```
dat=read.table("heightWeightData.txt")
names(dat)=c("gender","height","weight")
male=dat[which(dat[,1]==1),2:3]
co=cov(male)
mn=as.vector(sapply(male,mean))
library(mixtools)

mixtools package, version 1.0.4, Released 2016-01-11
This package is based upon work supported by the National Science
Foundation under Grant No. SES-0518772.

library(ggplot2)
library(MASS)
#ggplot(male,aes(x=height,y=weight))+geom_text(aes(label=rownames(male)),check_overlap =FALS
simu=as.data.frame(ellipse(mu=mn,sigma=co,alpha=0.05,npoints = 500,draw=F))
names(simu)=c("height","weight")
ggplot()+geom_text(data=male,aes(label=rownames(male),x=height,y=weight))+geom_path(data=sim
```
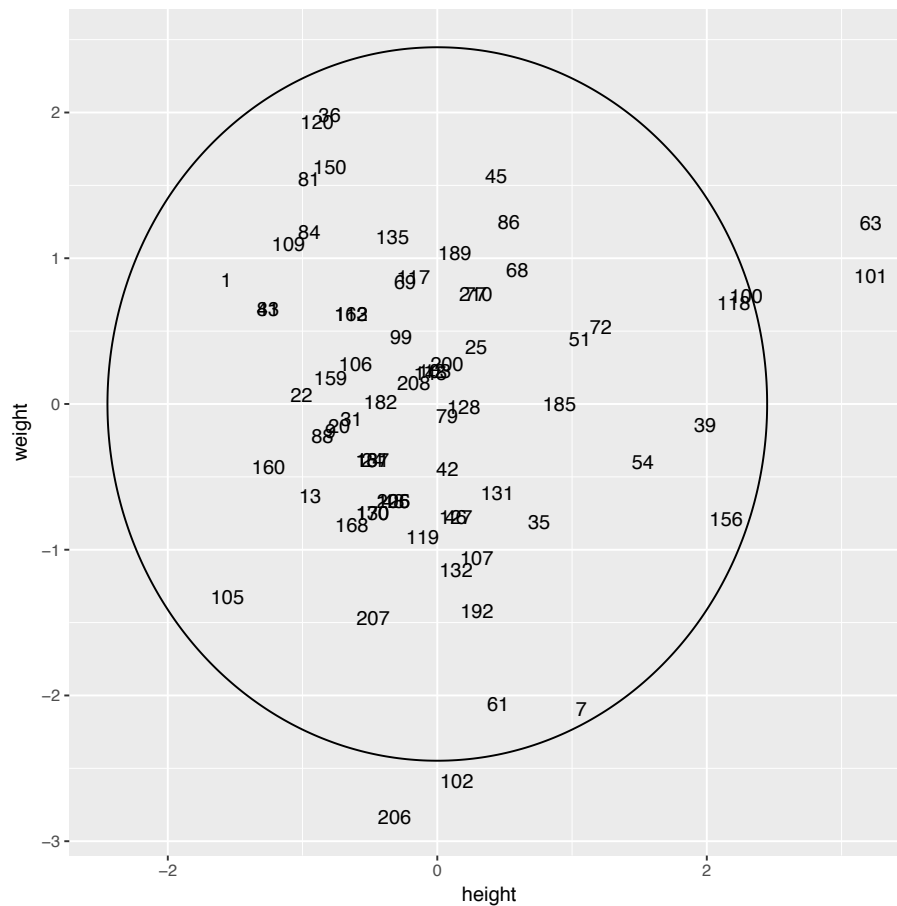
```
stmale=as.data.frame(scale(male))
simu=as.data.frame(ellipse(mu=0,cov(stmale),alpha=0.05,npoints = 500,draw=F))
names(simu)=c("height","weight")
ggplot()+geom_text(data=stmale,aes(label=rownames(stmale),x=height,y=weight))+geom_path(data
```
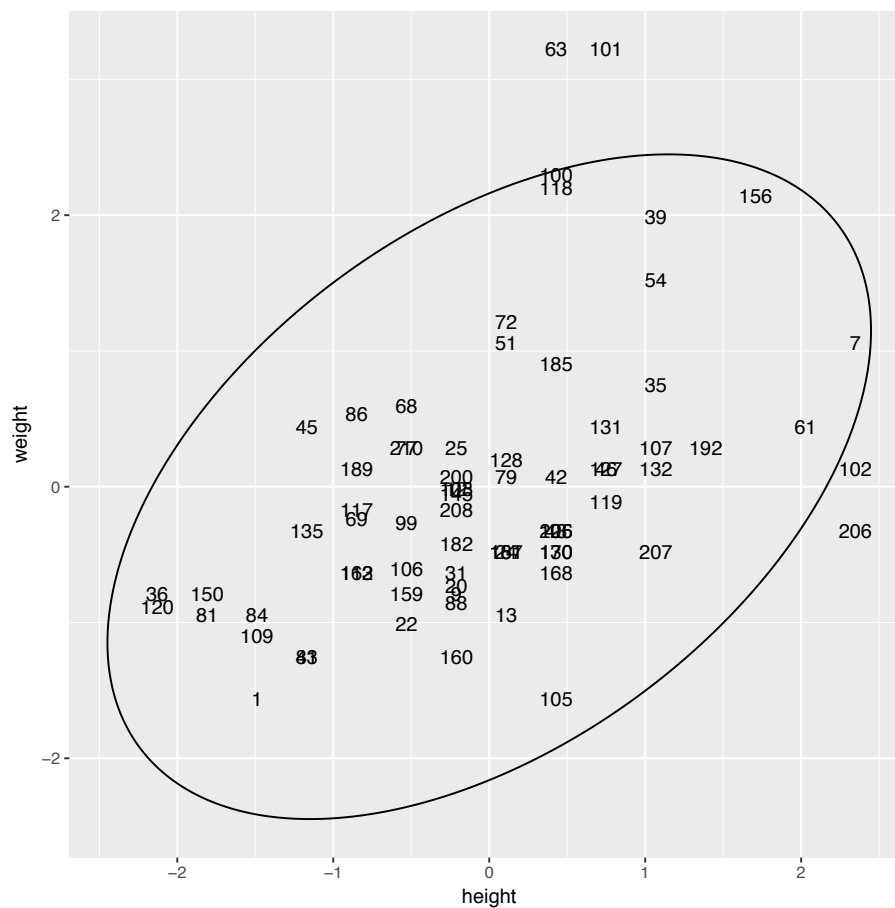
```
decom=eigen(t(scale(male,T,F)) %*% scale(male,T,F))
newmale=diag(decom$values^(-1/2)) %*% t(decom$vectors) %*% t(as.matrix(male))
newmale=as.data.frame(scale(t(newmale)),T,F)
names(newmale)=c("height","weight")
simu=as.data.frame(ellipse(apply(newmale,2,mean),cov(newmale),alpha=0.05,npoints = 500,draw=
names(simu)=c("height","weight")
ggplot()+geom_text(data=newmale,aes(label=rownames(stmale),x=height,y=weight))+geom_path(dat
```

```
dat=read.table("crabs.txt",header = T)
library(GGally)
datt=dat
datt$Species=factor(datt$Species)
datt$Sex=factor(datt$Sex)
ggpairs(datt,diag=list(continuous="barDiag"),upper = list(continuous = "points"),lower =list
```

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```

```
## 'stat_bin()' using 'bins = 30'.  Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'.  Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'.  Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'.  Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'.  Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'.  Pick better value with 'binwidth'.
```
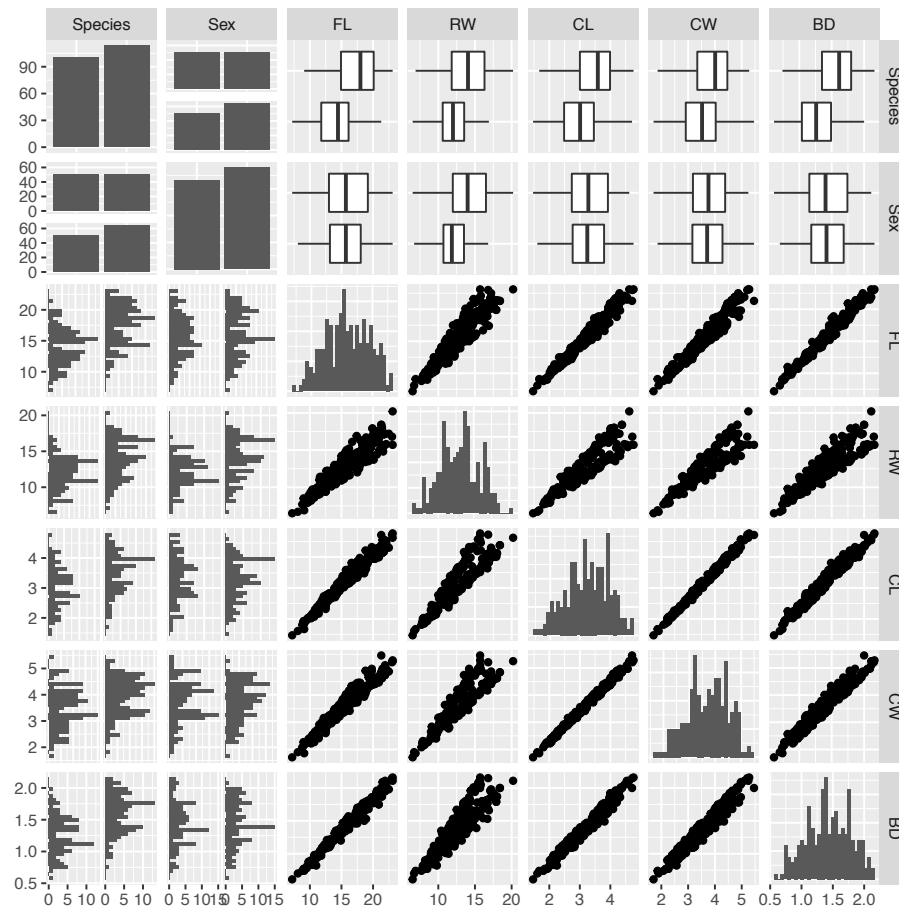


```r
pca1=princomp(dat[,3:7],cor=T)
loadings(pca1)[,1:3]
```

```
##          Comp.1       Comp.2       Comp.3
## FL -0.4518204 -0.1096168  0.53658552
## RW -0.4296613  0.8953731 -0.05670001
## CL -0.4526144 -0.2846850 -0.29343501
## CW -0.4504866 -0.2130314 -0.64480454
## BD -0.4510560 -0.2446697  0.45495505
```
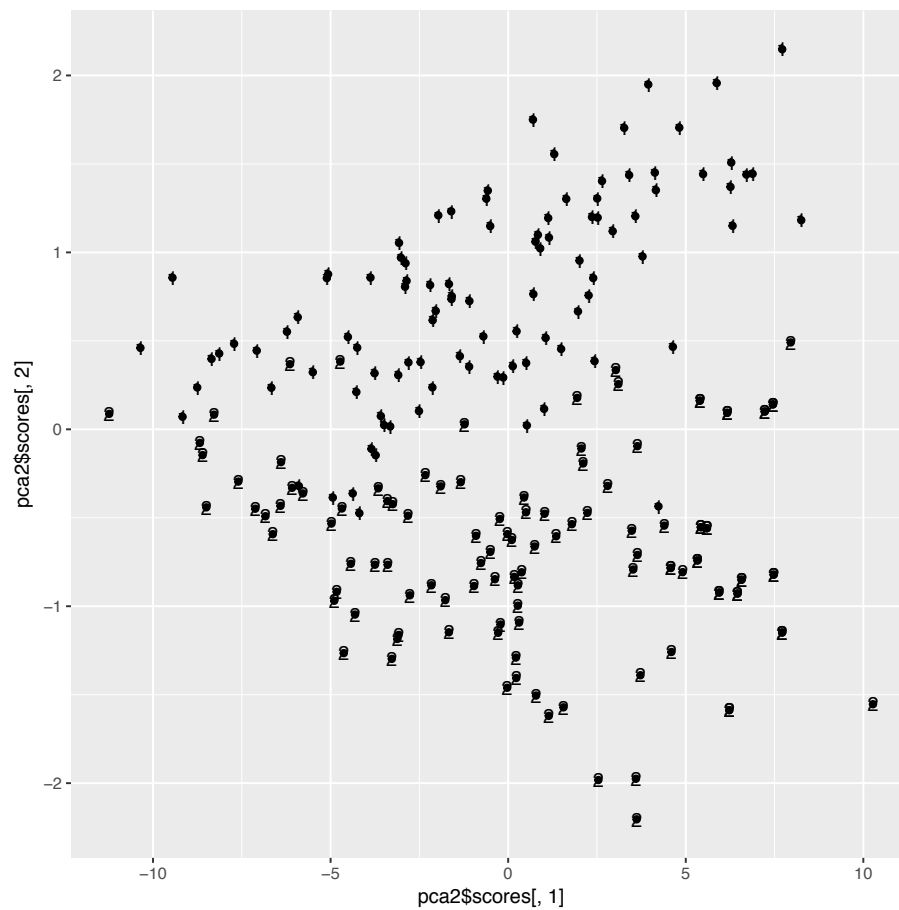
```r
qplot(x = pca1$scores[,1], y=pca1$scores[,2])+geom_text(aes(label=dat[,2]))
```
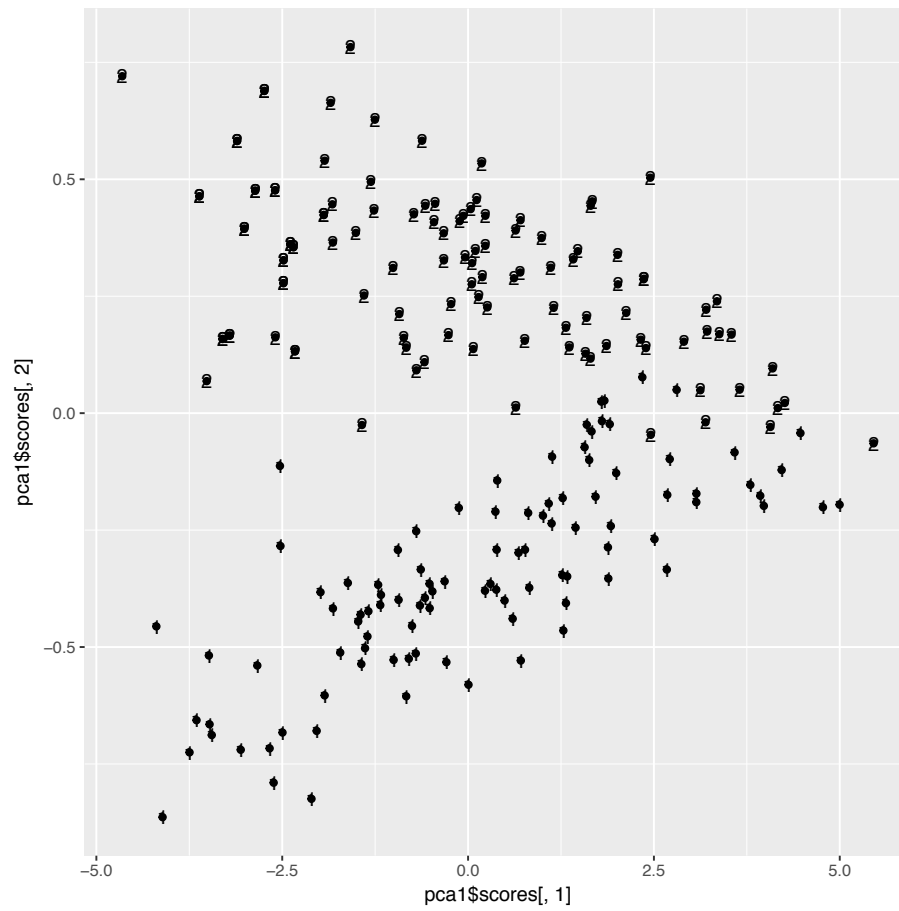


```r
pca2=princomp(dat[,3:7],cor=F)
loadings(pca2)[,1:3]

##        Comp.1      Comp.2      Comp.3
## FL 0.78236903  0.55458150  0.27004750
## RW 0.57470124 -0.81788273  0.01993393
## CL 0.15375104  0.11206356 -0.54771228
## CW 0.16836375  0.08548584 -0.78772673
## BD 0.07504306  0.06034167 -0.07858021

qplot(x = pca2$scores[,1], y=pca2$scores[,2])+geom_text(aes(label=dat[,2]))
```

```
#choose correlation, sexes are divided
summary(pca1)

## Importance of components:
##                          Comp.1     Comp.2      Comp.3      Comp.4
## Standard deviation     2.1911936 0.37610328 0.211188252 0.104263972
## Proportion of Variance 0.9602659 0.02829074 0.008920096 0.002174195
## Cumulative Proportion  0.9602659 0.98855661 0.997476709 0.999650904
##                            Comp.5
## Standard deviation     0.0417789539
## Proportion of Variance 0.0003490962
## Cumulative Proportion  1.0000000000

barplot(pca1$sdev^2)
```
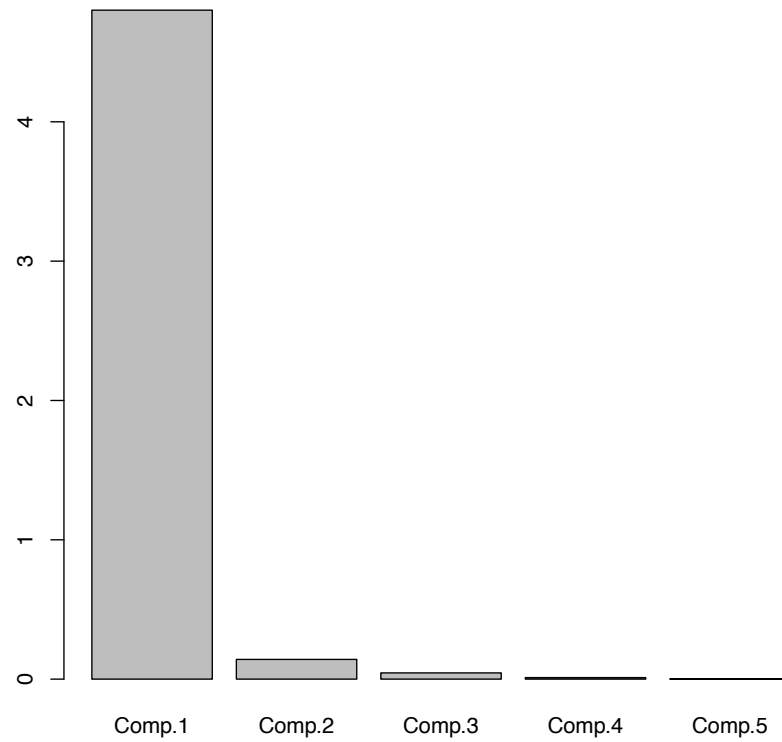
```
#retain 3 variables for precision of prediction
bs=c()
for(i in seq(500)){
  newdat=dat[sample(seq(dim(dat)[1]),dim(dat)[1],replace = T),3:7]
  pca=princomp(newdat,cor=T)
  perc=pca$sdev^2/sum(pca$sdev^2)
  bs[i]=sum(perc[1:2])
}
ci=quantile(bs,c(0.025,0.975))
ci
```

```
##      2.5%      97.5%
## 0.9859468 0.9906520
```

```
library(ggbiplot)
```

```
## Loading required package:  plyr
```

```
## Loading required package:  scales
## Loading required package:  grid
##
## Attaching package:  'grid'
## The following object is masked from 'package:mixtools':
##
##     depth
```

```
ggbiplot(pca1)
```