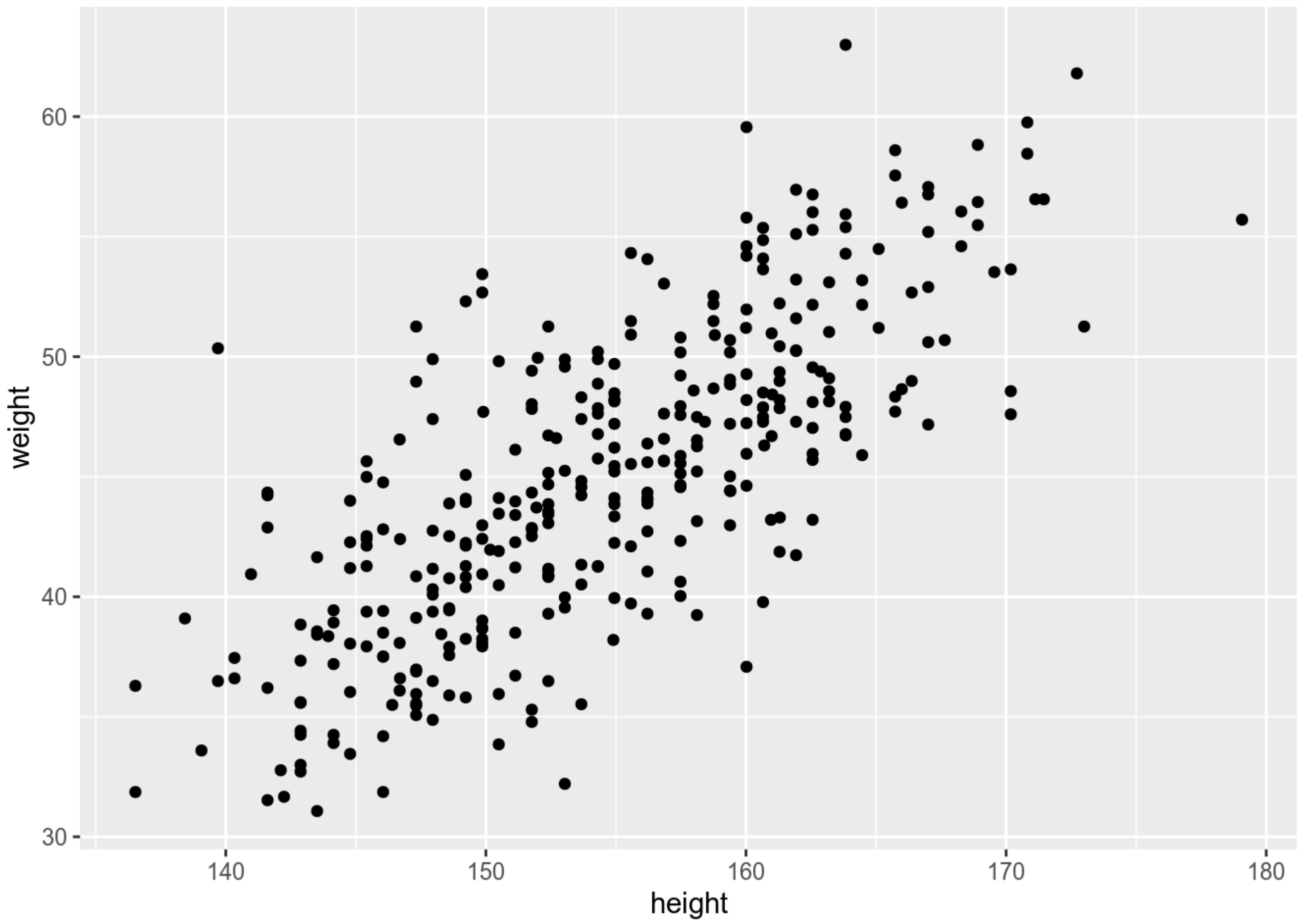
Howell example

Chang Tu

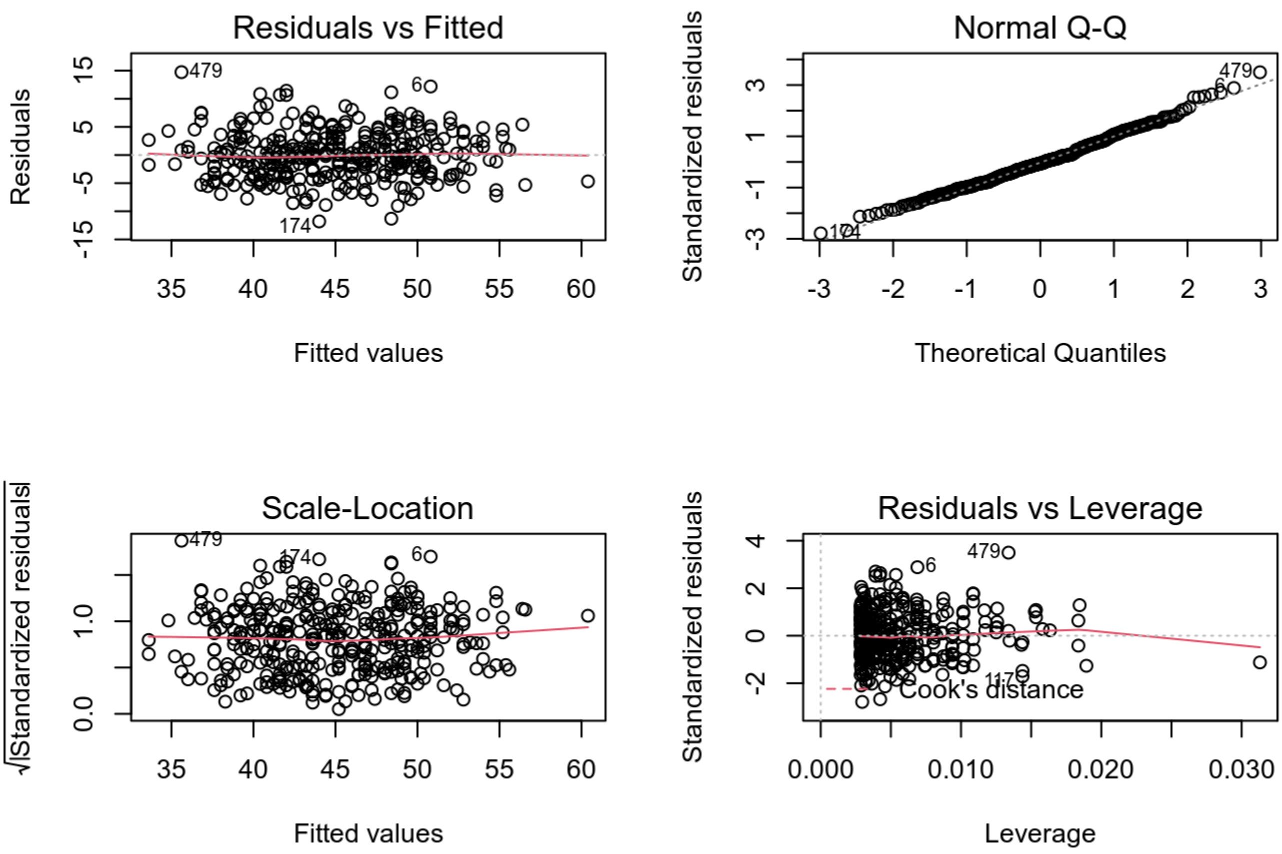

```r
rm(list=ls())
d <- read.csv("~/workspace/Baysian-inference/PART 1/Conjugate Normal Model and Linear
Regression/Howell Data from the Lecture.csv",sep=';')
d.adult <- d[d$age>=18,]
library(ggplot2)
ggplot(data=d.adult,aes(x=height,y=weight))+
  geom_point()
```




```r
m1 <- lm(weight ~ height, data=d.adult)
summary(m1)
```

```
##
## Call:
## lm(formula = weight ~ height, data = d.adult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.8022  -3.0183  -0.2293   2.8117  14.7348
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -52.31618    4.52650  -11.56   <2e-16 ***
## height        0.62942    0.02924   21.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.242 on 350 degrees of freedom
## Multiple R-squared:  0.5696, Adjusted R-squared:  0.5684
## F-statistic: 463.3 on 1 and 350 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2)); plot(m1)
```




```
# The Bayesian way
library(MCMCglmm)
```

```
## Loading required package: Matrix
```

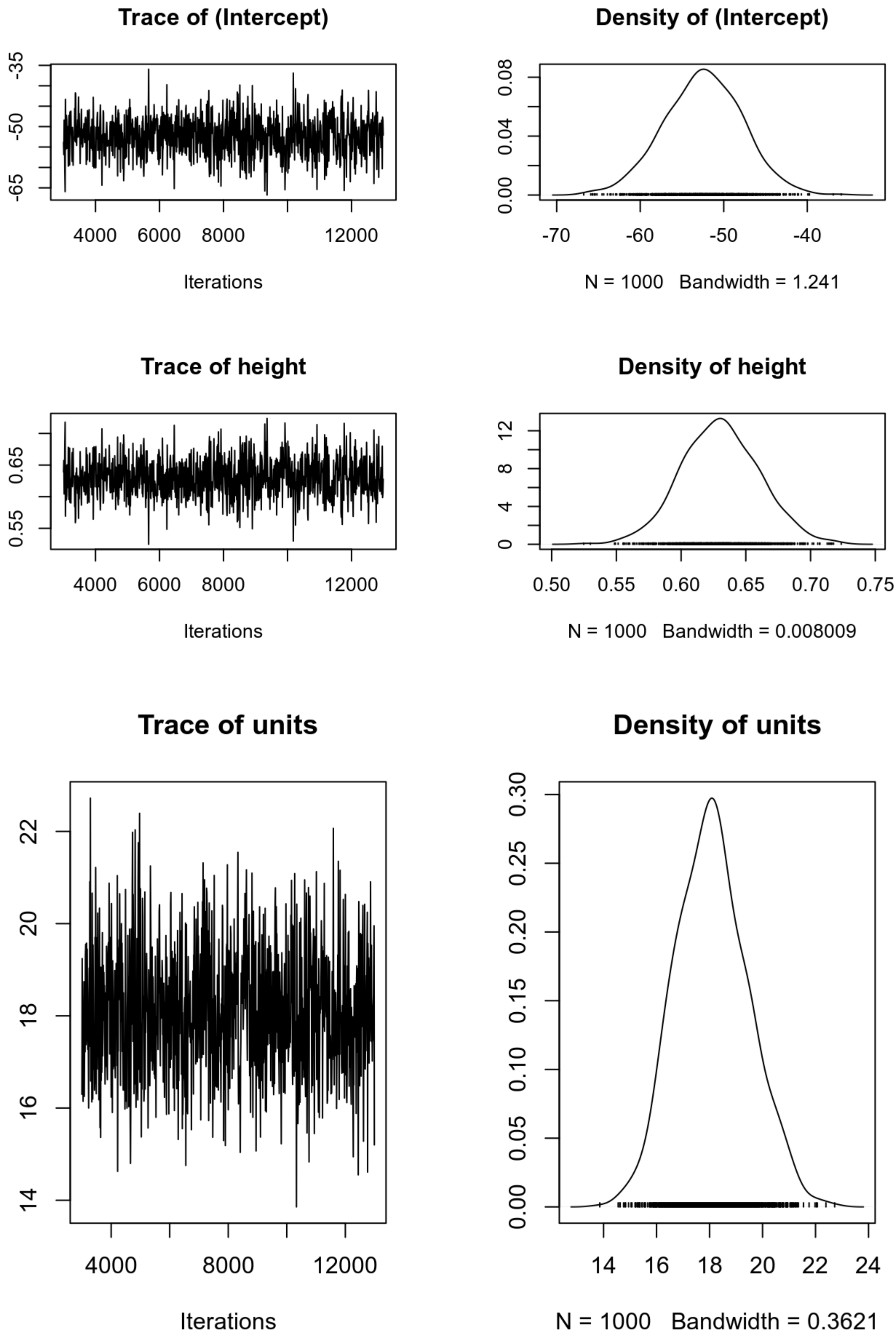
```
## Loading required package: coda
```

```
## Loading required package: ape
```

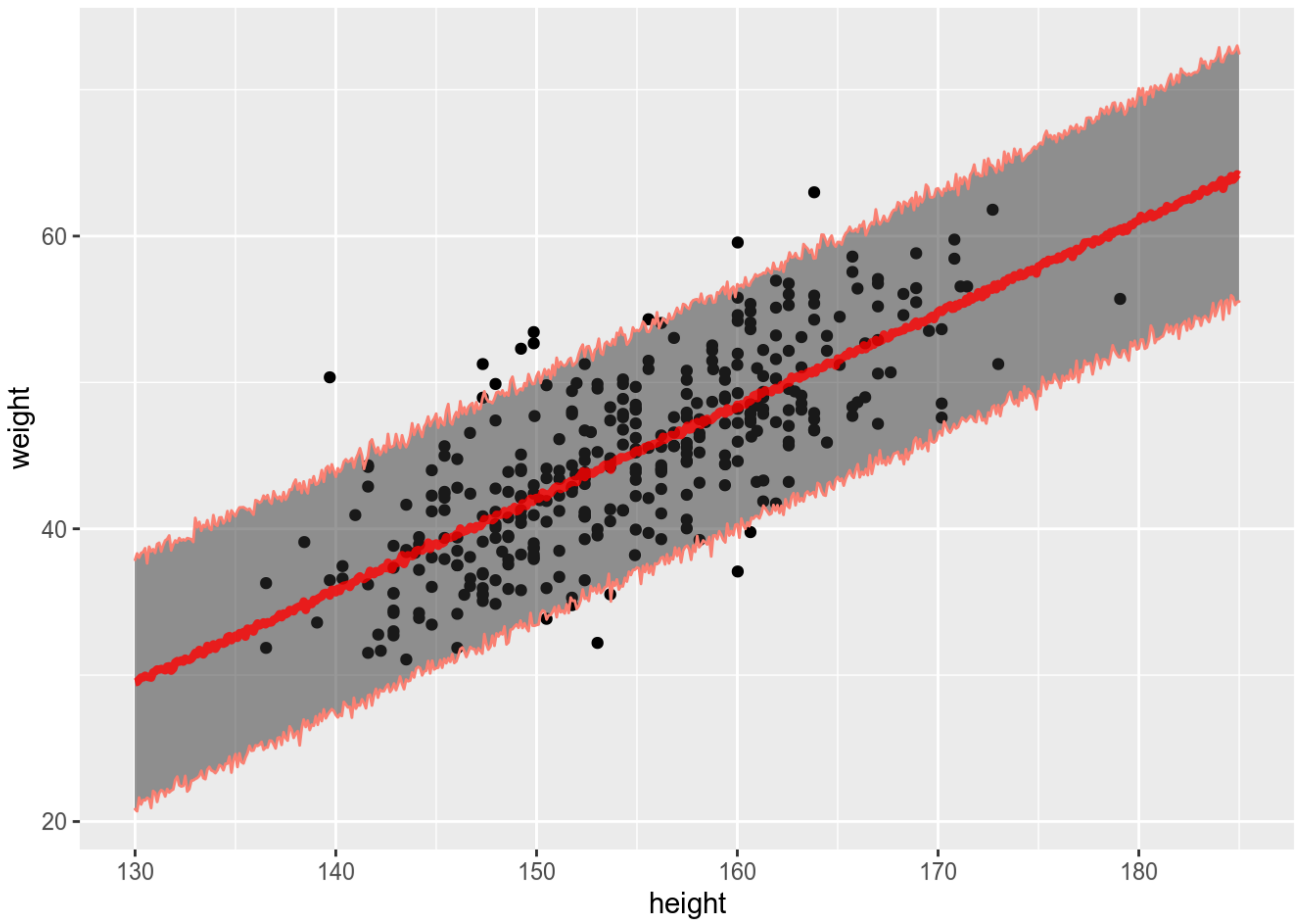
```
m1b <- MCMCglmm(weight ~ height, data=d.adult, verbose=F)
summary(m1b)
```

```
##
##  Iterations = 3001:12991
##  Thinning interval  = 10
##  Sample size  = 1000
##
##  DIC: 2020.148
##
##  R-structure:  ~units
##
##       post.mean l-95% CI u-95% CI eff.samp
## units      18.1    15.78    20.95     1000
##
##  Location effects: weight ~ height
##
##             post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)  -52.4603 -61.3505 -43.3182     1000 <0.001 ***
## height         0.6304   0.5702   0.6875     1000 <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

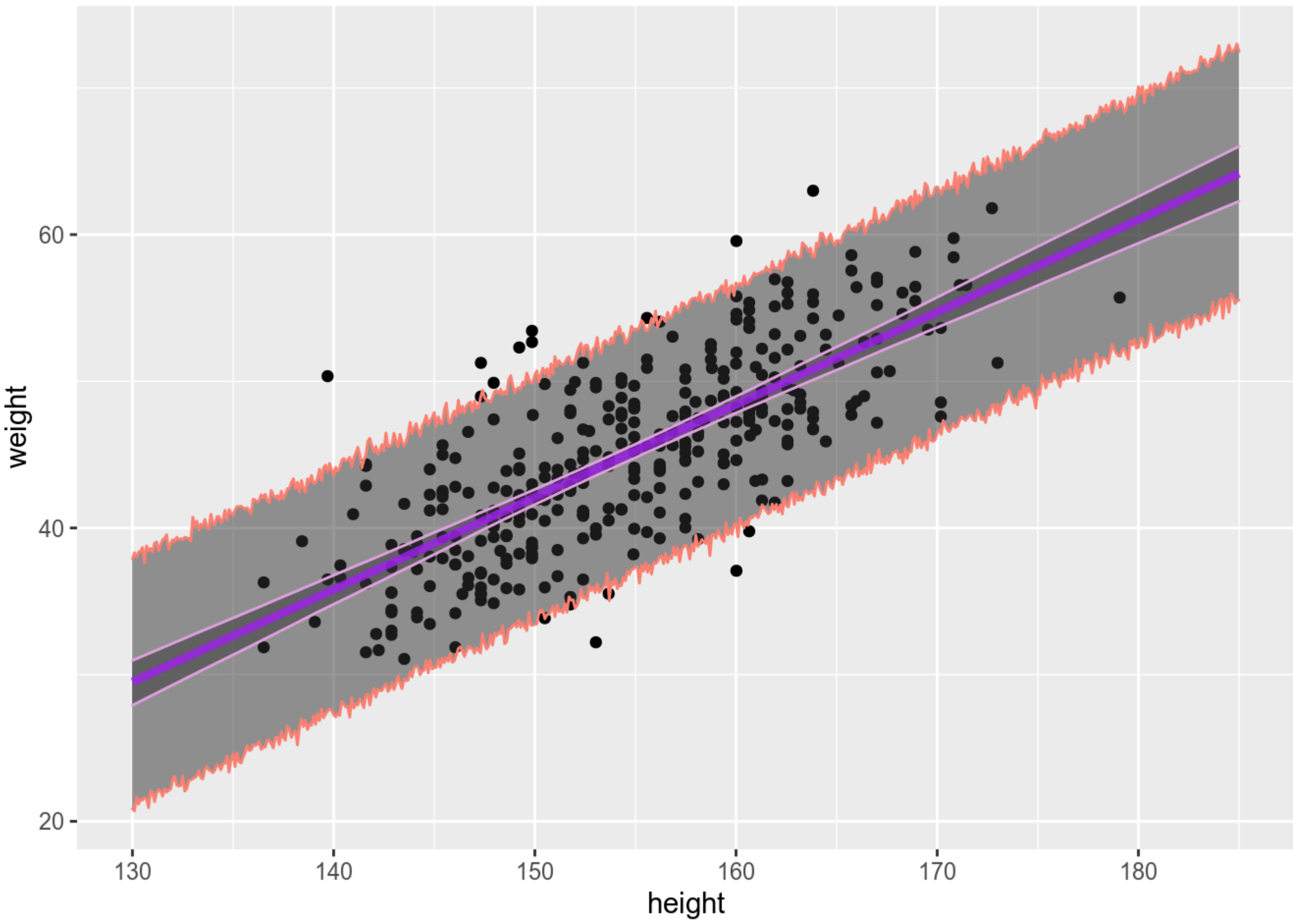
```
plot(m1b)
```


Trace of (Intercept)
Iterations
Density of (Intercept)
N = 1000 Bandwidth = 1.241
Trace of height
Iterations
Density of height
N = 1000 Bandwidth = 0.008009
Trace of units
Iterations
Density of units
N = 1000 Bandwidth = 0.3621

```
# Pr(beta[1]>0|data) > .999




### Model Fit: Using Posterior Predictive Distribution
# on a grid of reasonable values of height

dtf <- data.frame(height=seq(130,185,.1))

# a function to produce a posterior pred. estimate
pred.reg <- function(x,a,b,sigma){
  a+b*x+rnorm(length(a),0,sd=sigma)
}

m1b.pred <- sapply(dtf$height, pred.reg,
      a=m1b$Sol[,1],b=m1b$Sol[,2],sigma=sqrt(m1b$VCV))

dtf$pp.mean <- apply(m1b.pred,2,mean)
dtf$pp.q025 <- apply(m1b.pred,2,quantile,.025)
dtf$pp.q975 <- apply(m1b.pred,2,quantile,.975)

ggplot()+
  geom_point(data=d.adult,aes(x=height,y=weight))+
  geom_ribbon(data=dtf,aes(x=height,ymin=pp.q025,ymax=pp.q975),alpha=.5,col='salmon')
+
  geom_line(data=dtf,aes(x=height,y=pp.mean),size=1.5,alpha=.8,col='red')
```



```
# NB. for the posterior mean estimate (that's for mu, not for y):

pred.reg.mean <- function(x,a,b){
  a+b*x
}

m1b.post.mn <- sapply(dtf$height, pred.reg.mean,
                   a=m1b$Sol[,1],b=m1b$Sol[,2])

dtf$post.mean <- apply(m1b.post.mn,2,mean)
dtf$post.q025 <- apply(m1b.post.mn,2,quantile,.025)
dtf$post.q975 <- apply(m1b.post.mn,2,quantile,.975)

ggplot()+
  geom_point(data=d.adult,aes(x=height,y=weight))+
  geom_ribbon(data=dtf,aes(x=height,ymin=pp.q025,ymax=pp.q975),alpha=.5,col='salmon')
+
  #geom_line(data=dtf,aes(x=height,y=pp.mean),size=1.5,alpha=.8,col='red')+
  geom_ribbon(data=dtf,aes(x=height,ymin=post.q025,ymax=post.q975),alpha=.5,col='plu
m')+
  geom_line(data=dtf,aes(x=height,y=post.mean),size=1.5,alpha=.8,col='purple')
```




```
# note posterior mean and posterior predictive mean lines will be the same.
```