# STAT 638 HW 8

## Jack Cunningham

## 8.1)

a)

$Var(y_{i,j}|\mu, \tau^2)$ would be bigger than $Var(y_{i,j}|\theta_i, \sigma^2)$, this is because we expect samples from the same fixed group to be more alike than a sample from a more heterogeneous population. For example, we would expect students in the same school to be more alike than students at different schools.

b)

I would expect $Cov(y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2)$ to be zero since we know that $y_{1,j}, ..., y_{n_j,j}|\theta_j, \sigma^2$ are independent of each other. In other words if we know $\theta_j, \sigma$ we already have the information about the sampling distribution of subgroup $j$.

On the other hand $Cov(y_{i_1,j}, y_{i_2,j}|\mu, \tau)$ to be positive. This is because each observation gives us information about the subgroup $j$ . For example, if we see one student in a certain class scored higher than average on a national test we get some information about other students from the same class.

c)

I found it helpful to write out $Y_{ij} = \theta_j + e_{ij}$ where $e_{ij} \sim$ i.i.d. $N(0, \sigma^2)$

$$Var[y_{ij}|\theta_i, \sigma^2] = Var(\theta_i + e_{ij}|\theta_i) = \sigma^2$$

$$Var[y_{ij}|\mu, \tau^2] = Var(\theta_j + e_{ij}|\mu, \tau^2) = \tau^2 + \sigma^2$$

$$Var[\bar{y}_{.,j}|\theta_i, \sigma^2] = Var[\theta_j + \sum_{i=1}^{n} \frac{e_{ij}}{n}|\theta_i, \sigma^2] = \sigma^2/n$$

$$Var[\bar{y}_{.,j}|\mu, \tau^2] = Var[\theta_j + \sum_{i=1}^{n} \frac{e_{ij}}{n}|\mu, \tau^2] = \tau^2 + \sigma^2/n$$

$$Cov(y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2) = Cov(\theta_j + e_{i_1,j}, \theta_j + e_{i_2,j})|\theta_j, \sigma^2 = 0$$

$$Cov(y_{i_1,j}, y_{i_2,j}|\mu, \tau^2) = Cov(\theta_j + e_{i_1,j}, \theta_j + e_{i_2,j}) = Cov(\theta_j, \theta_j) = Var(\theta_j) = \tau^2$$

d)

We start with the joint distribution:

$$p(\mu, \theta_1, ..., \theta_m, \sigma^2, \tau^2, y_1, ..., y_m) = p(\mu)p(\theta_1, ..., \theta_m|\mu, \tau^2)p(y_1, ..., y_m|\theta_1, ...\theta_m, \sigma^2)$$

Then we condition on everything but $\mu$:

$$p(\mu|\theta_1, ..., \theta_m, \sigma^2, \tau^2, y_1, ..., y_m) = p(\mu)p(\theta_1, ..., \theta_m|\mu, \tau^2) = p(\mu|\theta_1, ..., \theta_m, \tau^2)$$

This tells us only information from $\theta_1, ..., \theta_j$ is necessary to know the posterior distribution of $\mu$. This makes sense based on the flow of information in the model. The data $y_1, ..., y_m$ flows to the subgroup average $\theta_1, ...\theta_j$, then those subgroup averages flow up to $\mu$. This means that as long as we have the subgroup averages and the across group sampling variance we have all the information we need to make posterior inference about the overall mean $\mu$.

## 8.3)

a)

```
school_1 <- read.table("school1.dat")
school_2 <- read.table("school2.dat")
school_3 <- read.table("school3.dat")
school_4 <- read.table("school4.dat")
school_5 <- read.table("school5.dat")
school_6 <- read.table("school6.dat")
school_7 <- read.table("school7.dat")
school_8 <- read.table("school8.dat")
schools <- list()
schools[[1]] <- school_1$V1
schools[[2]] <- school_2$V1
schools[[3]] <- school_3$V1
```

2

```r
schools[[4]] <- school_4$V1
schools[[5]] <- school_5$V1
schools[[6]] <- school_6$V1
schools[[7]] <- school_7$V1
schools[[8]] <- school_8$V1

#Priors
mu_0 <- 7
gamma_2_0 <- 5
tau_2_0 <- 10
eta_0 <- 2
sigma_2_0 <- 15
nu_0 <- 2

#school details
k <- length(schools)
n_i <- sapply(schools, length)
n_total <- sum(n_i)
y_bar <- sapply(schools,mean)

set.seed(10)
niter <- 50000

#gibbs sampling
theta <- y_bar
mu <- mean(theta)
tau_2 <- var(theta)
sigma_2 <- mean(sapply(schools, var))

trace_theta <- matrix(NA, nrow = niter, ncol = k)
trace_mu <- numeric(niter)
trace_tau2 <- numeric(niter)
trace_sigma2 <- numeric(niter)

id <- 0
for(j in 1:niter){
  for(i in 1:k){

    #Theta Sample
    y_i <- schools[[i]]
```

```
    n_i <- length(y_i)
    ybar_i <- mean(y_i)
    var_theta_i <- 1/(n_i/sigma_2 + 1/tau_2)
    mean_theta_i <- var_theta_i * (n_i * ybar_i / sigma_2 + mu / tau_2)
    theta[i] <- rnorm(1, mean_theta_i, sqrt(var_theta_i))
  }

  #mu sample
  var_mu <- 1 / (k / tau_2 + 1 / gamma_2_0)
  mean_mu <- var_mu * (sum(theta) / tau_2 + mu_0 / gamma_2_0)
  mu <- rnorm(1, mean_mu, sqrt(var_mu))

  #sigma^2 sample
  ssq <- 0
  for (i in 1:k) ssq <- ssq + sum((schools[[i]] - theta[i])^2)
  post_nu <- nu_0 + n_total
  post_scale_sigma <- (nu_0*sigma_2_0 + ssq) / 2
  sigma_2 <- 1/rgamma(1, post_nu/2, post_scale_sigma)

  #tau^2 sample
  ssq_theta <- sum((theta - mu)^2)
  post_eta <- eta_0 + k
  post_scale_tau <- (eta_0 * tau_2_0 + ssq_theta) / 2
  tau_2 <- 1/rgamma(1, post_eta/2, post_scale_tau)

  #Saving values
  id <- id + 1
  trace_theta[id,] <- theta
  trace_mu[id]     <- mu
  trace_tau2[id]   <- tau_2
  trace_sigma2[id] <- sigma_2
}
```
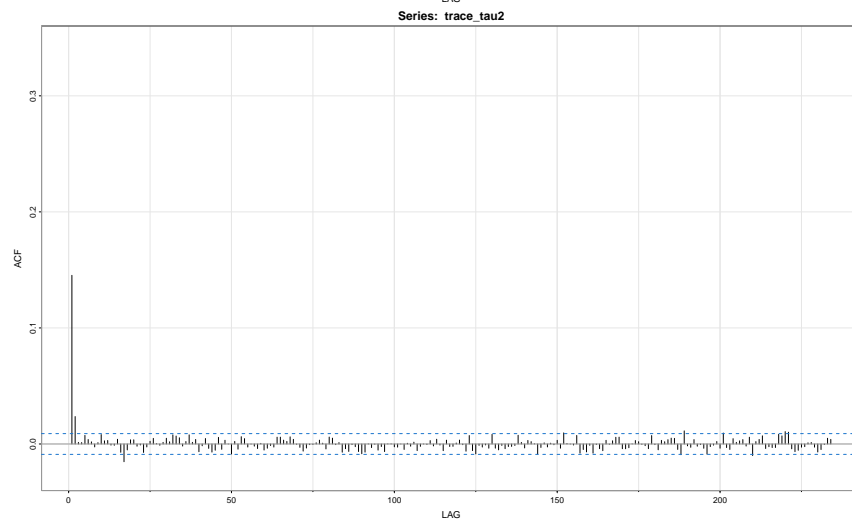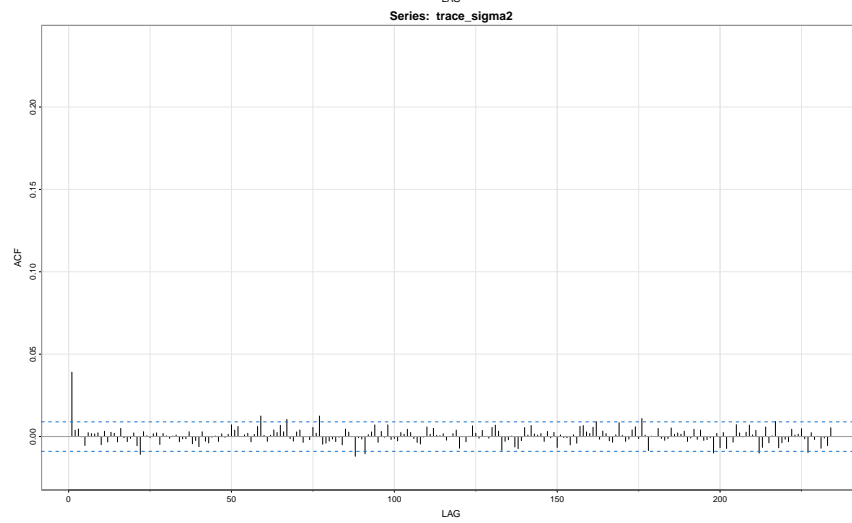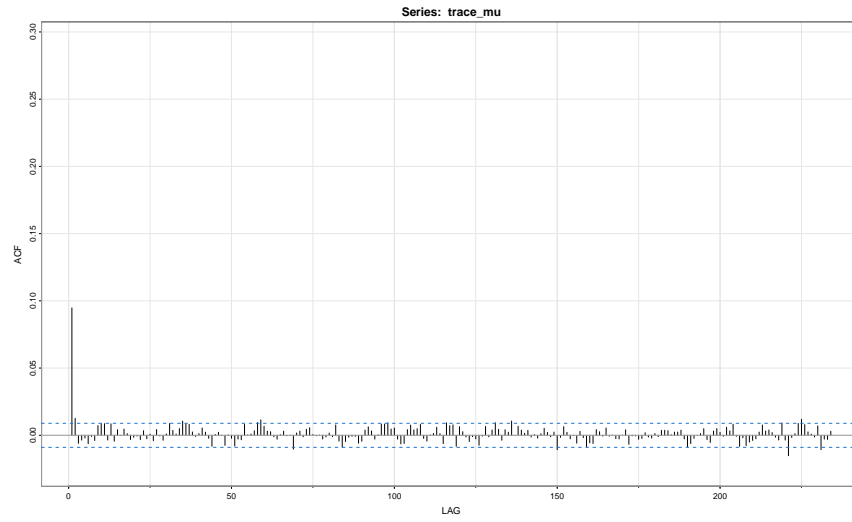
Assessing stationarity:

```
library(astsa)
par(mfrow = c(3,1))
acf1(trace_mu, plot = TRUE)
acf1(trace_sigma2, plot = TRUE)
acf1(trace_tau2, plot = TRUE)
```

**Series: trace_mu**



**Series: trace_sigma2**



**Series: trace_tau2**

The ACF quickly tends to zero for all three indicating that we have good convergence.

```
library(coda)
paste("mu effective size:", round(effectiveSize(trace_mu)))
```

```
[1] "mu effective size: 41346"
```

```
paste("sigma^2 effective size:", round(effectiveSize(trace_sigma2)))
```

```
[1] "sigma^2 effective size: 46239"
```

```
paste("tau^2 effective size:", round(effectiveSize(trace_tau2)))
```

```
[1] "tau^2 effective size: 37310"
```

The effective sample sizes also look good.

b)

```
paste("Posterior Mu:", round(mean(trace_mu),3), "95% CR:",round(quantile(trace_mu,c(.025))
```

```
[1] "Posterior Mu: 7.563 95% CR: 5.97 9.145"
```

```
paste("Posterior Sigma^2:", round(mean(trace_sigma2),3),
      "95% CR:", round(quantile(trace_sigma2,c(.025)),3),
      round(quantile(trace_sigma2,c(.975)),3))
```

```
[1] "Posterior Sigma^2: 14.489 95% CR: 11.732 17.862"
```

```
paste("Posterior Tau^2:", round(mean(trace_tau2),3), "95% CR:",
      round(quantile(trace_tau2,c(.025)),3), round(quantile(trace_tau2,c(.975)),3))
```

```
[1] "Posterior Tau^2: 5.516 95% CR: 1.91 14.471"
```

```r
mu_prior <- rnorm(1000000, mu_0, sd = sqrt(gamma_2_0))
dinvgamma <- function(x, shape, scale){
  dens <- ifelse(x > 0,
                 (scale^shape/gamma(shape))*x^(-shape - 1)*exp(-scale /x),
                 0
  )
}
x_prior <- seq(0.0001, 120, .01)
sigma_2_prior <- dinvgamma(x_prior, nu_0/2, ((nu_0 * sigma_2_0))/2)
tau_2_prior <- dinvgamma(x_prior, eta_0/2, (eta_0 * tau_2_0) / 2)

par(mfrow = c(3,2))
plot(density(mu_prior), main = "Prior Mu", xlim = c(0,15))
plot(density(trace_mu), main = "Posterior Mu" , xlim = c(0,15))

plot(x_prior, sigma_2_prior, type = "l", main = "Prior Sigma^2")
plot(density(trace_sigma2), main = "Posterior Sigma^2", xlim = c(5, 25))

plot(x_prior, tau_2_prior, main = "Prior Tau^2", type = "l")
plot(density(trace_tau2),  main = "Posterior Tau^2",xlim = c(0,40))
```
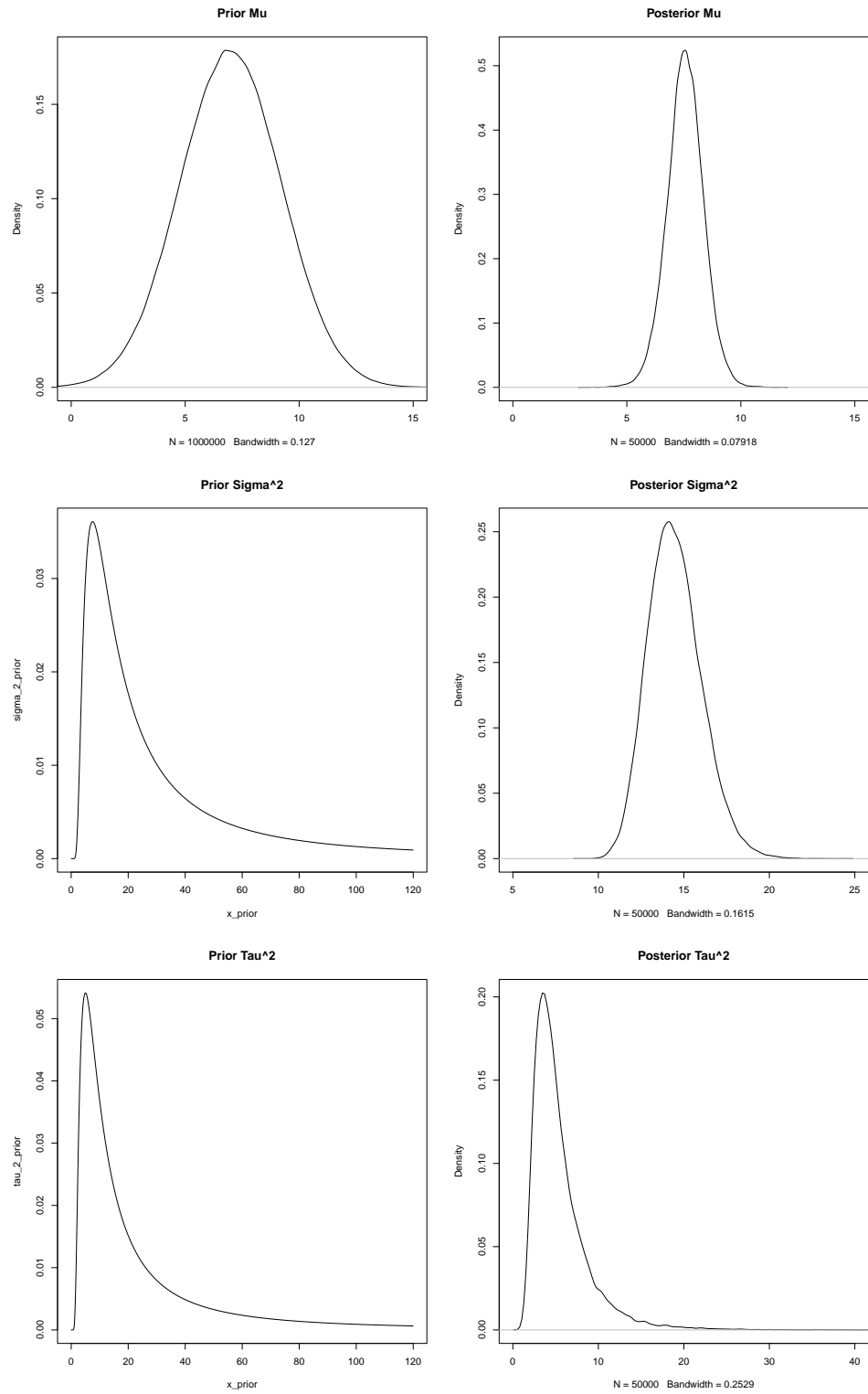
## Prior Mu

Density

N = 1000000   Bandwidth = 0.127

## Posterior Mu

Density

N = 50000   Bandwidth = 0.07918

## Prior Sigma^2

sigma_2_prior

x_prior

## Posterior Sigma^2

Density

N = 50000   Bandwidth = 0.1615

## Prior Tau^2

tau_2_prior

x_prior

## Posterior Tau^2

Density

N = 50000   Bandwidth = 0.2529

The prior for $\mu$ is a normal distribution centered at 7 with $\gamma_0^2$ chosen to keep almost all its density greater than 0. After receiving the data we see the center shift right slightly, at around 7.5. We see the distribution tighten around that value, with almost all density between 5 and 10.

The prior for $\sigma^2$ peaks near 15 and has a very long right tail. After receiving data we see the distribution come close to symmetry, centered still around 15 but with almost all mass between 10 and 20.

The prior for $\tau^2$ peaks near 10 and also has a very long right tail. Unlike the posterior for $\sigma^2$ this distribution keeps the right tail with visible density extending out to 30. The location of the peak is similar to the prior.

The posterior shows that we have learned a great deal about the feasible range for each statistic from the data. This helps us understand our uncertainty more accurately.

c)

```
par(mfrow = c(2,1))
set.seed(20)
R_posterior = trace_tau2/(trace_sigma2 + trace_tau2)
sigma_2_prior_sim = 1/rgamma(100000,nu_0/2, ((nu_0 * sigma_2_0))/2)
tau_2_prior_sim   = 1/rgamma(100000,eta_0/2, (eta_0 * tau_2_0) / 2 )
R_prior = tau_2_prior_sim/(sigma_2_prior_sim + tau_2_prior_sim)
plot(density(R_prior), main = "R prior", xlim = c(0,1))
plot(density(R_posterior), main = "R posterior", xlim = c(0,1))
```

## R prior



N = 100000   Bandwidth = 0.0257

## R posterior



N = 50000   Bandwidth = 0.00999
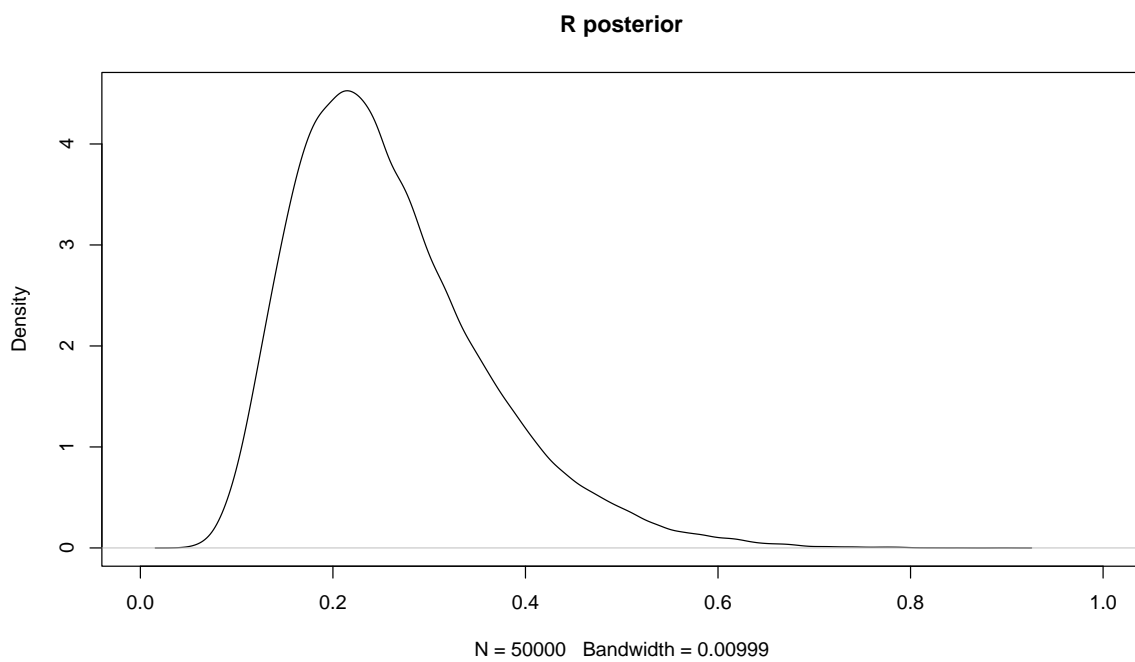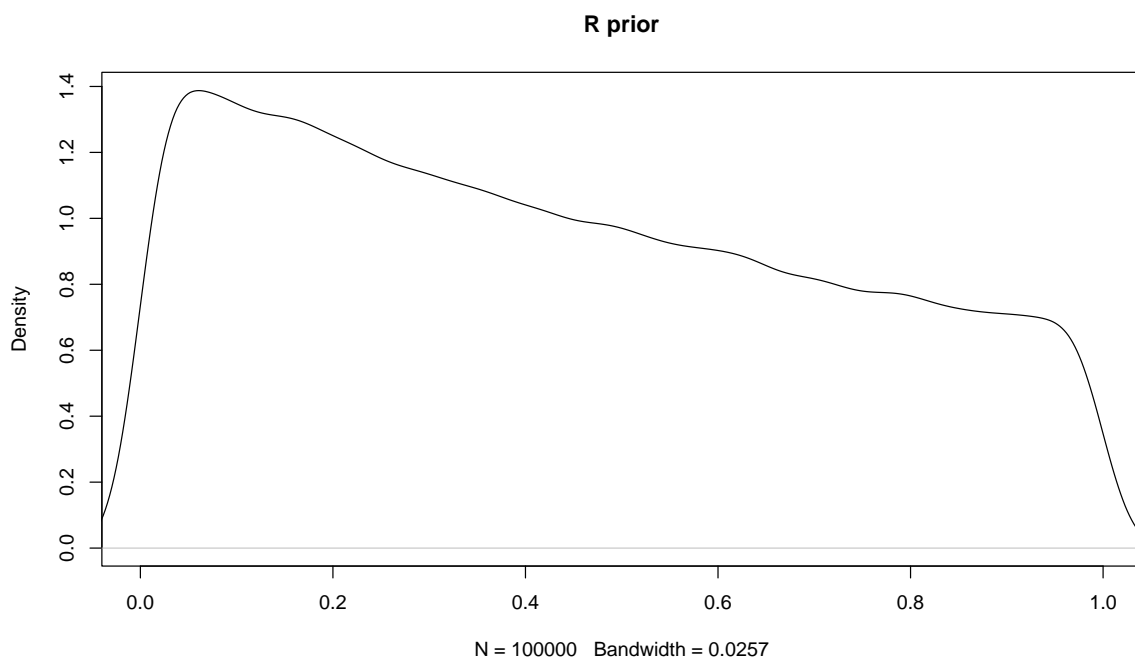
The R statistic represents the portion of variance coming from across school variation from all
the variation in the data. The prior represents a slightly informed view that the across school

variation makes up less of the total variation than the between school variation. This makes sense, considering we selected $\sigma_0^2 = 15$ and $\tau_0^2 = 10$.

After the data is observed we see that the posterior is centered around .2, saying that 20% of total variance is coming from across school variation and 80% is coming from between school variation. The distribution ranges from 0 to about .8 with a long right tail.

If we were wondering if across school variation makes up a smaller portion than between school variation, we can test this by seeing how many times $R_{post}$ is less than .5. This probability is 0.02732. Evidence suggests that most variance in the data is coming from between school variation.

d)

To obtain the posterior probability that $\theta_7 < \theta_6$ we can simply check the portion of times $\theta_{7,post}$ is less than $\theta_{6,post}$.

```
mean(trace_theta[,7] < trace_theta[,6])
```

```
[1] 0.5254
```

Similarly we can find the posterior probability that $\theta_7$ is the smallest of all $\theta$'s by checking the portion of times $\theta_7$ is the minimum of all $\theta$ in each posterior sample.

```
theta_mins = apply(trace_theta, 1, min)
mean(trace_theta[,7] == theta_mins)
```
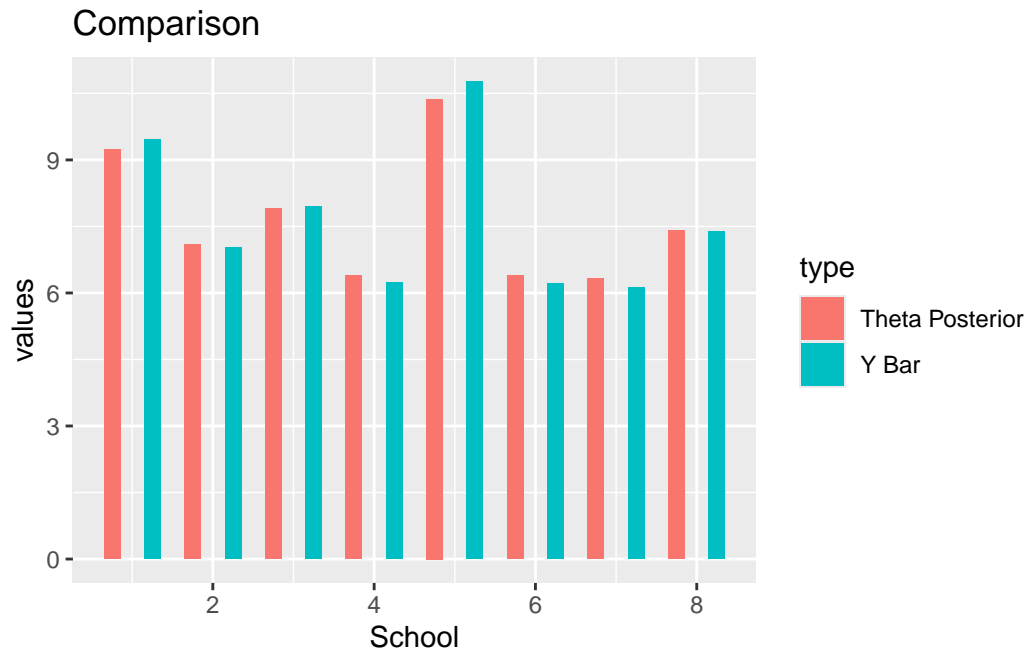
```
[1] 0.3237
```

e)

```
library(tidyverse)
theta_post_exp <- apply(trace_theta, 2, mean)
school_labels = rep(1:8,times = 2)
type = rep(c("Y Bar", "Theta Posterior"), each = 8)
values = c(y_bar, theta_post_exp)

comparison_df <- tibble(school_labels, values, type)

comparison_df |> ggplot(aes(x = school_labels, y = values, fill = type)) +
  geom_col(position = position_dodge(width = 1), width = .4) +
  labs(title = "Comparison", x = "School")
```

11

## Comparison



The comparison between the sample mean and posterior mean of $\mu$ is below:

```
paste("Sample Mean: ", round(mean(unlist(schools)),3),". Posterior mean of mu: ", round(me
```

```
[1] "Sample Mean:  7.691 . Posterior mean of mu:  7.563"
```