Sierra Bainter
Lab HW 8
Due 3/9/13

a) I believe a sensible prior for the ages of married couples sampled from the U.S. population should place the majority of ages between 18-95, with a mean age near 55 (slightly higher for men than women), correlated around .7.

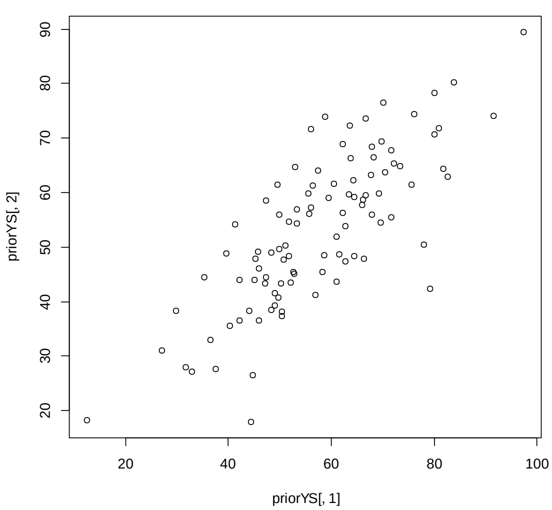
In line with these beliefs, I initially chose a semi-conjugate prior distribution with $\boldsymbol{\theta} = \begin{pmatrix} \theta_h \\ \theta_w \end{pmatrix} \sim N_2 \begin{pmatrix} 56 \\ 53 \end{pmatrix} \begin{pmatrix} 100 & 70 \\ 70 & 100 \end{pmatrix}$

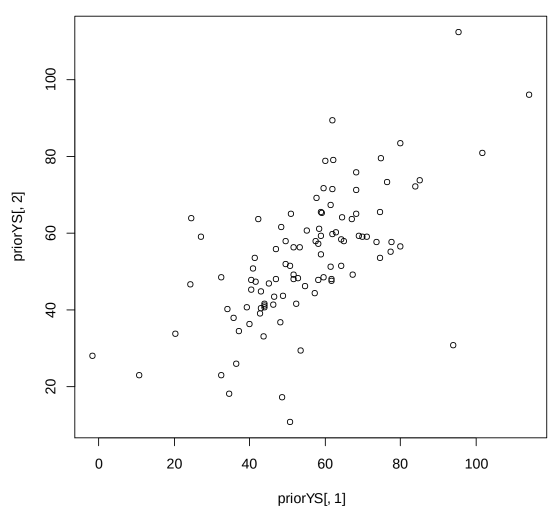
$$\Sigma \sim Inv-Wishart\left(4, \begin{pmatrix} 361 & 252.7 \\ 252.7 & 361 \end{pmatrix}^{-1}\right)$$

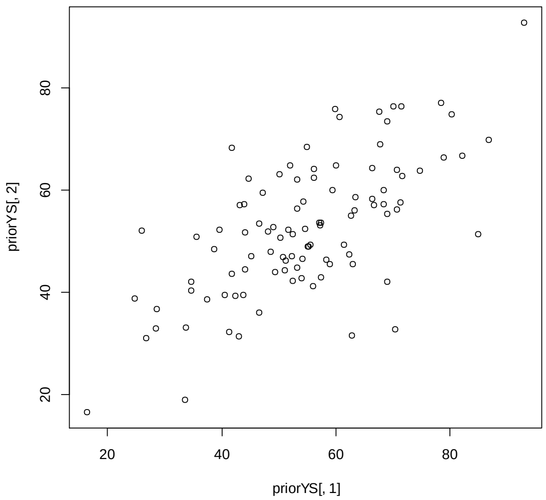
b) I generated prior predictive datasets from this prior distribution. The results were more variable than I expected, and there were frequently outliers that did not make any sense for the age data. After several iterations of adjusting the priors and inspecting scatterplots, I decided on the same θ distribution as in a) but set

$$\Sigma \sim Inv-Wishart\left(4, \begin{pmatrix} 196 & 137.2 \\ 137.2 & 196 \end{pmatrix}^{-1}\right)$$

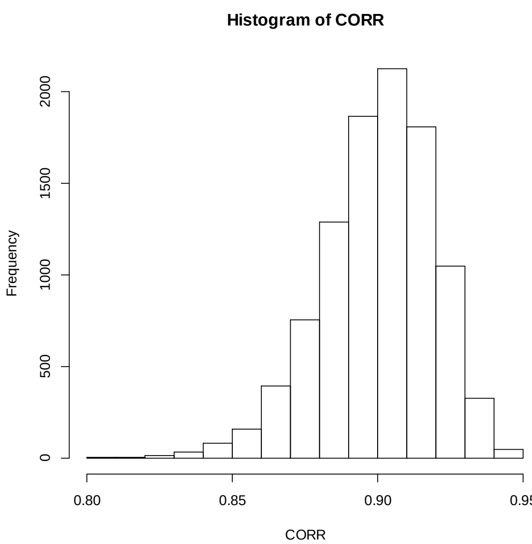
Three example scatterplots are shown below.

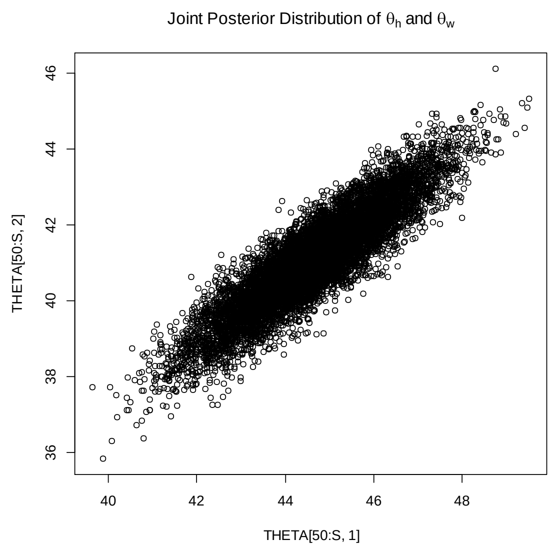






c) Using my prior distribution and the agehw.dat data, I used Gibbs sampling to obtain an MCMC approximation of 10,000 samples from the posterior distribution. I assessed convergence by inspecting trace plots. Because the chain started in a high density area for both parameters, convergence was almost immediate, however to be safe I discarded the first 50 iterations as burn-in for the following summary statistics.





The 95% posterior confidence interval for the mean age was between 38.60 and 43.60 for wives and between 41.96 and 47.32 for husbands. The 95% posterior confidence interval for the correlation between the ages of husbands and wives was between .858 and .932.

d) Using a diffuse prior, the posterior confidence intervals were slightly wider and shifted to slightly smaller values, with the mean age between 38.33 and 43.44 for wives and between 41.69 and 47.12 for husbands. The 95% posterior confidence interval for the correlation between the ages of husbands and wives was between .792 and .899. The distribution of the correlation was more effected by the diffuse prior than the distributions of the means.

e) The proportion of samples from the posterior distribution with the mean age of husbands greater than the mean age of wives was 1, $P[\theta_h > \theta_w \mid y_1, \ldots, y_{100}] = 1$. This is a strong weight of evidence that the mean age of husbands is greater. The average difference between the two means was 3.52, and a 95% confidence interval for the difference was [2.08, 4.96].

R Code:

```
# source("http://www.stat.washington.edu/~hoff/Book/Data/data/chapter7.r")
Y<-read.table("c:/users/hugh/documents/dropbox/Bayes/Lab/agehw.dat", header=T)

mu0<-c(0,0)
L0<-matrix( c(100000,0,0,100000),nrow=2,ncol=2)
nu0<-3
S0<-matrix( c(1000,0,0,1000),nrow=2,ncol=2)

#draws from prior predictive
priorYS<-NULL
for(t in 1:100)
{
thetap<-rmvnorm(1,mu0,L0)
Sigma<-solve( rwish(1, nu0, solve(S0)) )
priorYS<-rbind(priorYS,rmvnorm(1,thetap,Sigma))
}
plot(priorYS[,1],priorYS[,2])

n<-dim(Y)[1] ; ybar<-apply(Y,2,mean)
Sigma<-cov(Y) ; THETA<-SIGMA<-NULL
YS<-NULL
CORR<-NULL
set.seed(1)
S<-10000
for(s in 1:S)
{
  ###update theta
  Ln<-solve( solve(L0) + n*solve(Sigma) )
  mun<-Ln%*%( solve(L0)%*%mu0 + n*solve(Sigma)%*%ybar )
  theta<-rmvnorm(1,mun,Ln)
  ###update Sigma
  Sn<- S0 + ( t(Y)-c(theta) )%*%t( t(Y)-c(theta) )
  #  Sigma<-rinvwish(1,nu0+n,solve(Sn))
  Sigma<-solve( rwish(1, nu0+n, solve(Sn)) )
    ### save results
  YS<-rbind(YS,rmvnorm(1,theta,Sigma))
  THETA<-rbind(THETA,theta) ; SIGMA<-rbind(SIGMA,c(Sigma))
}

plot(SIGMA[,4], type="l", col="darkblue", ylab=expression(theta))
plot(THETA[50:S,1],THETA[50:S,2], main=expression(paste ("Joint Posterior Distribution of " ,theta[h], "
and ", theta[w])))

CORR<-SIGMA[50:S,2]/sqrt(SIGMA[50:S,1]*SIGMA[50:S,4])
hist(CORR)
quantile( SIGMA[,2]/sqrt(SIGMA[,1]*SIGMA[,4]), prob=c(.025,.5,.975) )
quantile( THETA[50:S,2], prob=c(.025,.5,.975) )
quantile( CORR, prob=c(.025,.5,.975) )

quantile(   THETA[50:S,1]-THETA[50:S,2], prob=c(.025,.5,.975) )
mean( THETA[50:S,1]-THETA[50:S,2])
mean( THETA[50:S,1]>THETA[50:S,2])
```