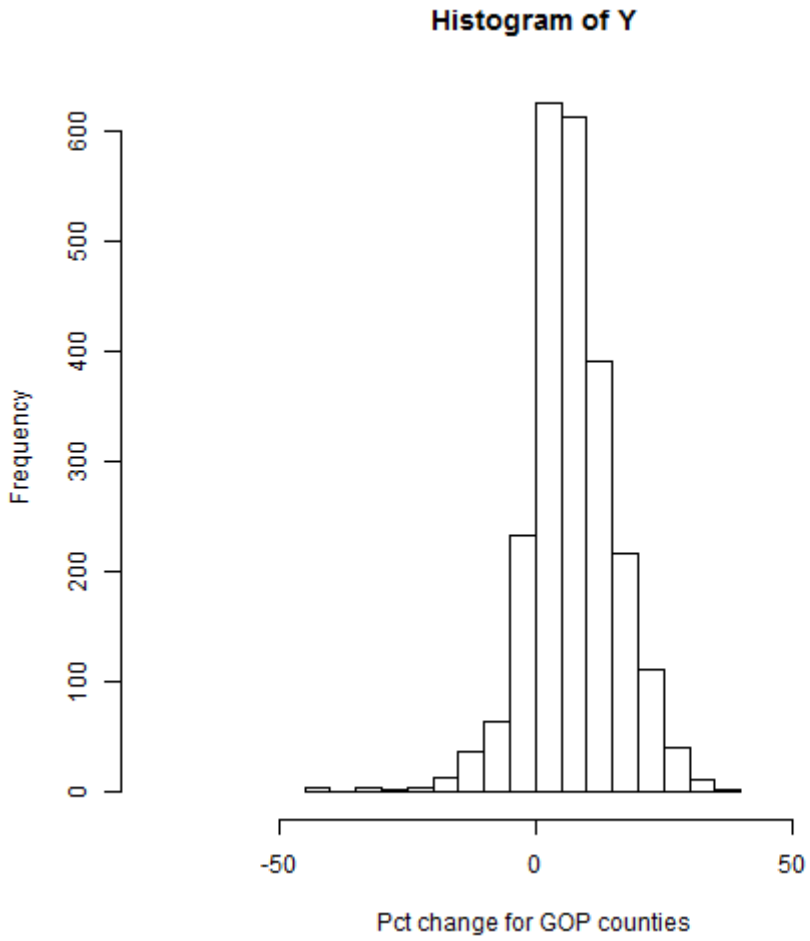
Bayesian two-sample t-test

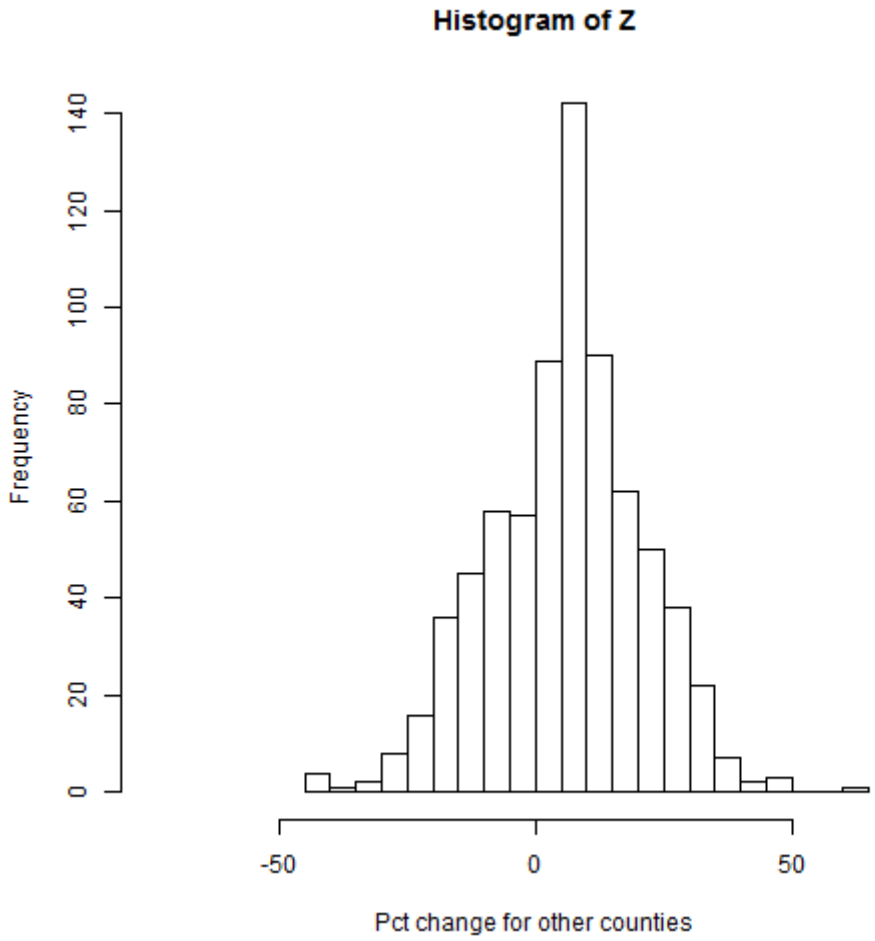
In this exercise we will test whether counties that favored GOP in 2012 have higher percent increase in support for GOP in 2016. Let $Y_1, \ldots, Y_{n_1} \sim \text{Normal}(\mu_Y, \sigma^2)$ be the percent change in support for the n_1 counties that favored GOP in 2012 and $Z_1, \ldots, Z_{n_2} \sim \text{Normal}(\mu_Z, \sigma^2)$ be the change in support for the remaining n_2 counties that voted Democrat in 2012. Our goal is to determine if the difference in means $\Delta = \mu_Y - \mu_Z$ is zero. We select priors

$$\mu_Y, \mu_Z \sim \text{Normal}(\mu_0, \sigma_0^2) \quad \sigma^2 \sim \text{InvGamma}(a, b).$$

```
load("S:\\Documents\\www\\ABA\\code\\election_2008_2016.RData")

change <- 100*(pct_gop_16/pct_gop_12-1)

Y  <- change[pct_gop_12>0.5]
Y  <- Y[!is.na(Y)]
n1 <- length(Y)

Z  <- change[pct_gop_12<0.5]
Z  <- Z[!is.na(Z)]
n2 <- length(Z)

hist(Y,breaks=20,xlim=c(-75,75),xlab="Pct change for GOP counties")
```




```
hist(Z,breaks=20,xlim=c(-75,75),xlab="Pct change for other counties")
```




```
summary(Y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -44.380   1.999   6.452   6.933  11.810  38.180
```

```
summary(Z)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -42.480  -3.822   6.835   5.900  15.210  61.680
```

```
# OK, maybe the variances aren't the same, but we persevere!
```

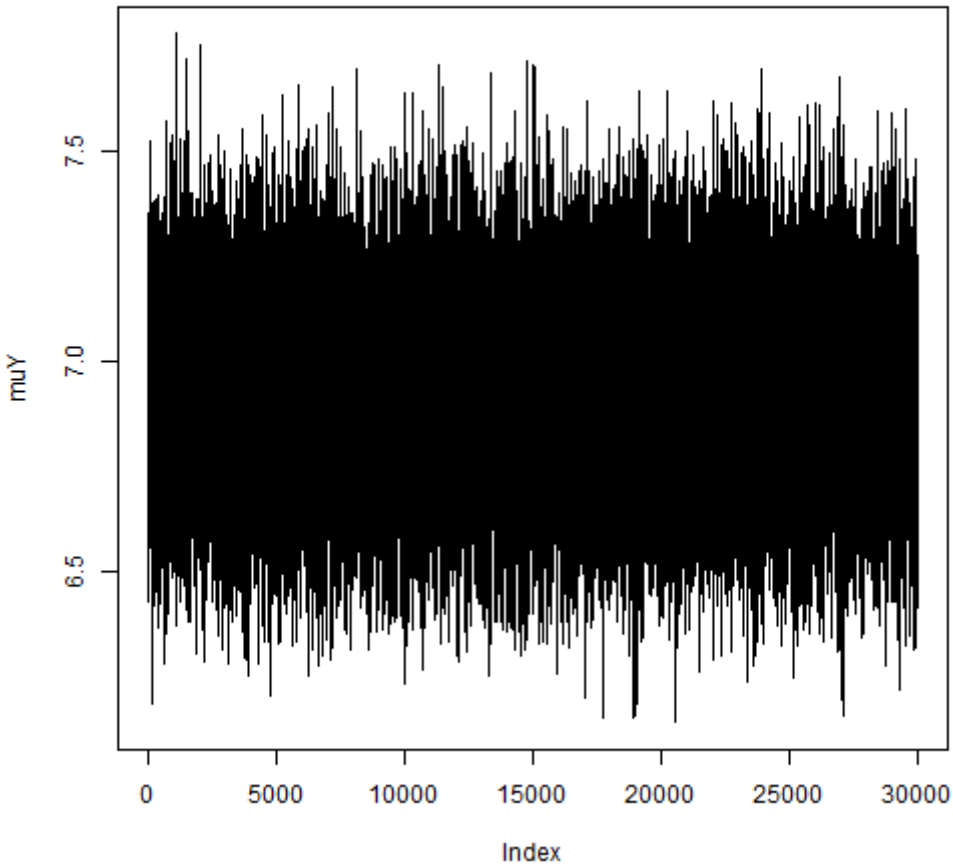
```
### Priors

  mu0 <- 0
  s20 <- 1000
  a   <- 0.01
  b   <- 0.01
```

MCMC!

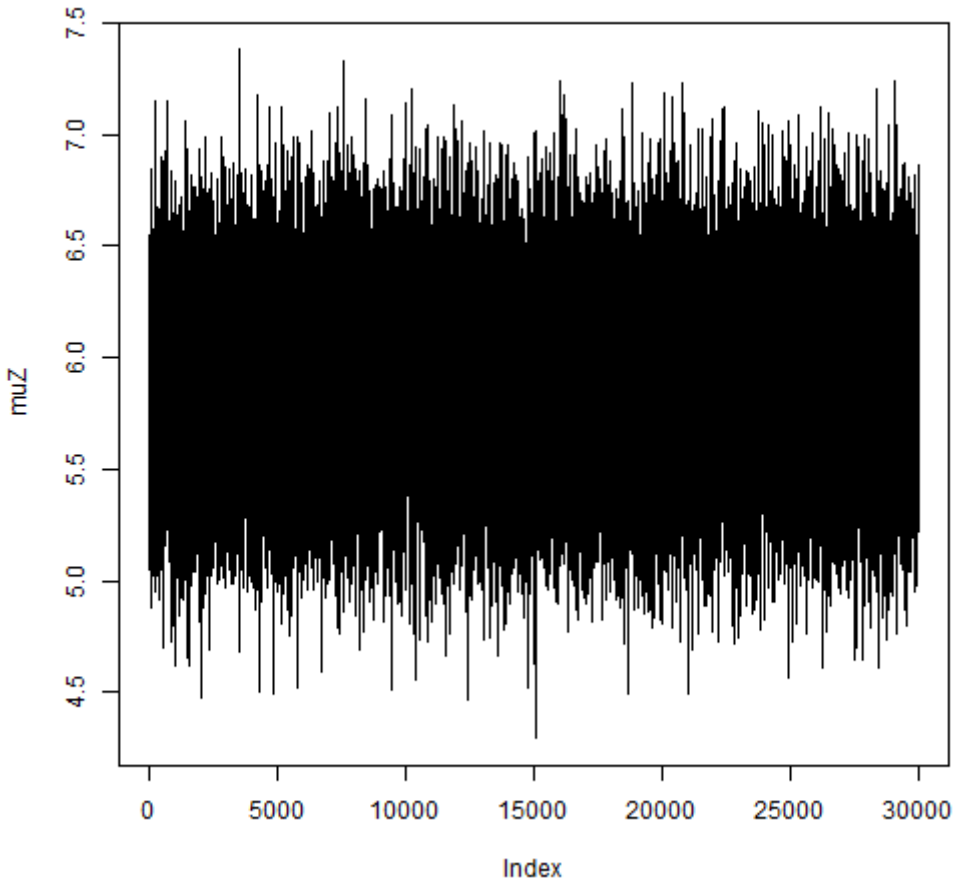
```
n.iters <- 30000
keepers <- matrix(0,n.iters,4)
colnames(keepers)<-c("muY","muZ","sigma2","Delta")

# Initial values
muY         <- mean(Y)
muZ         <- mean(Z)
s2          <- (var(Y)+var(Z))/2
keepers[1,] <- c(muY,muZ,s2,muY-muZ)

for(iter in 2:n.iters){

  # sample muY|s2,Y

   A   <- sum(Y)/s2+mu0/s20
   B   <- n1/s2+1/s20
   muY <- rnorm(1,A/B,1/sqrt(B))

  # sample muZ|s2,Z

   A   <- sum(Z)/s2+mu0/s20
   B   <- n2/s2+1/s20
   muZ <- rnorm(1,A/B,1/sqrt(B))

  # sample s2|mu,Y,Z

   A  <- n1/2+n2/2+a
   B  <- sum((Y-muY)^2)/2 + sum((Z-muZ)^2)/2+b
   s2 <- 1/rgamma(1,A,B)

  # keep track of the results
   keepers[iter,] <- c(muY,muZ,s2,muY-muZ)

}
```

Plots convergence

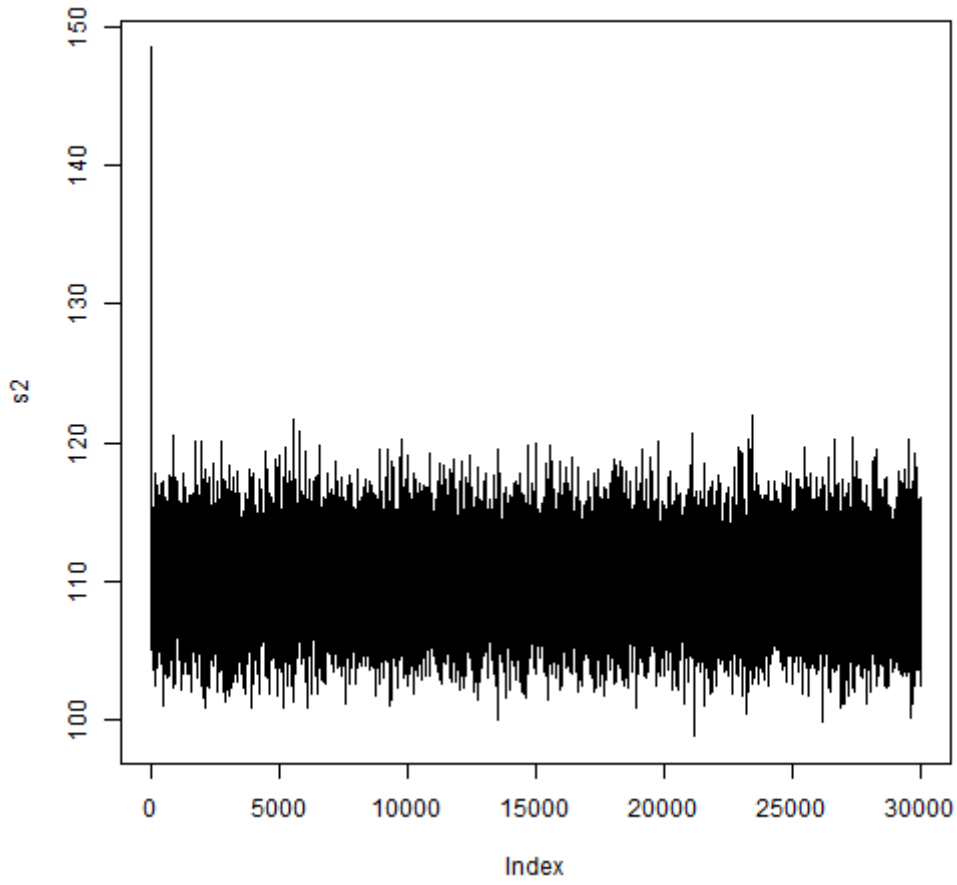
```
plot(keepers[,1],type="l",ylab="muY")
```



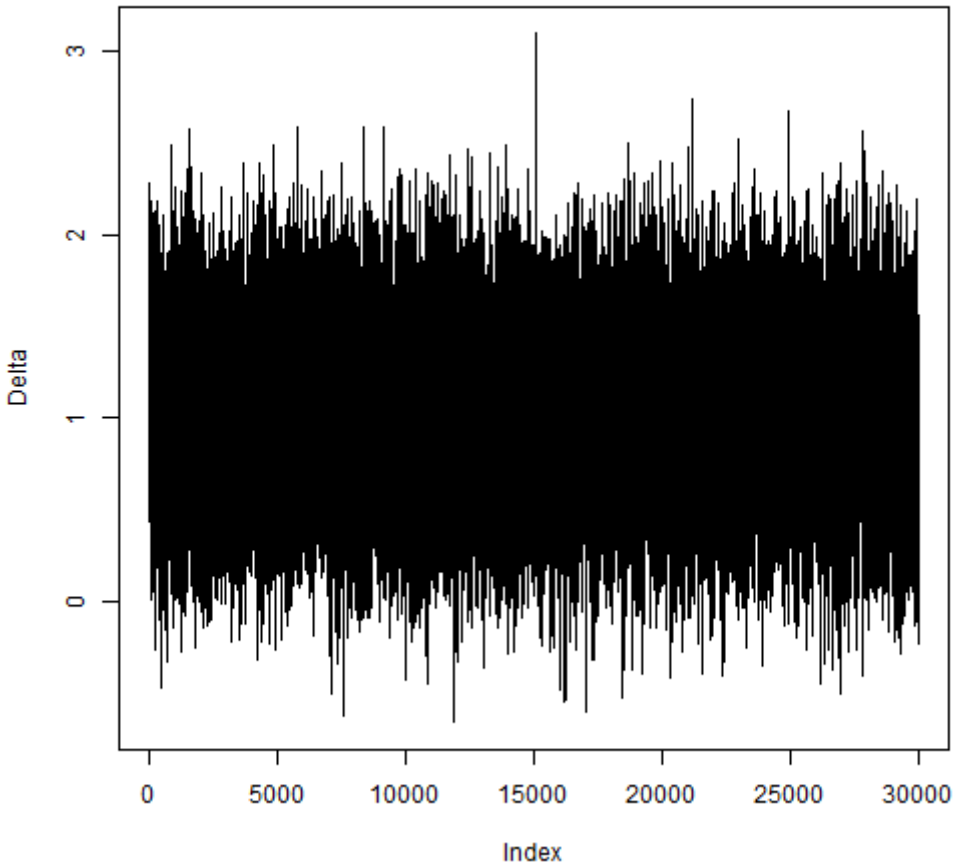

```
plot(keepers[,2],type="l",ylab="muZ")
```




```
plot(keepers[,3],type="l",ylab="s2")
```

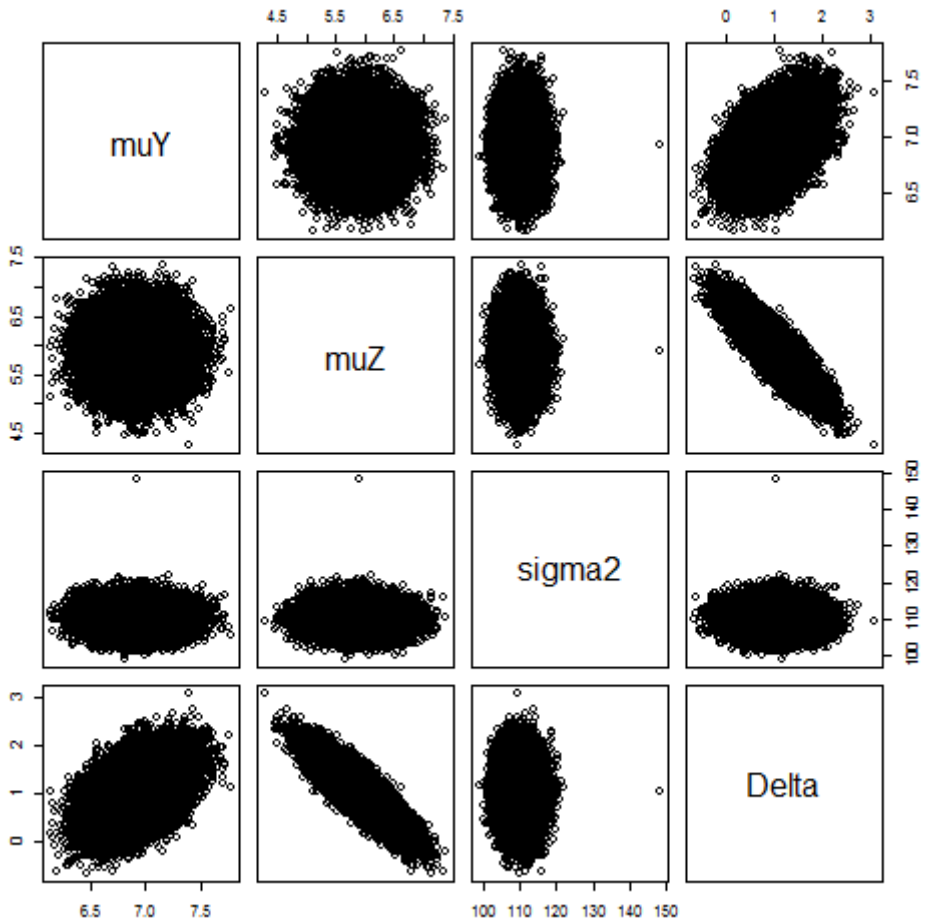



```
plot(keepers[,4],type="l",ylab="Delta")
```



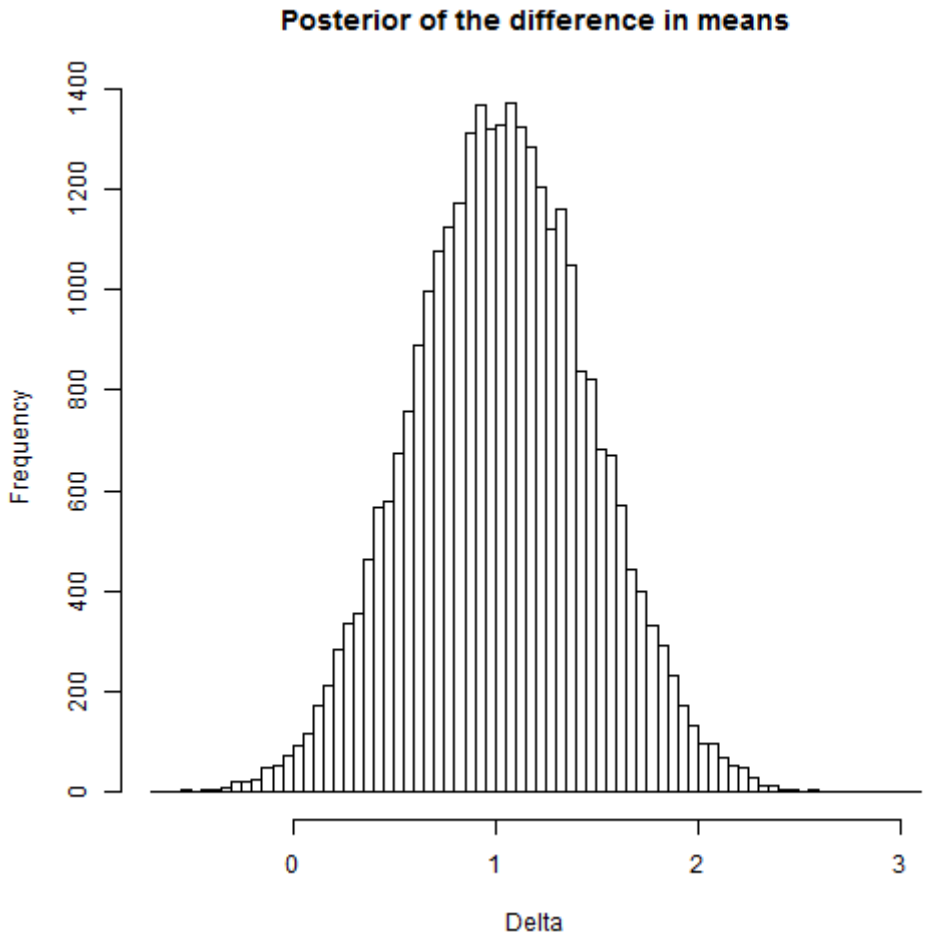
Plots of the joint posterior distribution.

```
pairs(keepers)
```



Plots of the marginal distribution of $\Delta = \mu_Y - \mu_Z$.

```
Delta <- keepers[,4]
hist(Delta,main="Posterior of the difference in means",breaks=100)
```



Compute the posterior probability that $\Delta > 0$ to test $H_0 : \Delta \leq 0$ versus $H_1 : \Delta > 0$

```
  mean(Delta>0)
```

```
## [1] 0.9909667
```

These data favor the alternative hypothesis that the GOP counties moved more towards GOP than other counties.