2

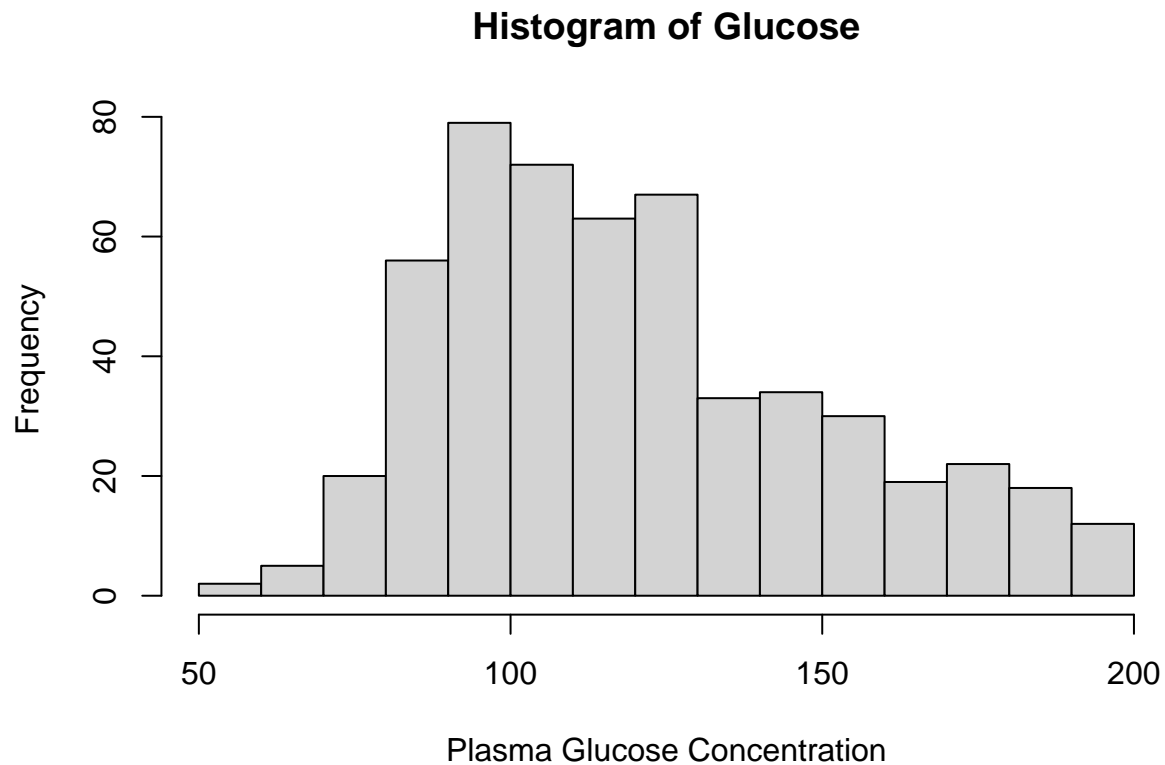**(a)**

```
glucose <- read.table("glucose.dat")
hist(glucose$V1, xlab = "Plasma Glucose Concentration",main = "Histogram of Glucose")
```

Histogram of Glucose

We can see from the histogram that the data is skewed to the right rather than noramlly distributed bell shape.

(c)

```
library(coda)
Y = glucose$V1
n = length(Y)
# set some values
a = b = 1
mu0 = 120
t20 = 200
s20 = 1000
nu0 = 10
S = 10000
THETA1 = numeric(S)
THETA2 = numeric(S)
YPRED = numeric(S)
p = 0.5
theta1  = mean(Y)
```

```r
theta2 = theta1
s21  = var(Y)
s22 = s21
# Gibbs
for (s in 1:S) {
  #sample x,p,thetas,sigma sqaures
  p1 = p * dnorm(Y, theta1, sqrt(s21))
  p2 = (1 - p) * dnorm(Y, theta2, sqrt(s22))
  bernp = p1 / (p1 + p2)
  X = rbinom(n, 1, bernp)
  n1 = sum(X)
  n2 = n - n1
  y1 = Y[X == 1]
  y2 = Y[X == 0]
  ybar1 = mean(y1)
  ybar2 = mean(y2)
  yvar1 = var(y1)
  yvar2 = var(y2)
  p = rbeta(1, a + n1, b + n2)
  t2n1 = 1 / (1 / t20 + n1 / s21)
  mun1 = (mu0 / t20 + n1 * ybar1 / s21) / (1 / t20 + n1 / s21)
  theta1 = rnorm(1, mun1, sqrt(t2n1))
  t2n2 = 1 / (1 / t20 + n2 / s22)
  mun2 = (mu0 / t20 + n2 * ybar2 / s22) / (1 / t20 + n2 / s22)
  theta2 = rnorm(1, mun2, sqrt(t2n2))
  nun1 = nu0 + n1
  s2n1 = (nu0 * s20 + (n1 - 1) * yvar1 + n1 * (ybar1 - theta1)^2) / nun1
  s21 = 1 / rgamma(1, nun1 / 2, s2n1 * nun1 / 2)
  nun2 = nu0 + n2
  s2n2 = (nu0 * s20 + (n2 - 1) * yvar2 + n2 * (ybar2 - theta2)^2) / nun2
  s22 = 1 / rgamma(1, nun2 / 2, s2n2 * nun2 / 2)

  # posterior
  xpred = runif(1) < p
  ypred = ifelse(xpred, rnorm(1, theta1, sqrt(s21)), rnorm(1, theta2, sqrt(s22)))
  THETA1[s] = theta1
  THETA2[s] = theta2
  YPRED[s] = ypred
}
# plot the autocorrelation functions and calculate effective sample sizes
theta_min = pmin(THETA1, THETA2)
theta_max = pmax(THETA1, THETA2)
acf(theta_min,main="Theta Min")
```
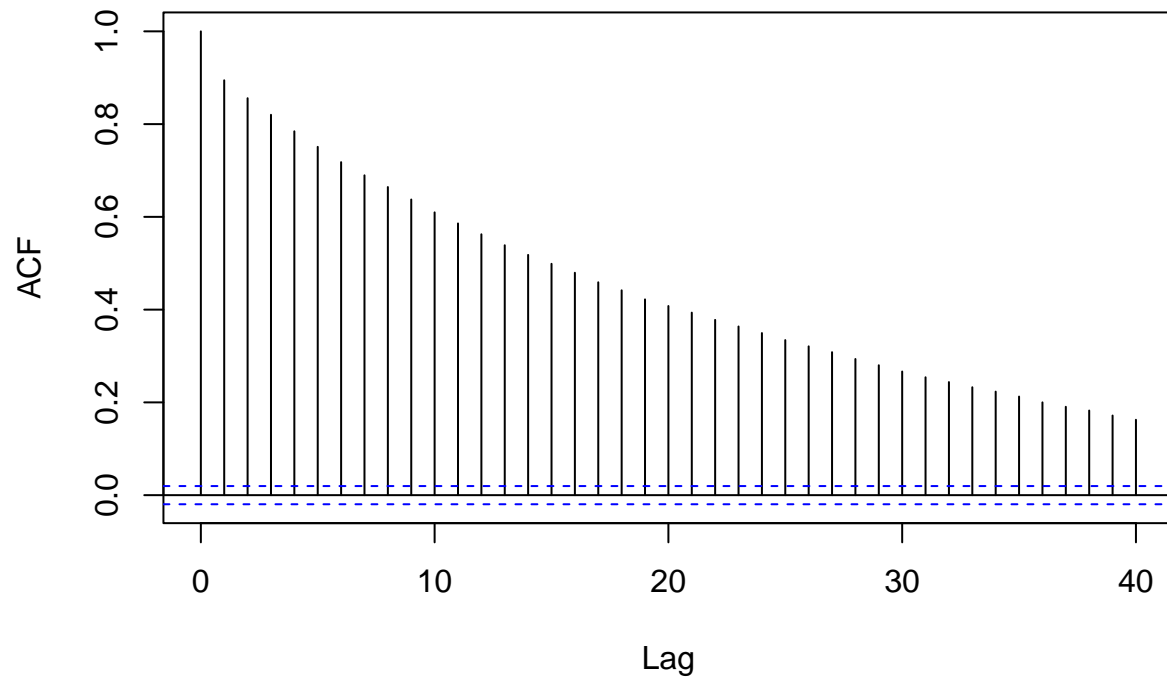
Theta Min
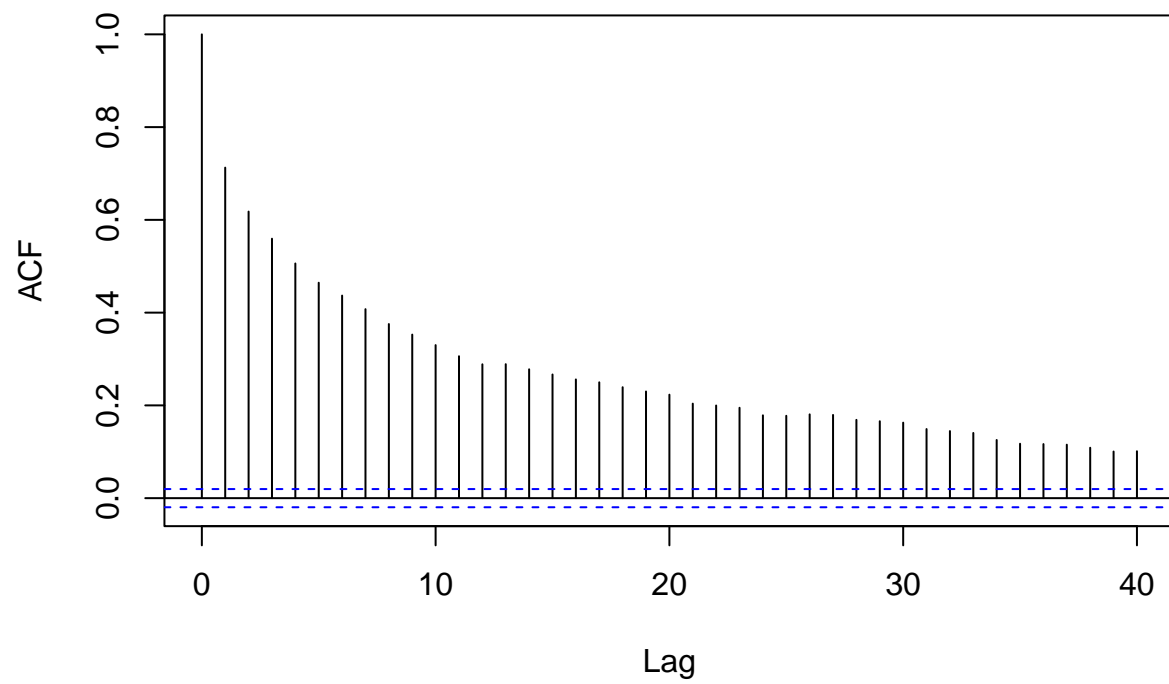

```
acf(theta_max,main="Theta Max")
```

Theta Max


```
effectiveSize(theta_min)
```

```
var1
364.998
```

```
effectiveSize(theta_max)
```

```
##      var1
## 219.6368
```

The plot of the autocorrelation functions and their effective sample sizes are above.

(d)

```
library(ggplot2)
pred = data.frame(y = YPRED, data = 'predictive')
orig = data.frame(y = Y, data = 'original')
Y_data = rbind(pred,orig)
ggplot(Y_data, aes(x = y, fill = data)) +
  geom_density(alpha = 0.66)+
  scale_fill_manual(values=c("navy","#E69F00"))+
  ggtitle("Density plot for Comparison of Original Y and Predicted Y")
```

Density plot for Comparison of Original Y and Predicted Y
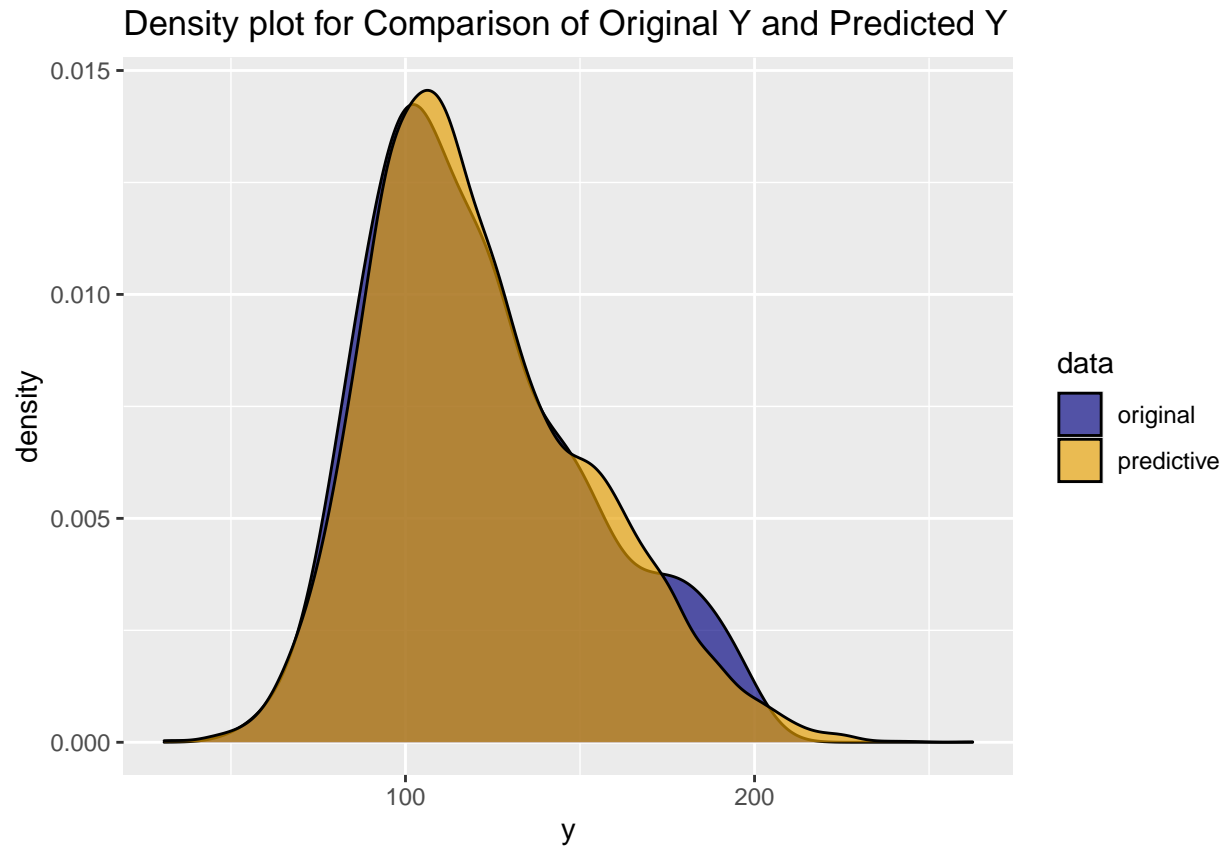
From the plot above, we can see that the densitiy of original dataset and the predicted posterior dataset are very similar to each other which indicates that our mixture model fits well.