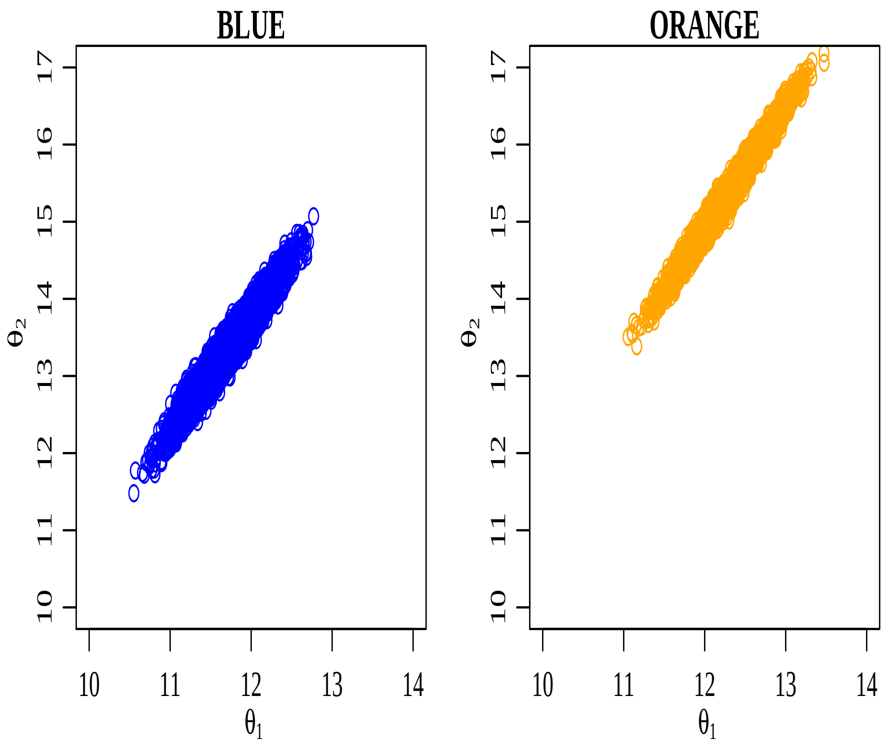
Introduction to Bayesian Data Analysis
Tutorial 7 - Solutions

(1) (a) The R code is:

```
blue_crab<-read.table("bluecrab.dat",header=F)
orange_crab<-read.table("orangecrab.dat",header=F)

y1_blue<-blue_crab[,1]
y2_blue<-blue_crab[,2]
y1_orange<-orange_crab[,1]
y2_orange<-orange_crab[,2]

mu0_blue<-apply(blue_crab,2,mean)
mu0_orange<-apply(orange_crab,2,mean)

L0_blue<-cov(blue_crab)
L0_orange<-cov(orange_crab)

S0_blue<-L0_blue
S0_orange<-L0_orange

nu0_blue<-nu0_orange<-4

n_blue<-dim(blue_crab)[1]
n_orange<-dim(orange_crab)[1]

ybar_blue<-mu0_blue
ybar_orange<-mu0_orange

Sigma_blue<-S0_blue
Sigma_orange<-S0_orange
```

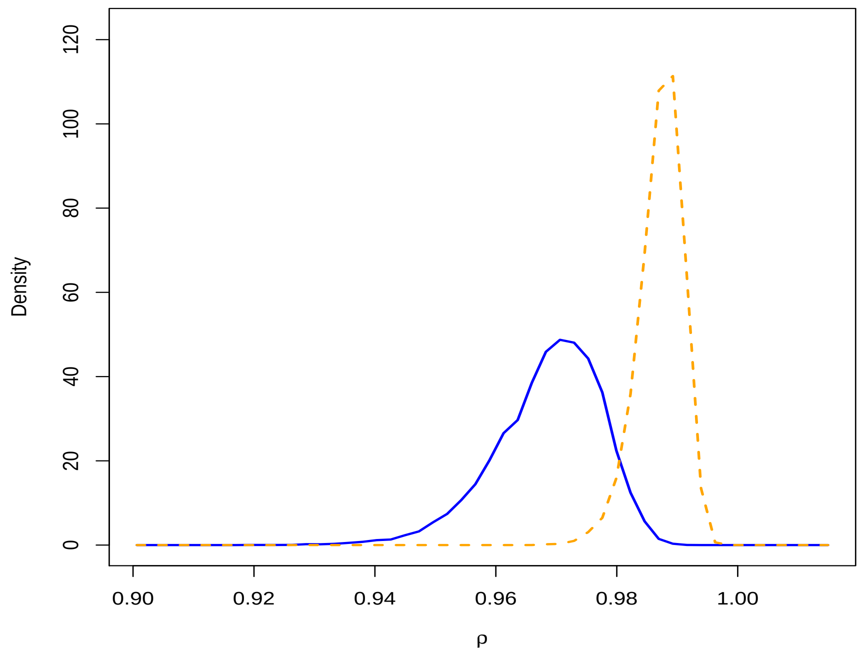
```
THETA_blue<-THETA_orange<-SIGMA_blue<-SIGMA_orange<-NULL
YS_blue<-YS_orange<-NULL
set.seed(1)
for (s in 1:10000)
{
 ###update theta
  Ln_blue<-solve( solve(L0_blue) + n_blue*solve(Sigma_blue) )
  mun_blue<-Ln_blue%*%( solve(L0_blue)%*%mu0_blue +
        n_blue*solve(Sigma_blue)%*%ybar_blue )
  theta_blue<-rmvnorm(1,mun_blue,Ln_blue)

  Ln_orange<-solve( solve(L0_orange) + n_orange*solve(Sigma_orange) )
  mun_orange<-Ln_orange%*%( solve(L0_orange)%*%mu0_orange +
         n_orange*solve(Sigma_orange)%*%ybar_orange )
  theta_orange<-rmvnorm(1,mun_orange,Ln_orange)
  ###

  ###update Sigma
  Sn_blue<- S0_blue +
      ( t(blue_crab)-c(theta_blue) )%*%t( t(blue_crab)-c(theta_blue) )
  Sigma_blue<-solve( rwish( nu0_blue+n_blue, solve(Sn_blue)) )

 Sn_orange<- S0_orange + ( t(orange_crab)-c(theta_orange) )%*%
             t( t(orange_crab)-c(theta_orange) )
  Sigma_orange<-solve( rwish( nu0_orange+n_orange, solve(Sn_orange)) )
  ###

  ###
  YS_blue<-rbind(YS_blue,rmvnorm(1,theta_blue,Sigma_blue))
   YS_orange<-rbind(YS_orange,rmvnorm(1,theta_orange,Sigma_orange))

  ### save results
THETA_blue<-rbind(THETA_blue,theta_blue) ;
SIGMA_blue<-rbind(SIGMA_blue,c(Sigma_blue))
THETA_orange<-rbind(THETA_orange,theta_orange) ;
 SIGMA_orange<-rbind(SIGMA_orange,c(Sigma_orange))
 }
```

(b) Comparing the two groups, for the same mean body depth, the mean rear width of orange crabs is greater than the mean width of blue crabs



(c) $Pr(\rho_{\text{blue}} < \rho_{\text{orange}} | \mathbf{y}_{\text{blue}}, \mathbf{y}_{\text{orange}}) \approx 0.98$. The rear width and body depth of orange crabs are more highly correlated than for blue crabs. The posterior distribution of ρ suggests we are more certain about the value of ρ_{orange}, than about the value of ρ_{blue}

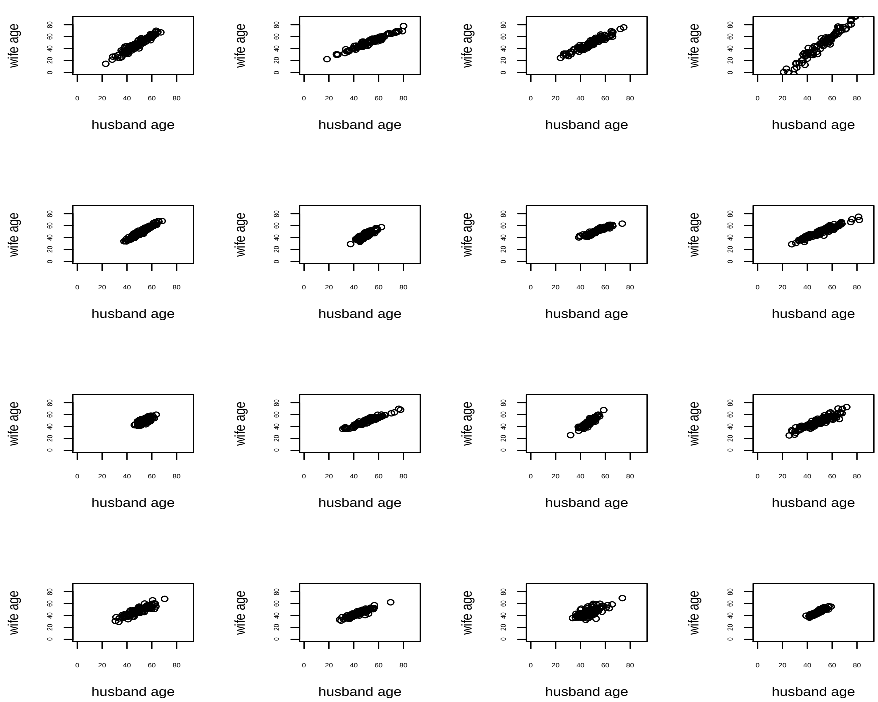
```
> mean(rho_blue<rho_orange)
[1] 0.9893
```



(2) (a)
```
mu_h<-50
mu_w<-48
mu0<-c(mu_h,mu_w)
nu0<-4
S0<-matrix(c(100,90,90,100),nrow=2,ncol=2)
L0<-matrix(c(9,8.1,8.1,9),nrow=2,ncol=2)
n<-100
```

A priori, assume the average age of married adult men is 50, and that on average, the husband is 2 years older than the wife. Expect high correlation between the age of married man and his wife, set prior correlation equal to 0.9. Assume a prior standard deviation for age of 10, which is set to cover 95% of ages of the population of married men and women. Assume a slightly lower prior standard deviation for the population mean.

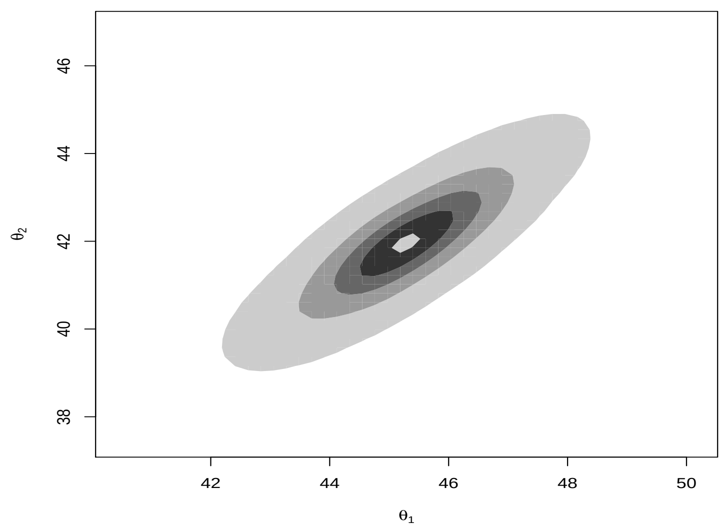
(b) The scatter plots of the prior predictive data sets show strong positive linear correlation and a range of 20 years - 80 years for both husband and wife ages. The prior predictive data sets roughly conform to prior beliefs.

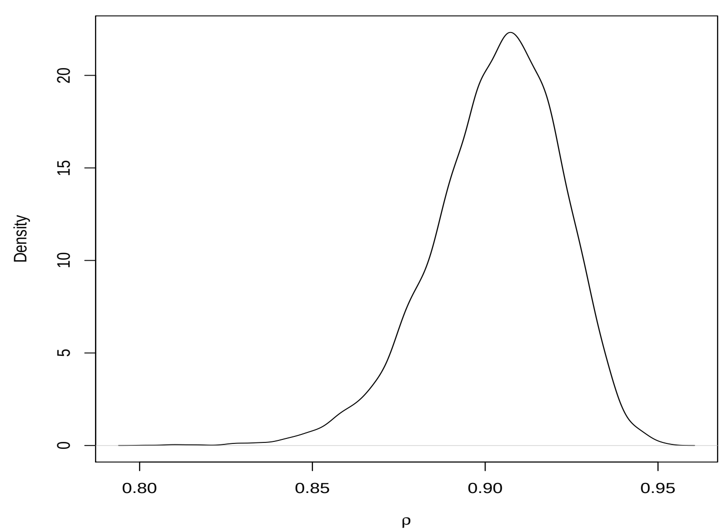


(c) Our prior estimates for θ_h and θ_w are not contained in the 95% posterior confidence intervals, specifically, our prior guess overestimated the average age of the husband and wife. Our prior guess on the correlation between husband and wife ages is contained in the 95% posterior confidence interval. That is, the data support the belief of a strong linear correlation between husband and wife ages as evidenced by the support on the interval (0.85,0.95) for the posterior density of ρ, and the strong linear correlation between posterior draws of θ_h and θ_w as shown in the scatter plot. Our prior guess also slightly underestimated the difference in average age of married man and the average age of a married woman.

	95% posterior confidence interval
θ_h	(42.9 , 47.8)
θ_w	(39.7 , 44.3)
$\theta_h - \theta_w$	(2.29 , 4.40)
ρ	(0.86 , 0.93)

```
> quantile(THETA[,1],c(0.025,0.975))
    2.5%    97.5%
42.86589 47.79614
> quantile(THETA[,2],c(0.025,0.975))
    2.5%    97.5%
39.65734 44.28761
> quantile(THETA[,1]-THETA[,2],c(0.025,0.975))
    2.5%    97.5%
2.292652 4.398711
> quantile(rho,c(0.025,0.975))
     2.5%     97.5%
0.8618795 0.9345871
```


46
44
42
40
38
42
44
46
48
50
θ1
θ2


20
15
10
5
0
0.80
0.85
0.90
0.95
ρ
Density