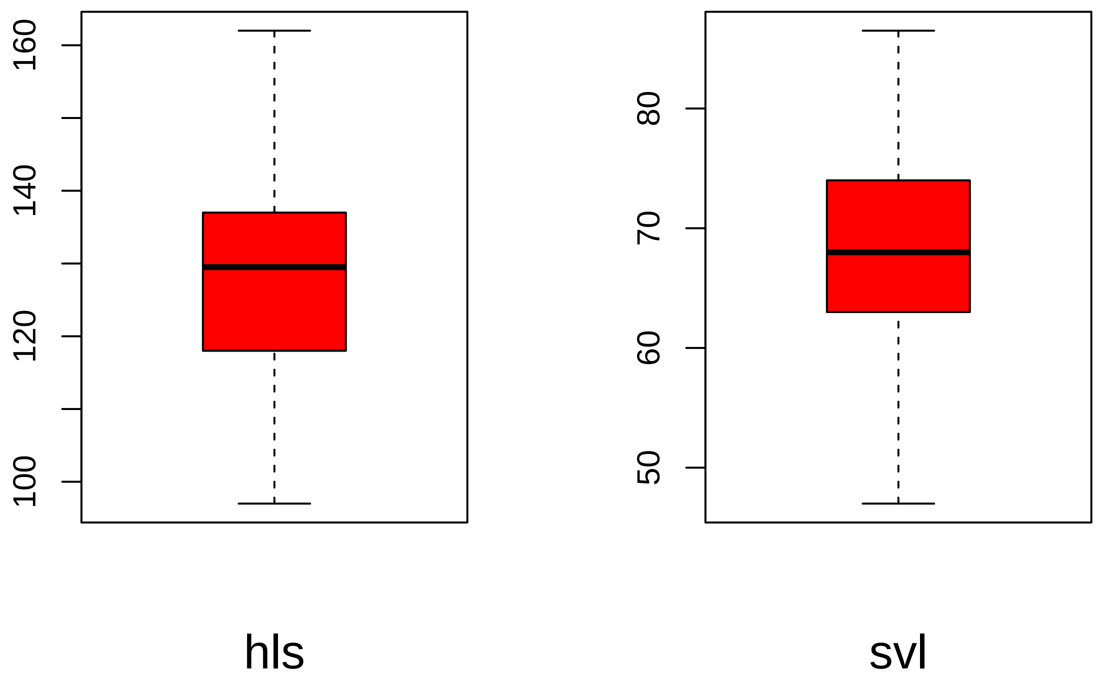
MATH 437 HW1

Drew Remmenga

1.a.

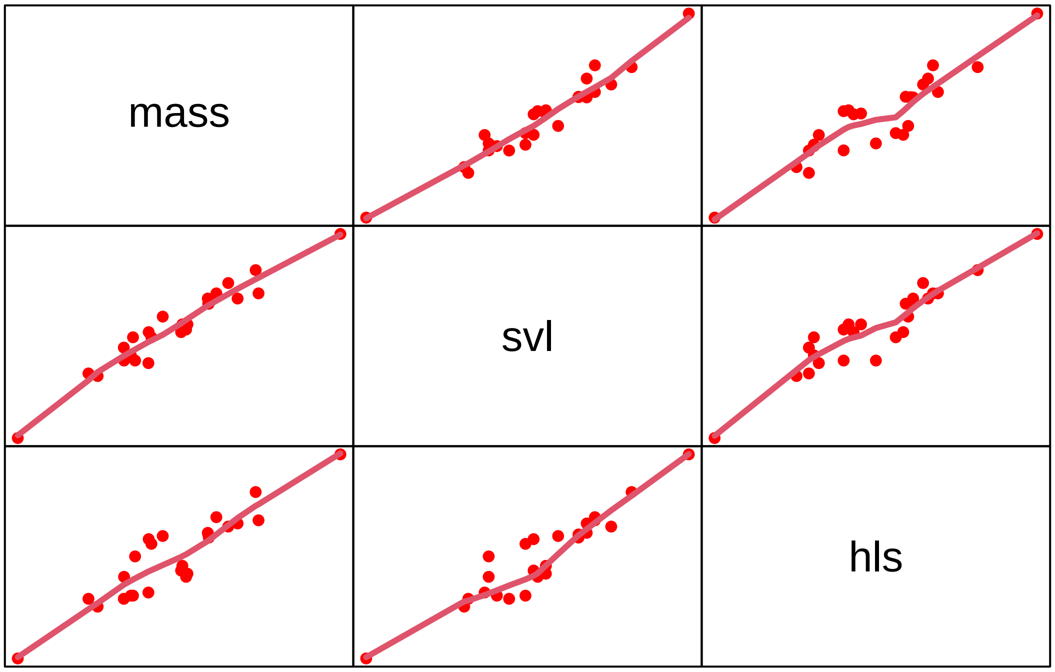
```
lizard <- read.csv("~/School/Math437/HW1/lizard.dat", sep="")
op <- par(mfrow = c(1, 2), cex.lab = 1.5)
boxplot(lizard$hls, xlab = "hls", col = "red")
boxplot(lizard$svl, xlab = "svl", col = "red")
```




```
par(op)
```

These two data have different medians. 1.b.

```
pairs(lizard[-1], lwd = 3, pch = 16, cex = 1.25, col = "red",
gap = 0, xaxt = "n", yaxt = "n",
panel = panel.smooth)
```



This data is mostly linear. 1.c. A bagplot is the generalized boxplot to multiple variables. 1.d.

```
library(MVA)
```

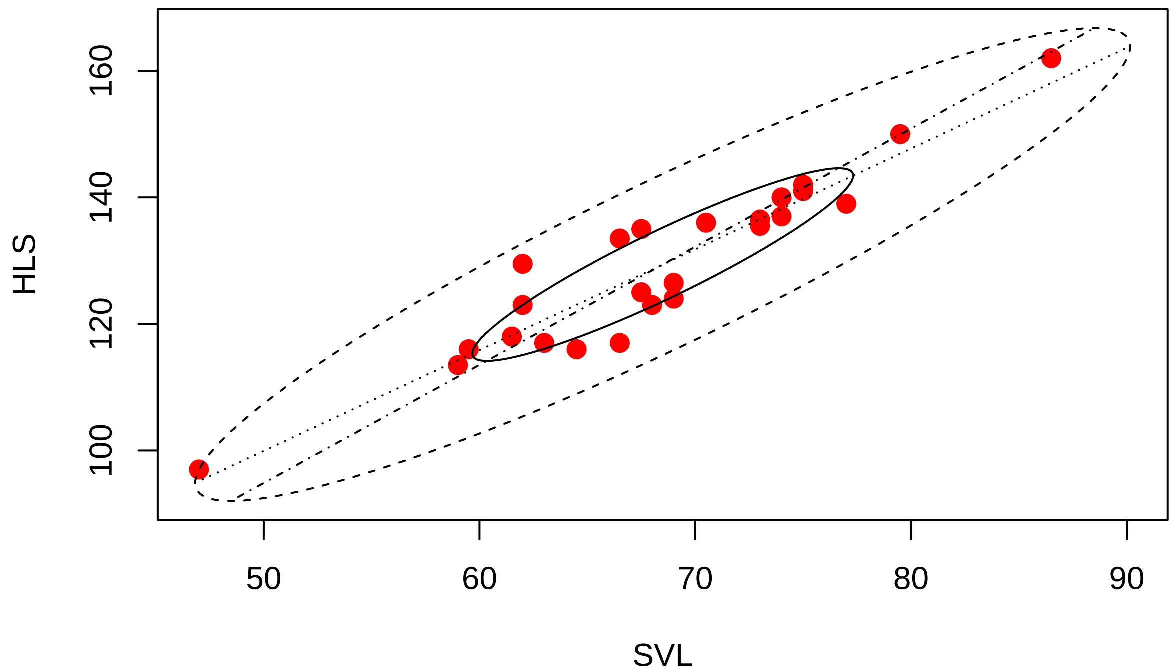
```
## Warning: package 'MVA' was built under R version 4.1.3

## Loading required package: HSAUR2

## Warning: package 'HSAUR2' was built under R version 4.1.3

## Loading required package: tools
```

```
bvbox(lizard[3:4], xlab = "SVL", ylab = "HLS",
pch = 19, cex = 1.25, col = "red")
```



1.e.

```
var(lizard[-1])
```

```
##            mass       svl       hls
## mass  7.292144  20.87785  33.80618
## svl  20.877854  63.77083 102.08542
## hls  33.806175 102.08542 185.83083
```

```
cor(lizard[-1])
```

```
##            mass       svl       hls
## mass 1.0000000 0.9681601 0.9183527
## svl  0.9681601 1.0000000 0.9377645
## hls  0.9183527 0.9377645 1.0000000
```

1.f.

```
mean(lizard$mass)
```

```
## [1] 8.6786
```

```
mean(lizard$svl)
```

```
## [1] 68.4
```

```r
mean(lizard$hls)
```

```
## [1] 129.32
```

```r
median(lizard$mass)
```

```
## [1] 8.953
```

```r
median(lizard$svl)
```

```
## [1] 68
```

```r
median(lizard$hls)
```

```
## [1] 129.5
```

1.g.

```r
sd(lizard$mass)
```

```
## [1] 2.700397
```

```r
sd(lizard$svl)
```

```
## [1] 7.985664
```

```r
sd(lizard$hls)
```

```
## [1] 13.63198
```

```r
mad(lizard$mass)
```

```
## [1] 2.802114
```

```r
mad(lizard$svl)
```

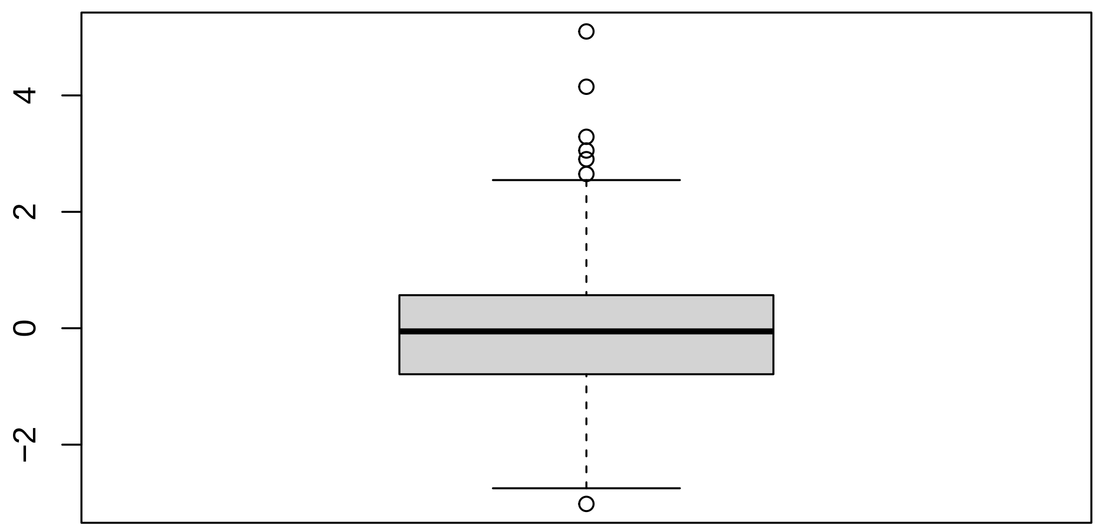
```
## [1] 8.8956
```

```r
mad(lizard$hls)
```

```
## [1] 14.0847
```

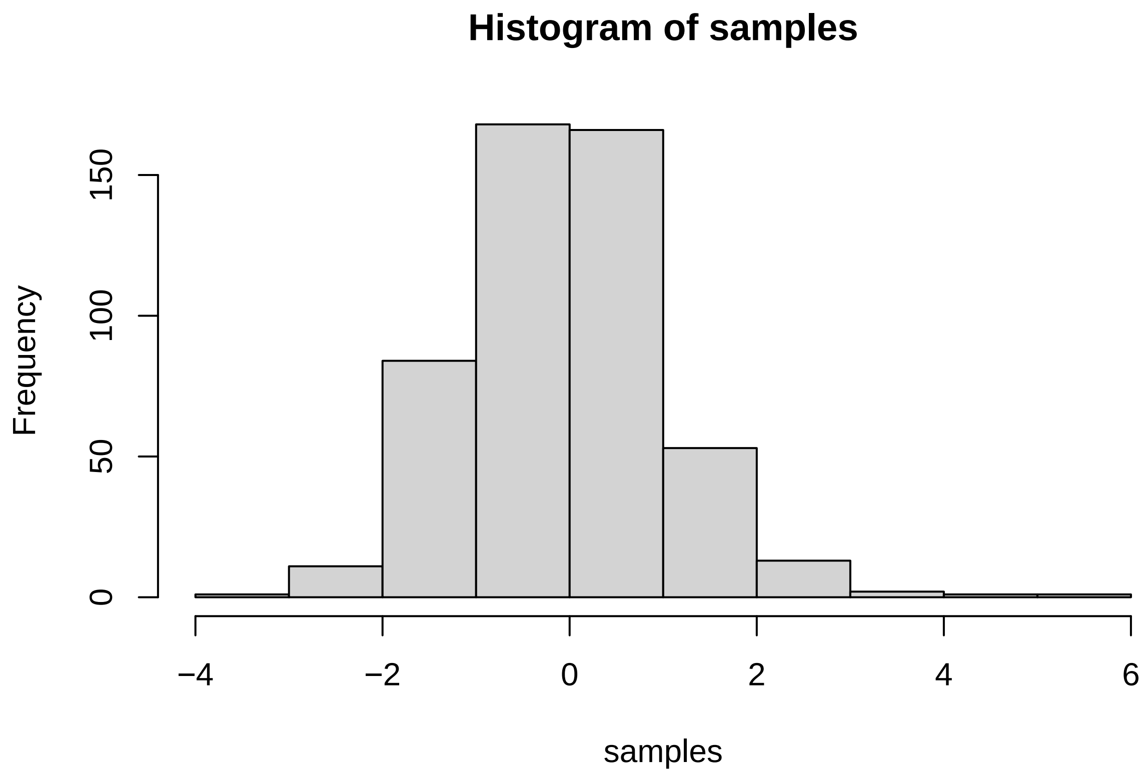
2.a.

```r
N = 500
components = sample(1:2,prob=c(.99,.01),size=N,replace=TRUE)
mus = c(0,3)
sds = sqrt(c(1,4))
samples = rnorm(n=N,mean=mus[components],sd=sds[components])
boxplot(samples)
```



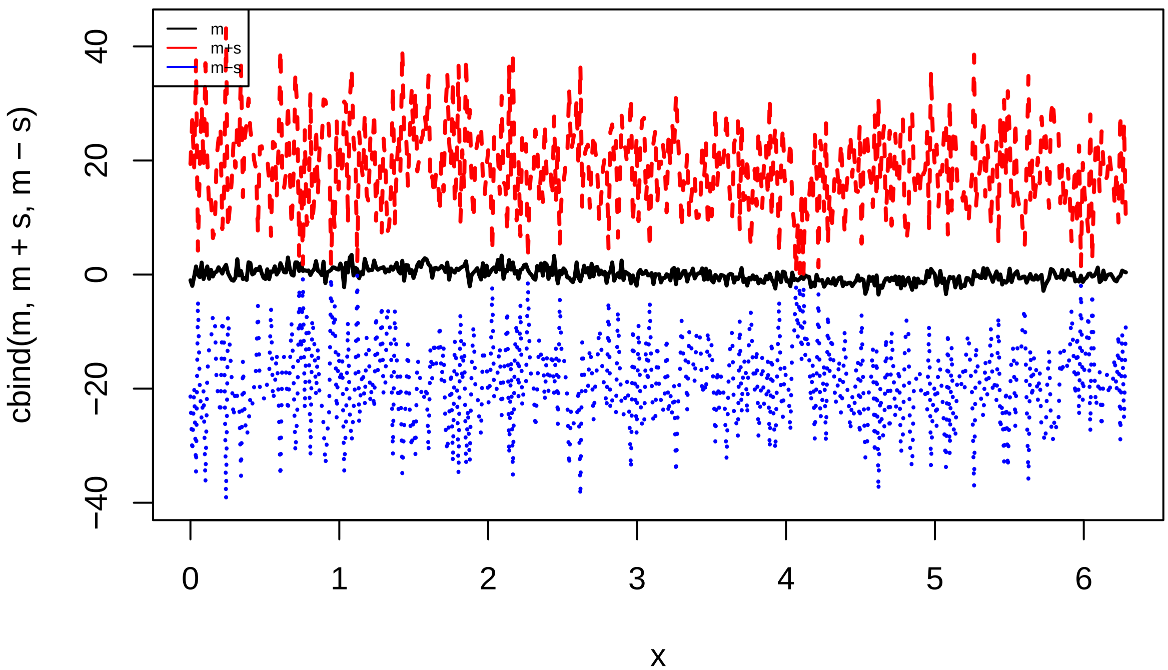
2.b.

```r
hist(samples)
```



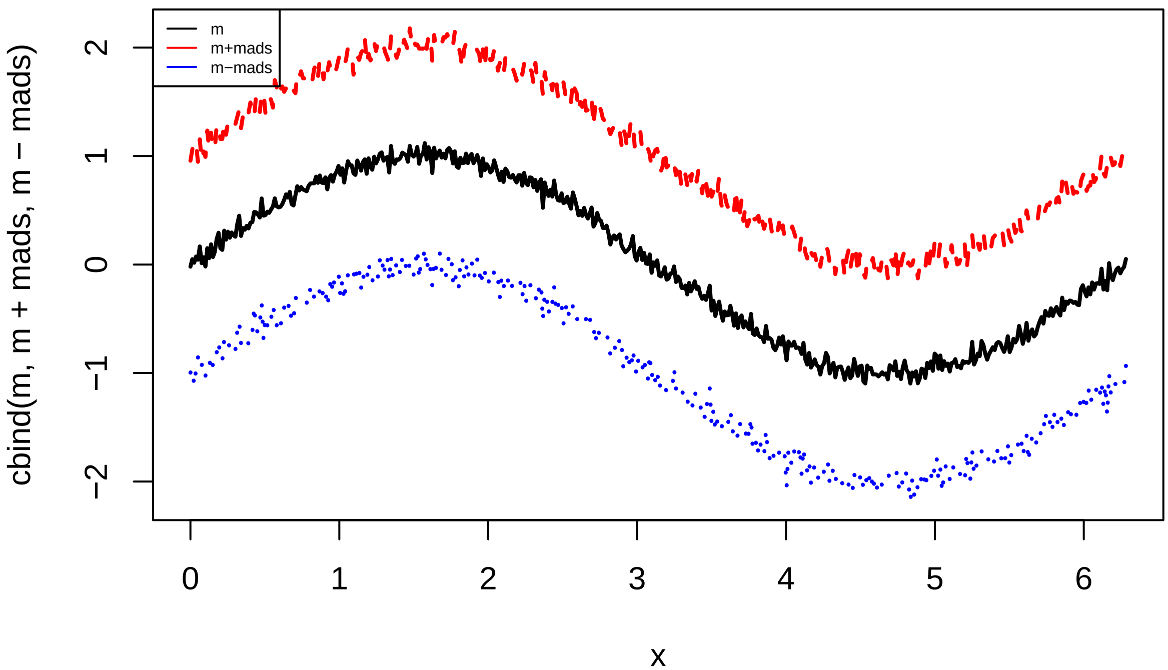
3.a.

```
x=seq(from=0,to=2*pi,length.out=N)
e1 = rnorm(N^2)
e2 = rnorm(N^2,0,200)
U1 = runif(N^2)
E = ifelse(U1<.99,e1,e2)
E = matrix(E,N,N)
means = apply(E[,1:500],2,mean)
s = apply(E[,1:500],2,sd)
m = sin(x)+means
matplot(x,cbind(m,m+s,m-s),type="l", col=c("black","red","blue"),lwd=2)
legend("topleft", legend=c("m","m+s","m-s"),col=c("black","red","blue"),lty=1:1,cex=.5)
```



3.b.

```
medians=apply(E[,1:500],2,median)
mads =apply(E[,1:500],2,mad)
m = sin(x)+medians
matplot(x,cbind(m,m+mads,m-mads),type="l", col=c("black","red","blue"),lwd=2)
legend("topleft", legend=c("m","m+mads","m-mads"),col=c("black","red","blue"),lty=1:1,cex=.5)
```



3.c. Mads and medians are much more stable than standard deviations and means. 4.a.

```
x = c(1.1,2.1,3.1,4.1,2.1,4.1,6.1,8.1,3.1,6.1,9.1,12.1)
A=matrix(x,nrow=3,ncol=4)
svd(A)
```

```
## $d
## [1] 20.109997  4.814510  2.128028
##
## $u
##            [,1]       [,2]        [,3]
## [1,] -0.4600358 -0.4195488  0.78252530
## [2,] -0.6087486 -0.4925456 -0.62195183
## [3,] -0.6463685  0.7624813  0.02881113
##
## $v
##            [,1]        [,2]        [,3]
## [1,] -0.1883717  0.18025563 -0.16729412
## [2,] -0.2891412  0.07719942  0.94941455
## [3,] -0.4843772 -0.86928361 -0.08227853
## [4,] -0.8039230  0.45375561 -0.25269534
```

4.b.

```r
svd_fact = svd(A)
d=svd_fact$d
u=svd_fact$u
v=svd_fact$v
d[1]*u[1:3,1]%*%t(v[1:4,1])
```

```
##          [,1]     [,2]     [,3]      [,4]
## [1,] 1.742686 2.674938 4.481128  7.437348
## [2,] 2.306033 3.539647 5.929713  9.841570
## [3,] 2.448543 3.758394 6.296162 10.449768
```