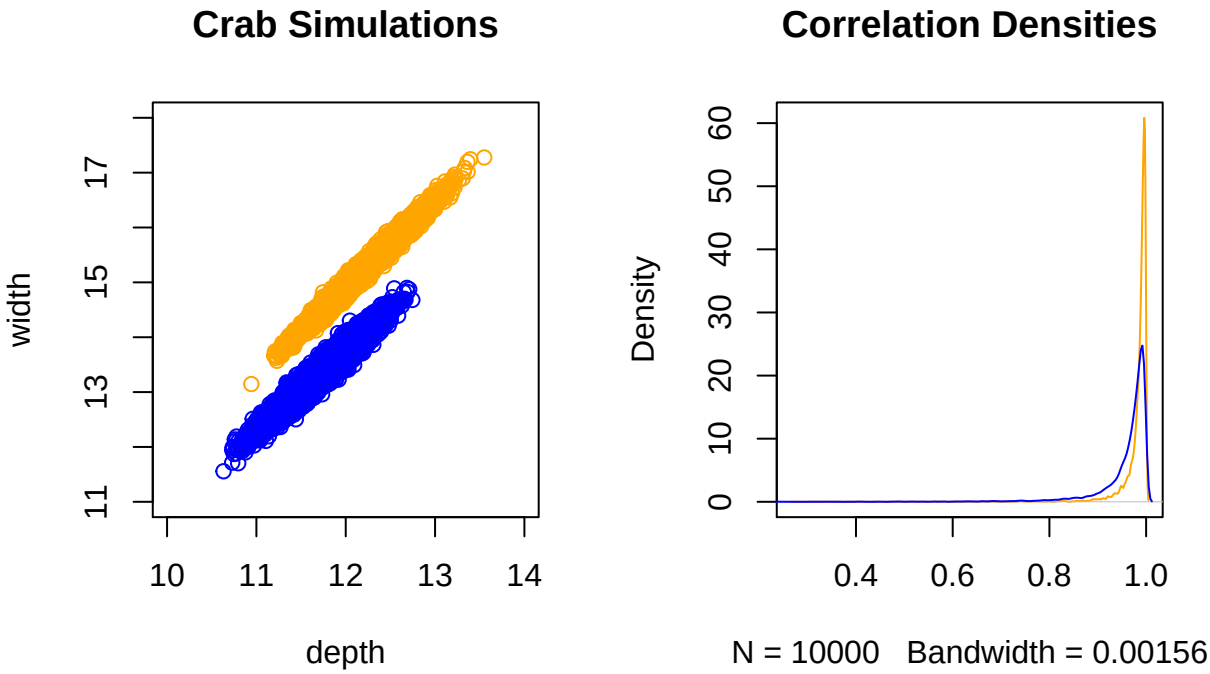
Homework 06
Joseph Blubaugh
jblubau1@tamu.edu

a)

```
library(MASS)

blue = read.table(file = "../Data/bluecrab.dat", sep = " ")
orange = read.table(file = "../Data/orangecrab.dat", sep = " ")

## Gibbs Sampling Blue
Mu.0 = colMeans(blue); S.o = cov(blue); V.o = 4

blue.sim = data.frame(); blue.cor = c()

for(i in 1:10000) {
  A.o = matrix(rWishart(n = 1, df = V.o, Sigma = S.o), nrow = 2)

  A.n = solve(A.o) + (50 * solve(S.o))
  B.n = solve(A.o) %*% Mu.0 + (50 * solve(S.o) %*% colMeans(blue))

  x = mvrnorm(n = 1, mu = solve(A.n) %*% B.n, Sigma = solve(A.n))

  blue.sim = rbind(blue.sim, x)
  blue.cor = c(blue.cor, A.o[1,2] / (sqrt(A.o[1,1]) * sqrt(A.o[2,2])))
}

colnames(blue.sim) = c("depth", "width")

## Gibbs Sampling Orange
Mu.0 = colMeans(orange); S.o = cov(orange); V.o = 4

orange.sim = data.frame(); orange.cor = c()

for(i in 1:10000) {
  A.o = matrix(rWishart(n = 1, df = V.o, Sigma = S.o), nrow = 2)

  A.n = solve(A.o) + (50 * solve(S.o))
  B.n = solve(A.o) %*% Mu.0 + (50 * solve(S.o) %*% colMeans(orange))

  x = mvrnorm(n = 1, mu = solve(A.n) %*% B.n, Sigma = solve(A.n))

  orange.sim = rbind(orange.sim, x)
  orange.cor = c(orange.cor, A.o[1,2] / (sqrt(A.o[1,1]) * sqrt(A.o[2,2])))
}

colnames(orange.sim) = c("depth", "width")
```

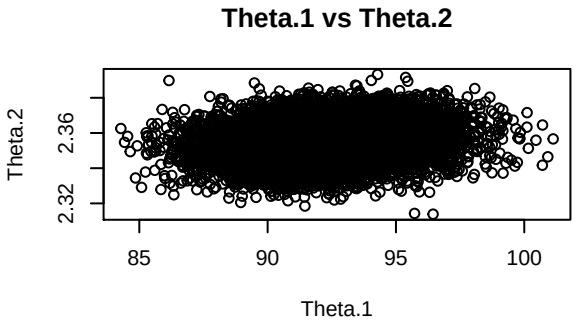
b) The orange crabs are much larger than the blue crabs. The orange crabs also have more variate in size, but their features are slightly more correlated than the blue crabs.

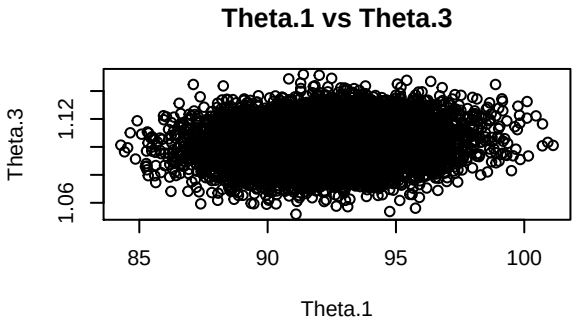


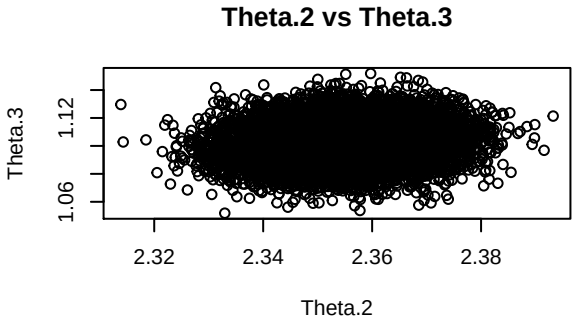
c) The proportion of simulations where the correlations of the blue crab feature were less than the correlation of the orange crab features was 7.052. On average the orange crabs have a larger width and depth compared to blue crabs. Their features are also more correlated than that of blue crabs.

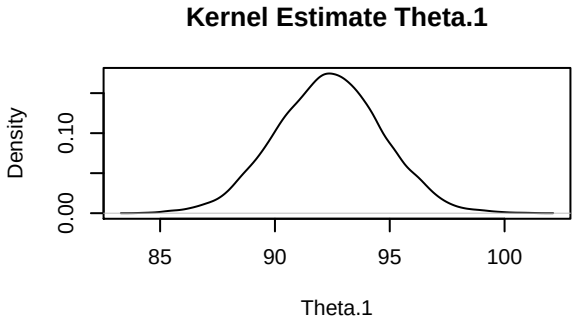
Missing Data

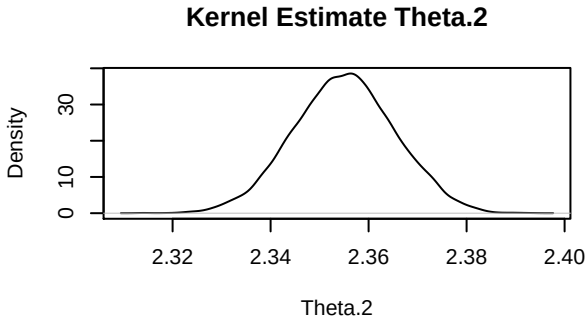
Summary of Mean Vectors

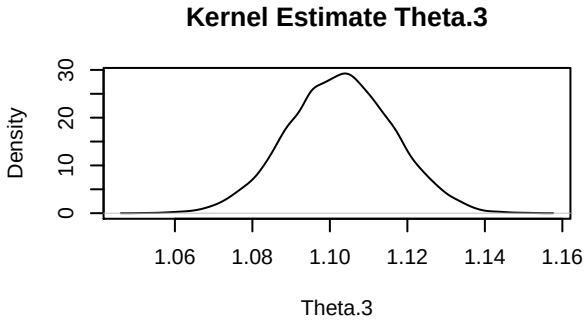




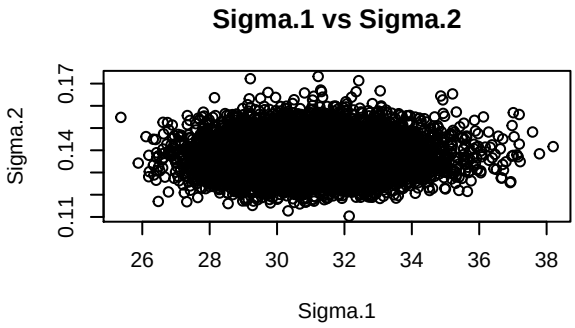


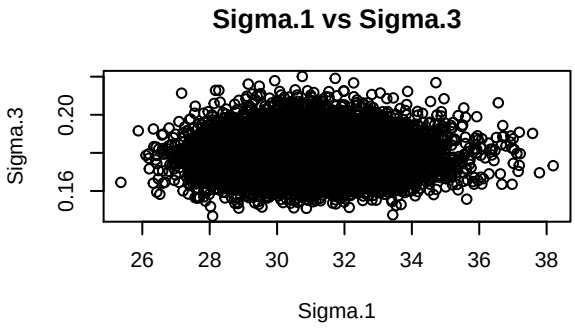


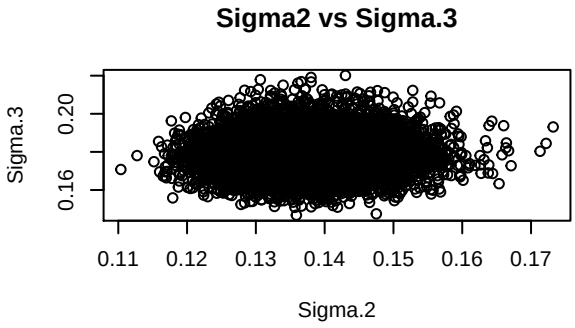


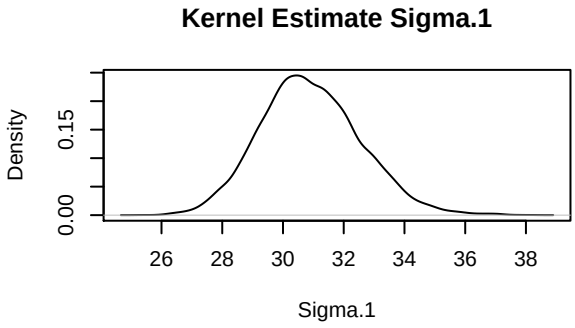


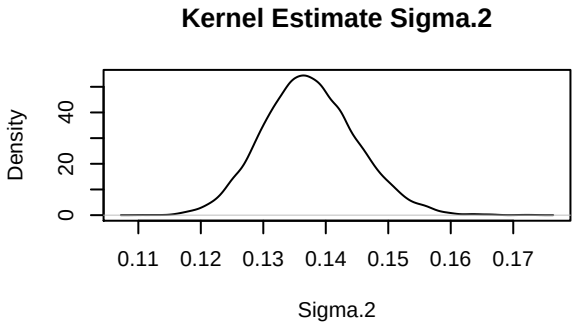
Summary of SD Vectors

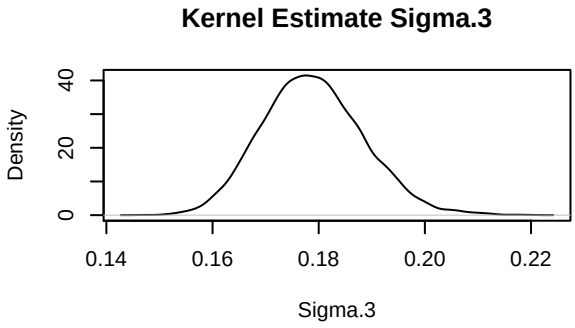






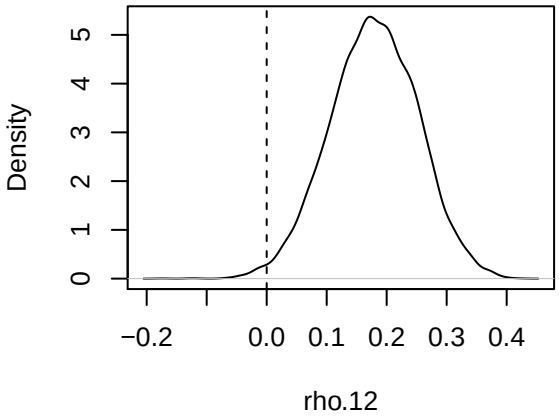




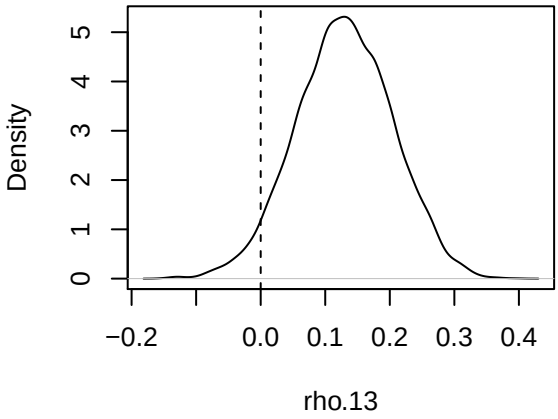


Are any of the variables correlated?

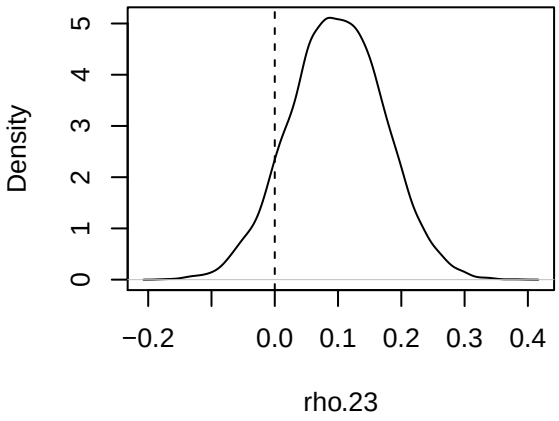
P(rho.12 < 0) = 0.0085



P(rho.13 < 0) = 0.0413



P(rho.23 < 0) = 0.0942



For the most part it appears that the two estimates are very close, but for missing value 1, the confidence interval built from the estimated parameters appears to be wider than the confidence interval built from the simulated values. For the second missing value, the confidence interval from the simulated values appear to be wider than the confidence interval construted from the parameter estimates. This could be because there is some correlation between theta.2 and theta.3

```
Missing Value 1:

Credible Region from Predicted Missing Values:

     2.5%     97.5%
0.7392187 1.4416295

Credible Region from mean +/- 1.96 * sd of estimated parameters:

[1] 0.7515836 1.4539277

Missing Value 2:

Credible Region from Predicted Missing Values:

    2.5%    97.5%
2.117770 2.633322

Credible Region from mean +/- 1.96 * sd of estimated parameters:

[1] 2.085259 2.624835
```