variance", "Proportion of variance", "Cumulative proportion")  
  
###The frst PC Vector kind of represents overall body metrics  
###IN the Second PC vector, the first half represents body lengths and the second  
### half represents the body girth measurements()  
  
#########Question 3  
data\_3=read.table("C:/Users/Nadim/Desktop/ST537/HW\_3/cancer\_dataset.txt",header = T,row.names = 1)  
data\_3.sd<-apply(data\_3,2,sd)  
std.data\_3<-scale(data\_3,center=T,scale=T)  
std.data\_3

## All.cancers Lung Colon Melanoma  
## Connecticut 0.84102053 -0.82426772 -0.05236419 0.78288392  
## Maine 1.15235697 0.27251757 0.11406903 0.17073532  
## Massachusetts 0.56544571 -0.26216526 -0.28814459 0.43043473  
## New-Hampshire 0.49392247 -0.41297324 -0.55166386 1.72893175  
## Rhode-Island 1.24491645 0.35477646 0.23889394 0.18928528  
## Vermont -0.30335125 0.08743505 -1.74443528 1.87733141  
## New-Jersey 1.21967296 -0.51579686 0.58562982 0.72723405  
## New-York 0.71901030 -0.78313827 0.23889394 -0.81241242  
## Pennsylvania 0.99248150 0.14912922 0.79367135 -0.60836289  
## Illinois 0.60962182 0.13541941 1.20975440 -0.60836289  
## Indiana -0.39380711 1.02655745 0.58562982 -0.51561310  
## Michigan 0.63907257 0.26566266 -0.06623363 -0.38576340  
## Ohio -0.14137216 0.79349058 0.30824112 -0.16316391  
## Wisconsin -0.83767190 -0.34442415 -0.88453030 -0.83096238  
## Iowa -0.15609753 0.09428996 0.36371886 -0.36721344  
## Kansas 0.18889690 0.24509794 0.40532717 0.26348511  
## Minnesota 0.56334208 -0.58434594 -0.23266685 0.67158417  
## Missouri 0.10895917 0.86889457 0.75206304 -0.38576340  
## Nebraska 0.12578817 -0.02224348 1.50101254 -0.88661225  
## North-Dakota 0.27724914 -0.93394625 2.84634775 -0.60836289  
## South-Dakota -1.26891495 -0.15934164 0.04472185 -1.68426042  
## Delaware 1.52680215 -0.01538857 0.46080491 1.22808290  
## DC 0.59700008 -0.33756925 0.55789095 -1.85121004  
## Florida -0.71776530 -0.40611833 -0.69035821 0.09653549  
## Georgia 0.32142525 0.71123168 0.03085242 0.35623490  
## Maryland -0.32018025 -0.22103581 -0.55166386 0.65303422  
## North-Carolina 0.29618176 0.94429856 -0.20492798 0.30058503  
## South-Carolina -0.05722718 0.84832984 -0.19105855 0.61593430  
## Virginia -0.10981779 0.23824303 -0.51005555 0.85708375  
## West-Virginia 1.03245037 1.45156175 1.40392649 -0.53416306  
## Alabama 0.67483419 1.62978936 0.93236570 0.31913498  
## Kentucky 1.28698895 2.34955470 1.75066237 1.17243303  
## Mississippi 1.51838765 2.17818200 1.02945175 -0.64546281  
## Tennessee 0.51916597 1.71890316 0.29437168 0.56028443  
## Louisiana 1.28278170 0.96486328 1.51488198 -1.22051149  
## Oklahoma 1.21546571 1.30075377 0.54402152 1.07968324  
## Texas -0.32018025 -0.20732599 -0.02462532 -0.94226213  
## Arizona -2.47429185 -1.22185238 -1.78604359 -0.88661225  
## Colorado -1.01648000 -1.45491926 -0.95387748 -0.18171387  
## Idaho -0.41484336 -1.03676987 -1.02322465 0.80143388  
## Montana -1.24156783 -0.50208704 -1.02322465 -0.12606400  
## New-Mexico -1.84110084 -1.38637018 -1.12031070 -0.86806230  
## Utah -1.55290427 -2.74364197 -1.67508810 2.48948001  
## Wyoming -1.39092517 -0.96822079 -1.38382997 -1.09066179  
## Alaska -2.24078951 0.24509794 -1.91086850 -2.92710758  
## California -0.89026252 -1.22870729 -0.42683894 0.54173447  
## Hawaii -1.18476996 -1.12588367 0.66884643 -0.42286332  
## Oregon -0.74300880 -0.50208704 -0.92613861 0.93128358  
## Washington -0.39591074 -0.67345974 -0.95387748 0.70868409  
## F.breast Pancreas Leukemia Ovarian  
## Connecticut 1.547001967 2.019202312 -0.65171339 0.10090625  
## Maine 0.922697105 0.054328313 2.29733268 -0.11406794  
## Massachusetts 1.272307828 0.181094378 0.31725889 -0.11406794  
## New-Hampshire 0.910211008 0.878307732 0.19087120 0.38753850  
## Rhode-Island 2.208765121 -0.769651105 0.48577581 -0.11406794  
## Vermont 1.522029773 0.624775603 -0.82023030 -0.83064856  
## New-Jersey 0.972641494 1.638904118 0.65429273 0.96080300  
## New-York 0.335850534 1.131839861 0.73855119 0.74582881  
## Pennsylvania 0.148559075 0.688158635 0.48577581 1.24743525  
## Illinois 0.385794923 0.624775603 0.02235429 0.45919656  
## Indiana -1.087564552 0.054328313 -0.73597185 -0.18572600  
## Michigan -0.300940426 1.131839861 0.31725889 0.81748687  
## Ohio -0.475745787 0.117711346 -0.82023030 0.02924819  
## Wisconsin -0.588120662 0.434626507 1.32836040 0.53085462  
## Iowa 0.161045173 -0.579502009 0.35938812 1.24743525  
## Kansas 0.298392242 -1.340098395 0.69642196 0.45919656  
## Minnesota 1.309766120 -1.023183234 2.08668653 0.10090625  
## Missouri -0.862814802 0.498009539 -0.60958416 -0.32904212  
## Nebraska -0.750439926 -1.403481427 0.94919733 -1.11728081  
## North-Dakota 0.410767118 0.941690764 2.12881576 0.67417075  
## South-Dakota -0.338398717 -1.276715363 -1.57855644 -0.40070018  
## Delaware 1.397168800 0.371243474 -0.06190417 2.46562230  
## DC 1.534515870 2.526266569 -2.33688257 -1.83386143  
## Florida -1.112536747 -0.199203815 -0.90448876 -0.11406794  
## Georgia 0.073642492 0.941690764 -0.65171339 0.88914493  
## Maryland 0.210989562 0.434626507 -0.73597185 -0.54401631  
## North-Carolina 0.235961756 -0.199203815 -0.35680878 0.02924819  
## South-Carolina -0.051218481 -0.516118977 -0.48319647 -0.40070018  
## Virginia 0.360822729 -0.706268073 -1.19939337 -0.47235825  
## West-Virginia -1.000161871 0.498009539 1.32836040 -0.04240987  
## Alabama -0.563148468 0.498009539 0.02235429 -0.61567437  
## Kentucky -0.138621161 -0.009054719 -0.10403340 -0.47235825  
## Mississippi -0.950217482 0.434626507 -0.60958416 -1.18893887  
## Tennessee -0.425801398 0.307860442 -0.56745493 -0.11406794  
## Louisiana -0.538176273 -0.199203815 0.31725889 -1.69054530  
## Oklahoma 0.435739312 0.498009539 2.08668653 0.74582881  
## Texas -1.137508941 -0.199203815 -0.18829186 -0.11406794  
## Arizona -2.698271097 -0.642885041 -1.78920258 -0.25738406  
## Colorado -0.001274092 -0.579502009 -0.01977494 0.96080300  
## Idaho 0.273420048 0.054328313 0.86493888 2.17899005  
## Montana -0.588120662 -0.896417170 -0.35680878 -0.68733243  
## New-Mexico -1.474633567 -1.340098395 0.06448351 -0.04240987  
## Utah -2.248771596 -1.720396588 0.06448351 -2.04883561  
## Wyoming -1.549550150 -2.227460846 0.19087120 -3.33868073  
## Alaska 0.398281020 -2.227460846 -0.98874722 0.81748687  
## California 0.086128589 -0.325969880 -0.31467955 0.24422238  
## Hawaii 0.148559075 1.321988957 -1.15726414 0.17256431  
## Oregon 1.010099786 -0.516118977 -0.73597185 0.53085462  
## Washington 0.310878340 -0.009054719 0.78068042 0.38753850  
## Cervix Prostate Liver  
## Connecticut -0.62301855 0.71178769 0.342294084  
## Maine 0.03919206 0.31841050 -1.242500572  
## Massachusetts -1.41767129 0.38489679 0.826536895  
## New-Hampshire -1.55011341 -0.07496668 -1.110434351  
## Rhode-Island -0.29191325 0.56219355 0.782514821  
## Vermont 0.30407630 -0.66226219 0.210227862  
## New-Jersey 1.16495010 1.28246163 0.342294084  
## New-York 0.63518161 0.63976088 1.618934223  
## Pennsylvania 0.63518161 0.40705888 0.738492748  
## Illinois 0.56896055 0.27962684 0.210227862  
## Indiana -0.02702900 -1.54874598 -0.274014949  
## Michigan 0.37029736 0.47354517 -0.229992875  
## Ohio 0.03919206 -0.65672166 -0.802279834  
## Wisconsin -1.35145023 -0.95590994 -0.538147392  
## Iowa -0.88790280 -1.18861194 -1.462610941  
## Kansas -0.35813431 -0.06388563 -0.362059097  
## Minnesota -1.15278704 1.51516362 -1.022390203  
## Missouri 0.63518161 -0.97807204 0.298272010  
## Nebraska 0.37029736 0.14111375 -1.154456424  
## North-Dakota 0.30407630 0.26300527 -1.286522646  
## South-Dakota -1.35145023 0.07462746 0.386316158  
## Delaware 0.70140267 2.44043108 0.694470674  
## DC 2.48937132 2.61772783 1.486868002  
## Florida 0.90006585 -1.23293613 0.386316158  
## Georgia 0.63518161 0.50124779 -0.009882506  
## Maryland -0.68923962 0.14111375 0.122183715  
## North-Carolina -0.15947113 -0.31874972 0.078161641  
## South-Carolina 0.30407630 0.03030327 -0.450103244  
## Virginia -0.82168174 0.74503083 -0.670213613  
## West-Virginia 1.16495010 -0.18577715 -0.494125318  
## Alabama 1.42983434 0.29070789 -0.141948728  
## Kentucky 1.03250797 -0.77307266 -0.626191539  
## Mississippi 0.96628691 1.17165115 0.474360305  
## Tennessee 0.23785524 -0.23010134 -0.053904580  
## Louisiana 0.90006585 1.16057010 0.914581043  
## Oklahoma 1.89338177 0.67854454 0.518382379  
## Texas 1.49605540 -0.45726281 1.971110813  
## Arizona -0.22569219 -2.21360882 0.210227862  
## Colorado -0.35813431 0.01922223 0.122183715  
## Idaho -1.81499766 0.41259941 -1.770765457  
## Montana -1.55011341 -0.09712877 -1.902831678  
## New-Mexico -0.35813431 -1.22185508 1.618934223  
## Utah -1.94743978 0.63422035 -1.682721310  
## Wyoming -0.82168174 0.57881512 -0.846301908  
## Alaska -0.82168174 -2.10279835 -0.758257760  
## California 0.43651842 -0.65672166 1.706978371  
## Hawaii -0.35813431 -1.77036693 2.587419846  
## Oregon 0.30407630 -0.81739685 -0.229992875  
## Washington -1.02034492 -0.26888501 0.474360305  
## attr(,"scaled:center")  
## All.cancers Lung Colon Melanoma F.breast Pancreas   
## 544.120408 66.424490 52.777551 23.679592 121.510204 13.014286   
## Leukemia Ovarian Cervix Prostate Liver   
## 15.246939 11.959184 7.440816 159.953061 9.022449   
## attr(,"scaled:scale")  
## All.cancers Lung Colon Melanoma F.breast Pancreas   
## 47.536999 14.588088 7.210099 5.390848 8.008908 1.577709   
## Leukemia Ovarian Cervix Prostate Liver   
## 2.373649 1.395516 1.510094 18.048835 2.271588

dat\_3.pca<- prcomp(std.data\_3)  
summary(dat\_3.pca)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6 PC7  
## Standard deviation 1.8872 1.4110 1.2070 1.04549 0.92966 0.8028 0.69483  
## Proportion of Variance 0.3238 0.1810 0.1324 0.09937 0.07857 0.0586 0.04389  
## Cumulative Proportion 0.3238 0.5048 0.6372 0.73660 0.81517 0.8738 0.91765  
## PC8 PC9 PC10 PC11  
## Standard deviation 0.59711 0.52480 0.48278 0.20192  
## Proportion of Variance 0.03241 0.02504 0.02119 0.00371  
## Cumulative Proportion 0.95007 0.97511 0.99629 1.00000

###Covariance deals with similarity of variables of same kinds of fields. Correlation   
### provides a metric for variables from even different fields  
  
#####Since we have different cancer types the metrics maybe different. Hence   
#####I think it would better if we use correlation  
  
###########Question 4  
###in some variables high values indicate better performances and in some lower values represent better performances  
data("heptathlon", package = "HSAUR3")  
a=heptathlon[,8]  
heptathlon[,8]=NULL  
heptathlon[,1]<-(max(heptathlon[,1])-heptathlon[,1])  
heptathlon[,4]<-(max(heptathlon[,4])-heptathlon[,4])  
heptathlon[,7]<-(max(heptathlon[,7])-heptathlon[,7])  
ggcorr(heptathlon, low = "#3B9AB2", mid = "#FFFFFF", high = "#990000",label = T, label\_color = "black",label\_size = 3, label\_round = 3)



##we can observe obvious patterns or high corr scores for sports of similar nature  
##but also we can see the correlation between sports of different natures like shot and longjump  
##########  
#are correlation matrices standardised  
std.data\_4<-scale(heptathlon,center=T,scale=T)  
dat\_4.pca<-prcomp(std.data\_4)  
summary(dat\_4.pca)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6  
## Standard deviation 2.1119 1.0928 0.72181 0.67614 0.49524 0.27010  
## Proportion of Variance 0.6372 0.1706 0.07443 0.06531 0.03504 0.01042  
## Cumulative Proportion 0.6372 0.8078 0.88223 0.94754 0.98258 0.99300  
## PC7  
## Standard deviation 0.2214  
## Proportion of Variance 0.0070  
## Cumulative Proportion 1.0000

dat\_4.pca$rotation

## PC1 PC2 PC3 PC4 PC5  
## hurdles -0.4528710 0.15792058 -0.04514996 0.02653873 -0.09494792  
## highjump -0.3771992 0.24807386 -0.36777902 0.67999172 0.01879888  
## shot -0.3630725 -0.28940743 0.67618919 0.12431725 0.51165201  
## run200m -0.4078950 -0.26038545 0.08359211 -0.36106580 -0.64983404  
## longjump -0.4562318 0.05587394 0.13931653 0.11129249 -0.18429810  
## javelin -0.0754090 -0.84169212 -0.47156016 0.12079924 0.13510669  
## run800m -0.3749594 0.22448984 -0.39585671 -0.60341130 0.50432116  
## PC6 PC7  
## hurdles -0.78334101 0.38024707  
## highjump 0.09939981 -0.43393114  
## shot -0.05085983 -0.21762491  
## run200m 0.02495639 -0.45338483  
## longjump 0.59020972 0.61206388  
## javelin -0.02724076 0.17294667  
## run800m 0.15555520 -0.09830963

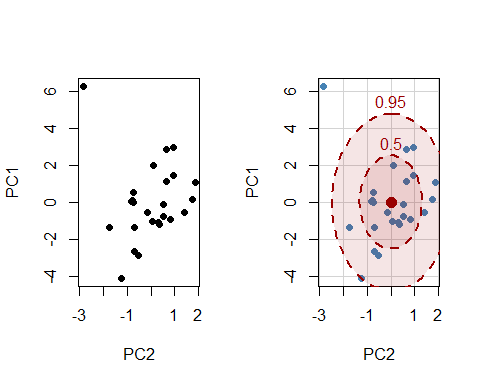
#####We can observe in the first PC that there except for javelin, the rest of the variables  
#####have significant components.(which i think reflects the core and cardio strength)  
####the second PC has a huge javelin component.(which i think reflects the upper body strength)   
par(mfrow=c(1,2))  
plot(dat\_4.pca$x[,2:1],pch=19)  
library(car)

## Warning: package 'car' was built under R version 3.5.2

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.5.2

dataEllipse(dat\_4.pca$x[,2:1], pch=19, col=c("steelblue","#990000"),lty=2,ellipse.label=c(0.5,0.95),levels=c(0.5,0.90),fill=TRUE,fill.alpha=0.1)



plot(a,dat\_4.pca$x[,1],xlab = 'Official\_scores',ylab = 'PCA\_1',pch=19)

