Take Home Exam 2

Ben Buzzee

November 6, 2020

Problem 1

Formulate semiconjugate priors

Since we haven't looked at the data, we don't have much information to incorporate into the priors. Using a vaguely informed guess for the ages and very large values for the variances, I would give μ the prior of MVN(μ, Λ) where $\mu = (35, 35)^T$ and

$$\Lambda = \begin{pmatrix} 400 & 0 \\ 0 & 400 \end{pmatrix}$$

This errs on the side of being too vague, and gives both elements of μ a standard deviation of 20. Although I believe there is probably a positive correlation in ages, I'd rather let the data speak for itself rather than assume it at the beginning. Maybe this is a non-random sample and there is a negative correlation in ages.

Similarly for Σ, I would use an IW(S_0^-1, ν_0) prior where S_0^-1 is a 2x2 Identity matrix*100 and $\nu_0 = 4$. I believe a variance of 10 is plausible and again make no judgments about covariances.

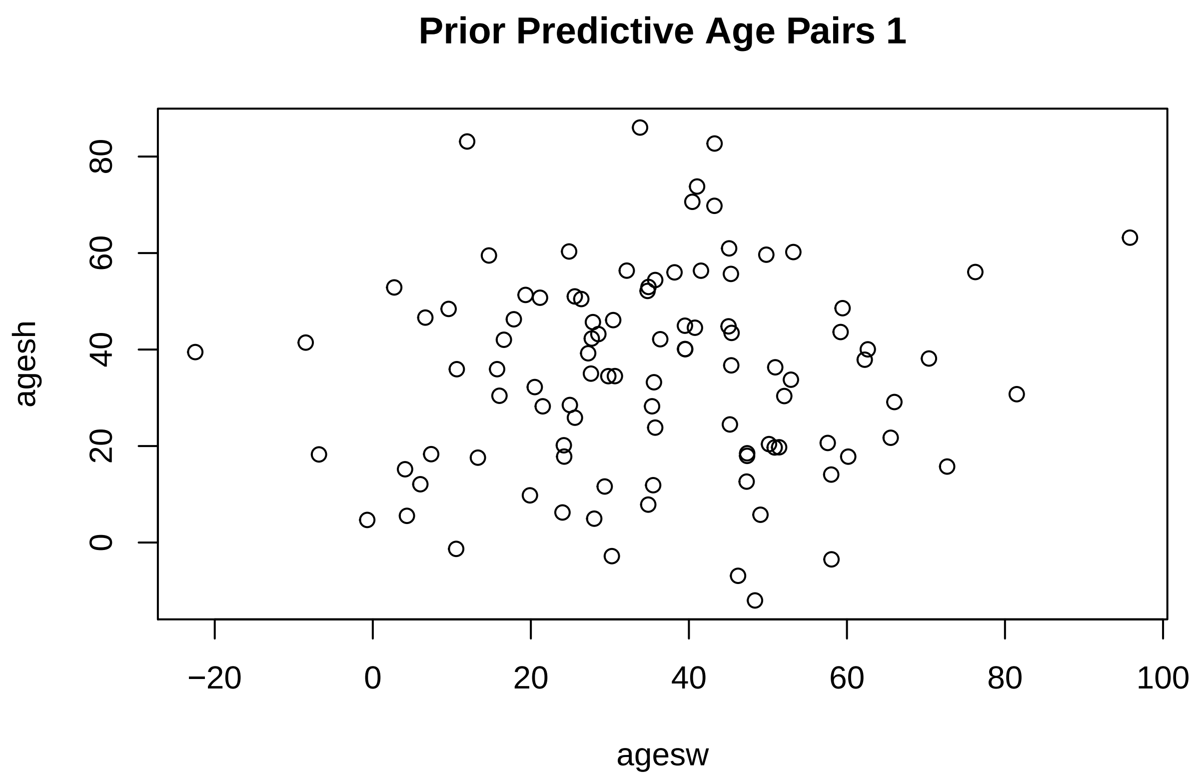
Problem 2

Generate a prior predictive set of size n = 100.

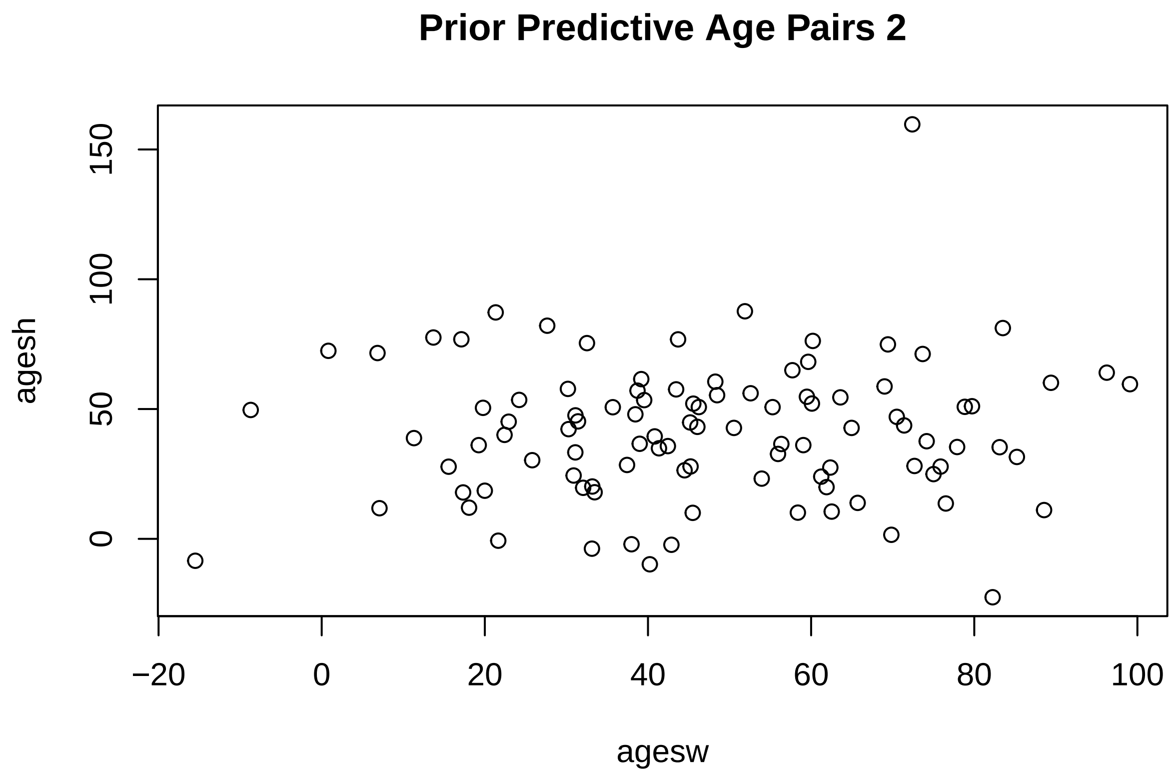
We will use the following modified code to do this:

```r
library(MASS) # for mvrnorm;  MASS is in the default R installation
set.seed(1729)
n <- 100 # sample of this many couples?

# prior for (mu.h,mu.w):  N( mu0, Lambda0 )
mu0 <- c( 35,35 )  # mean(h,w)
Lambda0 <- matrix( c(400,0,0,400), nrow=2,ncol=2)
Lambda0.inv <- solve(Lambda0)

# prior for Sigma:  IW( nu.0, S0.inv )
v.h.0 <- 100 # 186 # 10
v.w.0 <- 100 # 164 # 10
S0 <- matrix( c( v.h.0, 0,0, v.w.0 ),
              nrow=2,ncol=2 )
S0.inv <- solve(S0)
nu0 <- 4 # wide prior for covariance matrix?

my.data <- list( S0=S0, nu0=nu0,
                 mu0=mu0, Lambda0.inv=Lambda0.inv)

my.inits <- list( mu=c(35,35) )
my.fname <- "prior_model.txt"
```

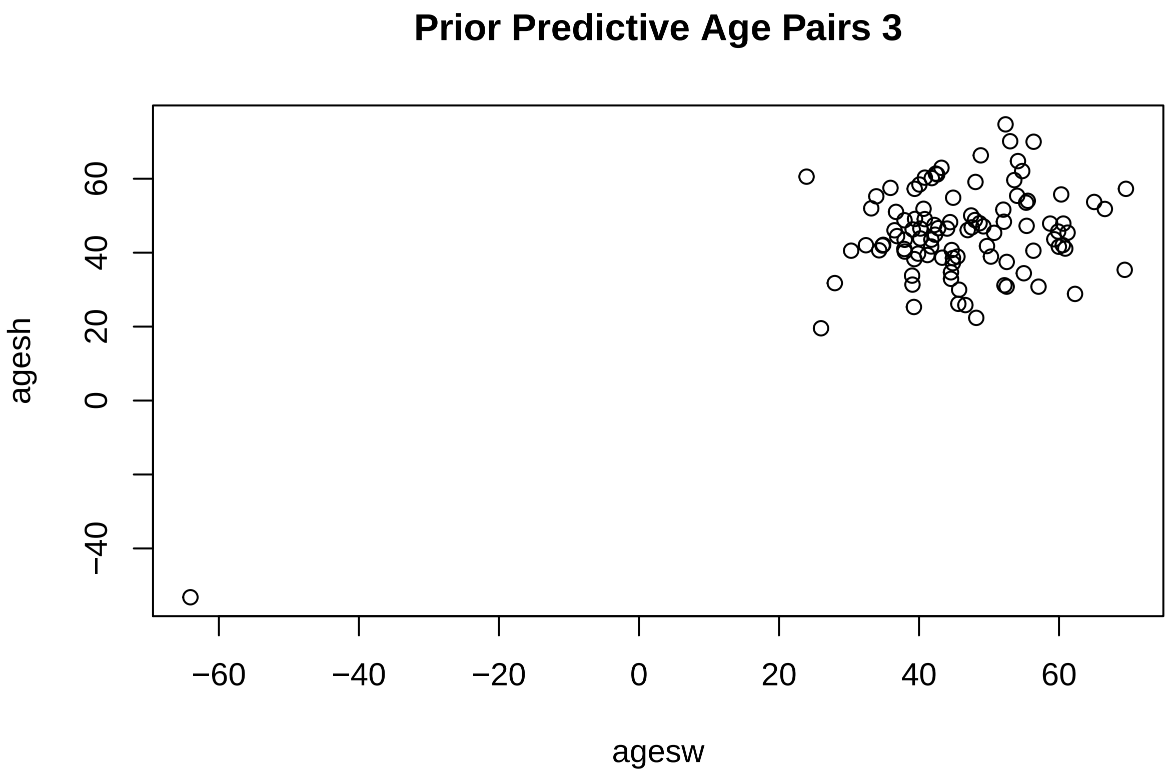
```r
library(rjags)
my.jags.model <- jags.model(
  file=my.fname, data=my.data, inits=my.inits,
  n.chains=1, n.adapt=1000, quiet=FALSE)
my.variables <- c("mu","Sigma","ypred")
my.coda.samples <- coda.samples(my.jags.model,
                                my.variables, 1200)

# Extract just a few samples, and this will be one simulated data set:
which <- seq( 100,1200, length=n )
agesh <- my.coda.samples[[1]][,"ypred[1]"][which] ##############
agesw <- my.coda.samples[[1]][,"ypred[2]"][which] ##############
```



The above is the first prior predictive dataset produced. The core of the data seem plausible, but there are some very unrealistic extreme cases, including marriages in which one of the spouses is under age 0. This can be fixed by tightening the prior and or increasing the prior mean ages. Lets try adjusting the mean age only to 45:



We still have ages below zero, now lets try also reducing the prior variances:



This scatterplot seems much more plausible. The final parametrization is $\mu \sim \text{MVN}(\mu, \Lambda)$ where $\mu = (45, 45)^T$ and $\Lambda = 75$*Identity Matrix. S_0 is 50 times the identity matrix, and $\nu_0 = 4$.

Problem 3

Find the Joint Prior and Correlation

```
n <- 100 # sample of this many couples?

# prior for (mu.h,mu.w):  N( mu0, Lambda0 )
mu0 <- c( 40,39 )  # mean(h,w)
Lambda0 <- matrix( c(25,20,20,25), nrow=2,ncol=2 )
Lambda0.inv <- solve(Lambda0)

# prior for Sigma:  IW( nu.0, S0.inv )
v.h.0 <- 10
v.w.0 <- 10
rho.0 <- 0.1
cov.0 <- rho.0*sqrt( v.h.0 * v.w.0 )
S0 <- matrix( c( v.h.0, cov.0,cov.0, v.w.0 ),
              nrow=2,ncol=2 )
S0.inv <- solve(S0)
nu0 <- 4
```

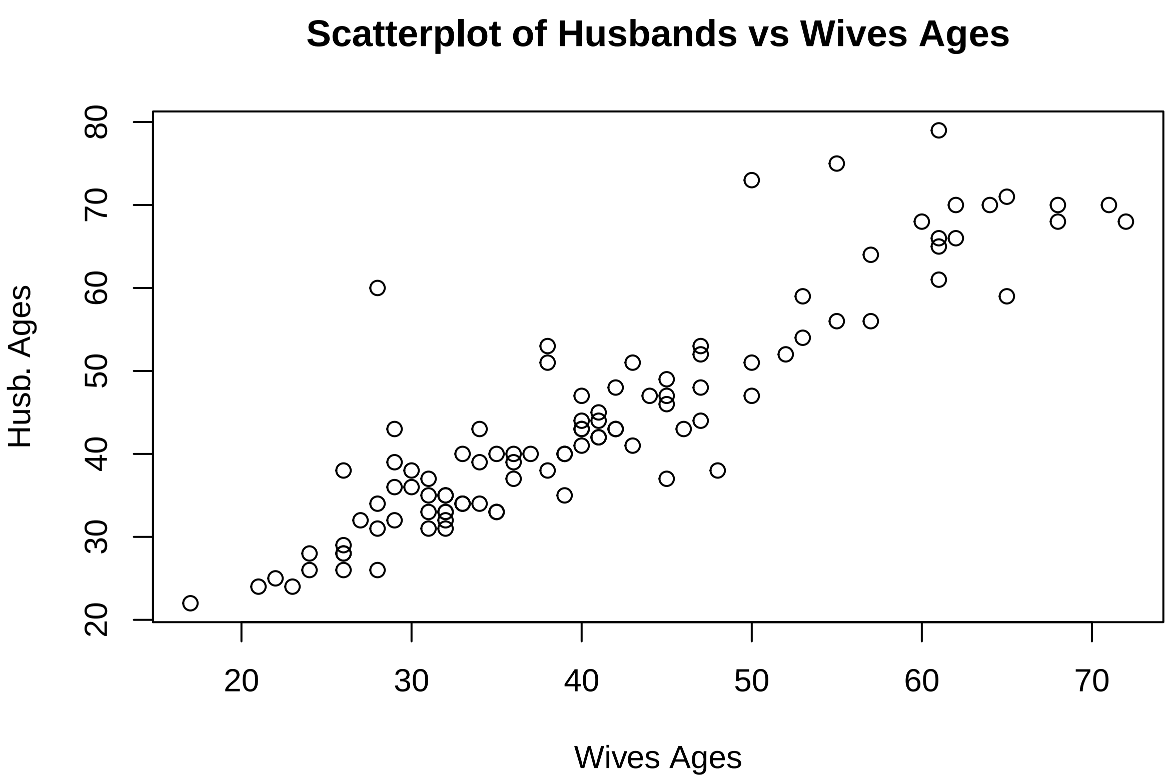
In the above code, the joint prior distribution for (μ_h, μ_w) is a multivarate normal distribution with mean $(40, 39)$ and

$$\Lambda = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix}$$

The prior correlation is $\frac{COV(\mu_h,\mu_w)}{\sqrt{(Var(\mu_h)Var(\mu_w))}} = \frac{20}{25} = .8$

Problem 4

Read in and examine the data



There appears to be a very strong correlation (.9035) between husband and wife ages, which I did not incorporate into my priors. Husbands ages range from 22 to 79, with a mean age of 44.42. Wives ages range from 17 to 72, with a mean age of 40.89.

Problem 5

Fit the full model and summarize results

```
library(MASS)
set.seed(17291)
```

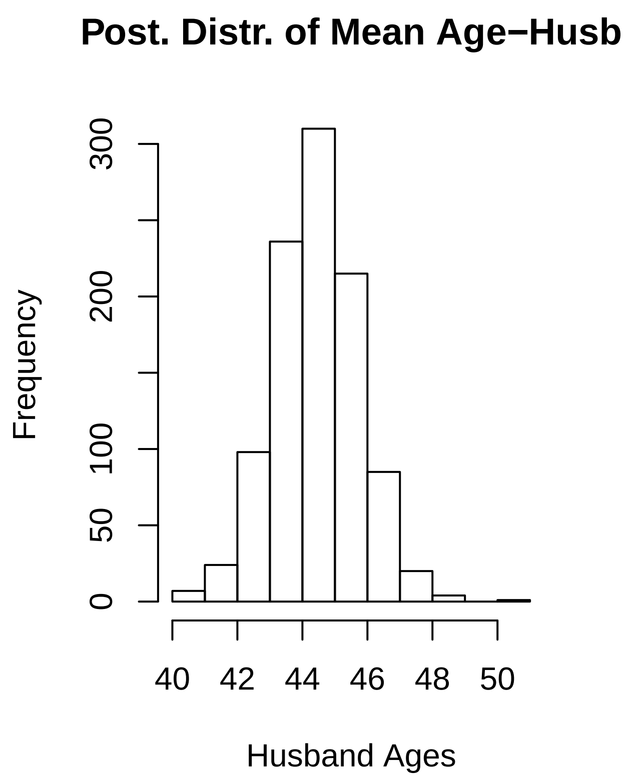
```r
n <- 100 # sample of this many couples?

# prior for (mu.h,mu.w)
N <- nrow(dat)
mu0 <- c( 45,45 )  # mean(h,w)
Lambda0 <- matrix( c(75,0,0,75), nrow=2,ncol=2)
Lambda0.inv <- solve(Lambda0)

# prior for Sigma:  IW( nu.0, S0.inv )
v.h.0 <- 50
v.w.0 <- 50
S0 <- matrix( c( v.h.0, 0,0, v.w.0 ),
              nrow=2,ncol=2 )
S0.inv <- solve(S0)
nu0 <- 4

my.data <- list( S0=S0, nu0=nu0,
                 mu0=mu0, Lambda0.inv=Lambda0.inv)

my.inits <- list( mu=c(45,45) )
my.fname <- "post_model.txt"
my.data <- list(y = dat, S0 = S0,
                N = nrow(dat),
                nu0 = 4,
                Lambda0.inv = Lambda0.inv,
                mu0 = mu0)

library(rjags)

my.jags.model <- jags.model(
                    file=my.fname,
                    data=my.data,
                    inits=my.inits,
                    n.chains=1,
                    n.adapt=1000,
                    quiet=FALSE)

my.variables <- c("mu","Sigma","ypred", "rho")
my.coda.samples <- coda.samples(my.jags.model,
                                my.variables,
                                n.iter = 10000,
                                thin = 10)

muh <- my.coda.samples[[1]][,5]
muw <- my.coda.samples[[1]][,6]
rho <- my.coda.samples[[1]][,7]

sigma11 <- my.coda.samples[[1]][,1]
sigma22 <- my.coda.samples[[1]][,4]
sigma12 <- my.coda.samples[[1]][,2]

# head(my.coda.samples)
```

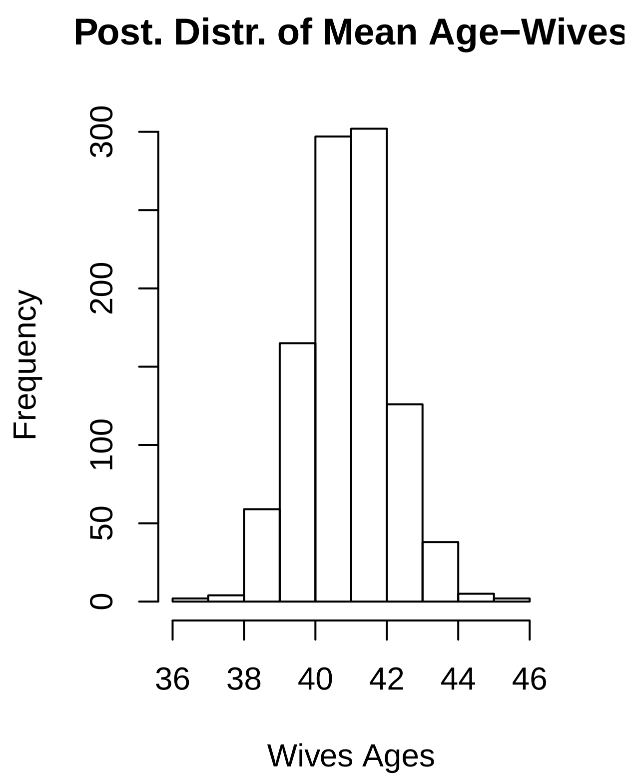
Results

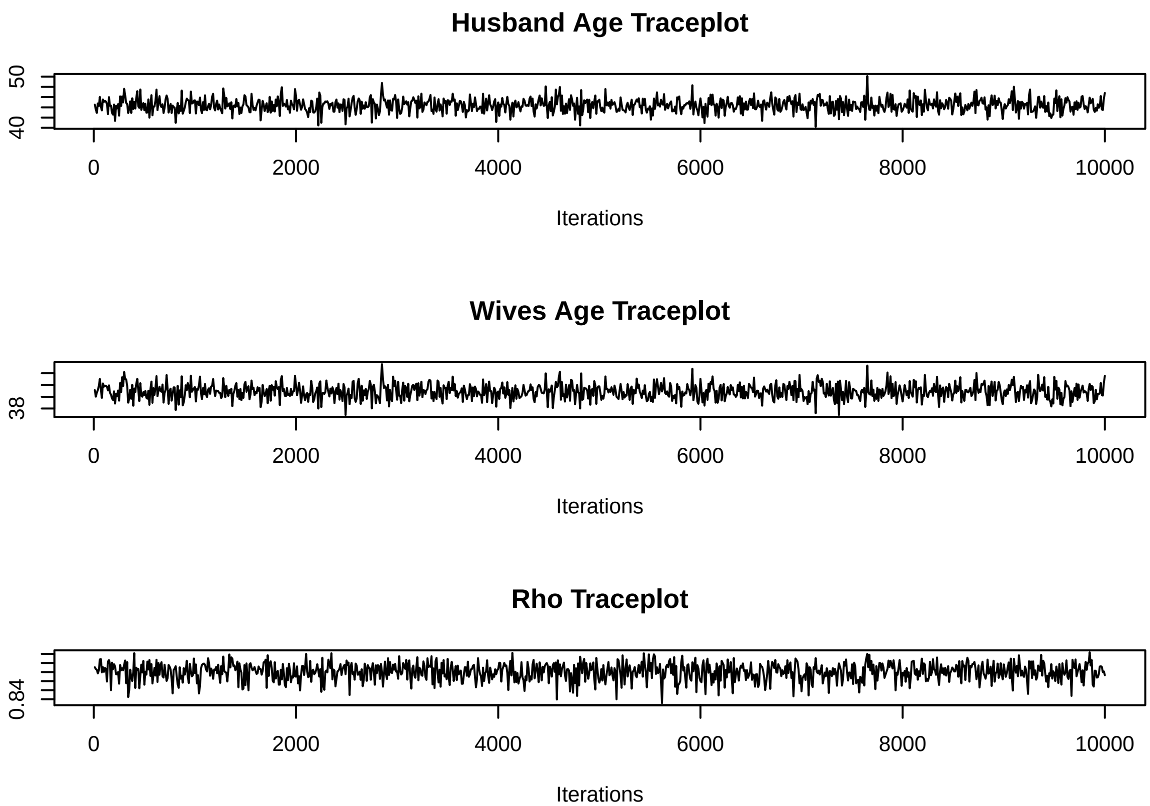
Summary Statistics

title	mean_post_age	standard_dev	lower_95_credible_bound	upper_95_credible_bound
husband	44.5	1.300	41.900	47.000
wife	41.0	1.200	38.500	43.300
correlation	0.9	0.019	0.856	0.933

Graphs

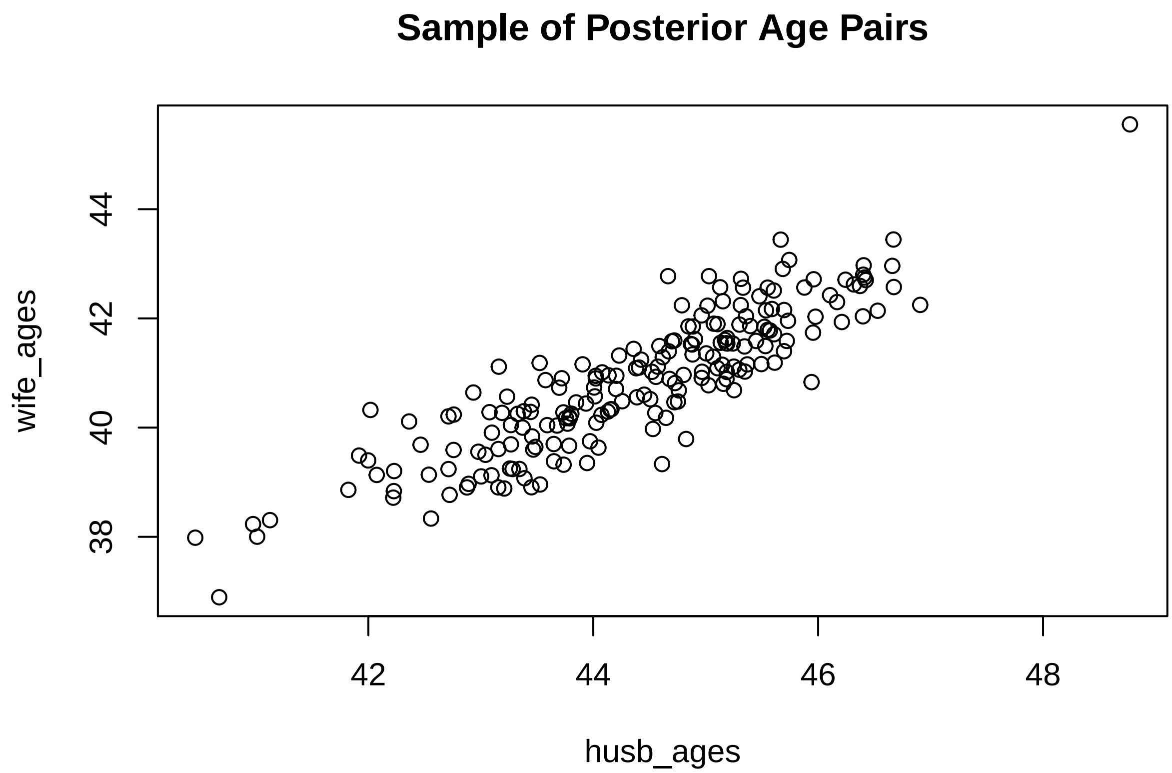






All three trace plots look good. Neither has any noticeable trend.

```r
par(mfrow = c(1,1))
husb_ages <- muh[200:400]
wife_ages <- muw[200:400]

plot(husb_ages, wife_ages, main = "Sample of Posterior Age Pairs")
```



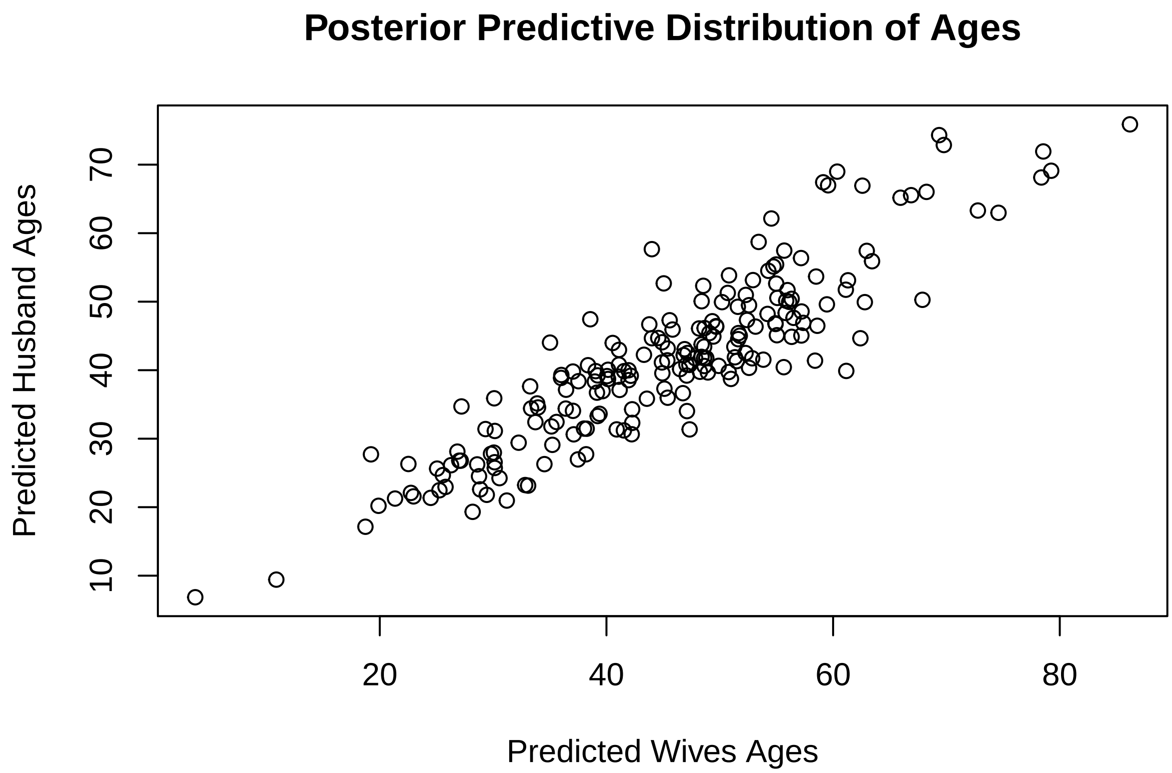
The scatterplot of the posterior samples looks very similar to our original data.

Correlation

From the above results table, we see that the correlation between the husband and wives ages is .9. This is very close to what we observed in the original dataset.

Problem 6

```
# head(my.coda.samples)

ypred1 <- my.coda.samples[[1]][,8][200:400]
ypred2 <- my.coda.samples[[1]][,9][200:400]


plot(ypred1, ypred2, xlab = "Predicted Wives Ages", ylab = "Predicted Husband Ages", main = "Posterior P
```



The predictive distribution is much more spread out than the posterior distribution. This can be attributed to the higher variance associated with predicting future values. The correlation seems roughly the same.

Problem 7

Create credible intervals when one spouses age is known

First we will find a credible interval for husbands ages when the wife is 43 years old:

```
set.seed(17291)


n <- 100

# prior for (mu.h,mu.w)
N <- nrow(dat)
mu0 <- c( 45,45 )   # mean(h,w)
Lambda0 <- matrix( c(75,0,0,75), nrow=2,ncol=2)
Lambda0.inv <- solve(Lambda0)

# prior for Sigma:  IW( nu.0, S0.inv )
v.h.0 <- 50
v.w.0 <- 50
S0 <- matrix( c( v.h.0, 0,0, v.w.0 ),
              nrow=2,ncol=2 )
```

```r
S0.inv <- solve(S0)
nu0 <- 4


my.inits <- list( mu=c(45,45) )
my.fname <- "post_model_y2.txt"
my.data <- list(y = dat, S0 = S0,
                N = nrow(dat),
                nu0 = 4,
                Lambda0.inv = Lambda0.inv,
                mu0 = mu0,
                y2 = 43) # add y2 to data list


my.jags.model <- jags.model(
                    file=my.fname,
                    data=my.data,
                    inits=my.inits,
                    n.chains=1,
                    n.adapt=1000,
                    quiet=FALSE)

my.variables <- c("mu","Sigma","ypred", "y1") # add y1 to variable list
my.coda.samples <- coda.samples(my.jags.model,
                                my.variables,
                                n.iter = 10000,
                                thin = 10)




y1 <- my.coda.samples[[1]][,7]

quantile(y1, probs = c(.025, .975))
```

And we find the credible interval for husbands ages when wives are age 43 is [35, 57.8]. Next we will do the same thing, except find an interval for wives ages when the husband is 43:

```r
library(MASS) # for mvrnorm;  MASS is in the default R installation
set.seed(17291)


n <- 100 # sample of this many couples?

# prior for (mu.h,mu.w)
N <- nrow(dat)
mu0 <- c( 45,45 )
Lambda0 <- matrix( c(75,0,0,75), nrow=2,ncol=2)
Lambda0.inv <- solve(Lambda0)

# prior for Sigma:  IW( nu.0, S0.inv )
v.h.0 <- 50
v.w.0 <- 50
S0 <- matrix( c( v.h.0, 0,0, v.w.0 ),
              nrow=2,ncol=2 )
```

```r
S0.inv <- solve(S0)
nu0 <- 4


my.inits <- list( mu=c(45,45) )
my.fname <- "post_model_y1_known.txt"
my.data <- list(y = dat, S0 = S0,
                N = nrow(dat),
                nu0 = 4,
                Lambda0.inv = Lambda0.inv,
                mu0 = mu0,
                y1 = 43) # add y1 here

library(rjags)

my.jags.model <- jags.model(
                    file=my.fname,
                    data=my.data,
                    inits=my.inits,
                    n.chains=1,
                    n.adapt=1000,
                    quiet=FALSE)

my.variables <- c("mu","Sigma","ypred", "y2") # add y2 to variables
my.coda.samples <- coda.samples(my.jags.model,
                                my.variables,
                                n.iter = 10000,
                                thin = 10)


y2 <- my.coda.samples[[1]][,7]

quantile(y2, probs = c(.025, .975))
```

The credible interval for wives ages when husbands are age 43 is [28.11, 50.7].