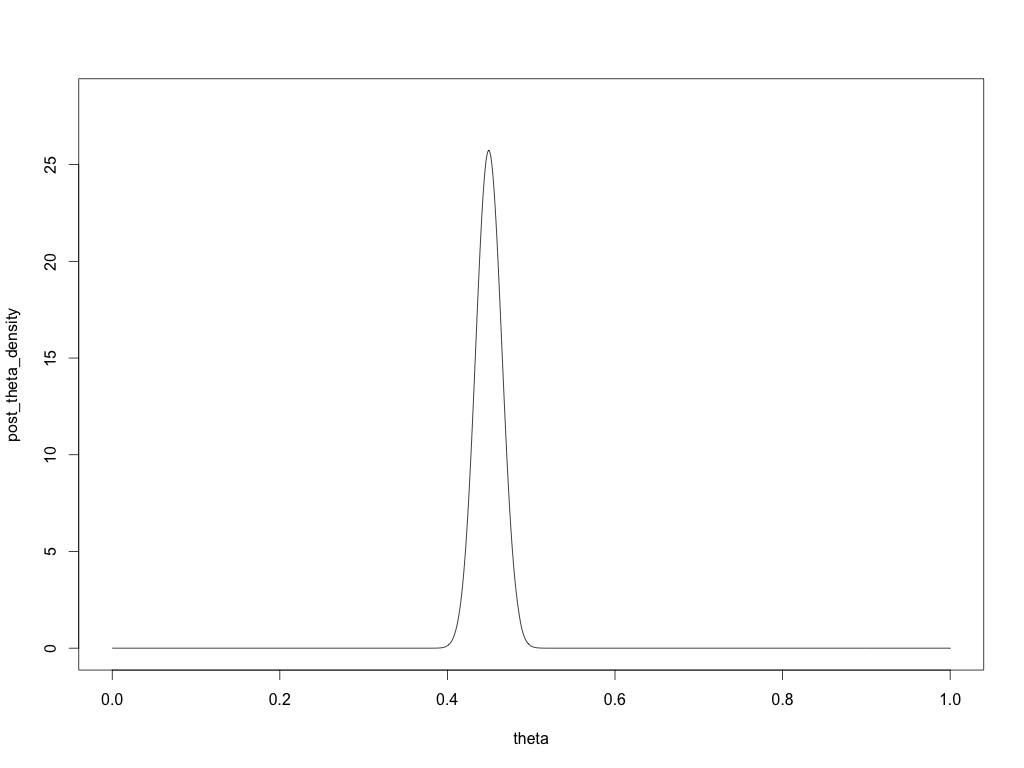
3. (a)



(b).

The posterior mean is 0.4495196 and the posterior variance is 0.0002376864

(c).

The posterior median is 0.449

(d).

The 95% central posterior interval is (0.419, 0.48).

library(triangle)

# The value of theta

thetas <- seq(0, 1, 0.001)

# Set prior P(theta), simulate the witch's hat distribution

prior\_density <- 0.5 \* dunif(thetas, 0, 1) + 0.5 \* dtriangle(thetas, 0.385, 0.585)

names(prior\_density) <- thetas

# Plot the Prior dist of theta

plot(prior\_density ~ thetas, ylim = c(0, 6), type = 'l')

# The density of y|theta

likelihood <- lapply(thetas, function(x) {dbinom(437, size = 980, prob = x)})

likelihood <- unlist(likelihood)

names(likelihood) <- thetas

# The marginal dist of y

marginal\_y <- sum(prior\_density \* likelihood)

# Posterior dist

post\_theta\_prob <- prior\_density \* likelihood / marginal\_y

post\_theta\_density <- post\_theta\_prob / (0.001 \* sum(post\_theta\_prob))

names(post\_theta\_prob) <- thetas

names(post\_theta\_density) <- thetas

# Plot of posterior density

plot(post\_theta\_density ~ thetas, ylim = c(0, 1.1 \* max(post\_theta\_density)), type = 'l')

# posterior mean and variance

mean <- sum(thetas \* post\_theta\_prob)

variance <- sum(thetas ^ 2 \* post\_theta\_prob) - mean ^ 2

# posterior quantile function

quantile <- function(percentage) {

cummu\_prob\_1 <- 0

cummu\_prob\_2 <- 0

for (theta in thetas) {

cummu\_prob\_1 <- cummu\_prob\_1 + post\_theta\_prob[as.character(theta)]

if ((cummu\_prob\_1 > percentage) && (cummu\_prob\_2) < percentage ) {

print(theta)

break

} else {

cummu\_prob\_2 <- cummu\_prob\_2 + post\_theta\_prob[as.character(theta)]

}

}

}

# Posterior median

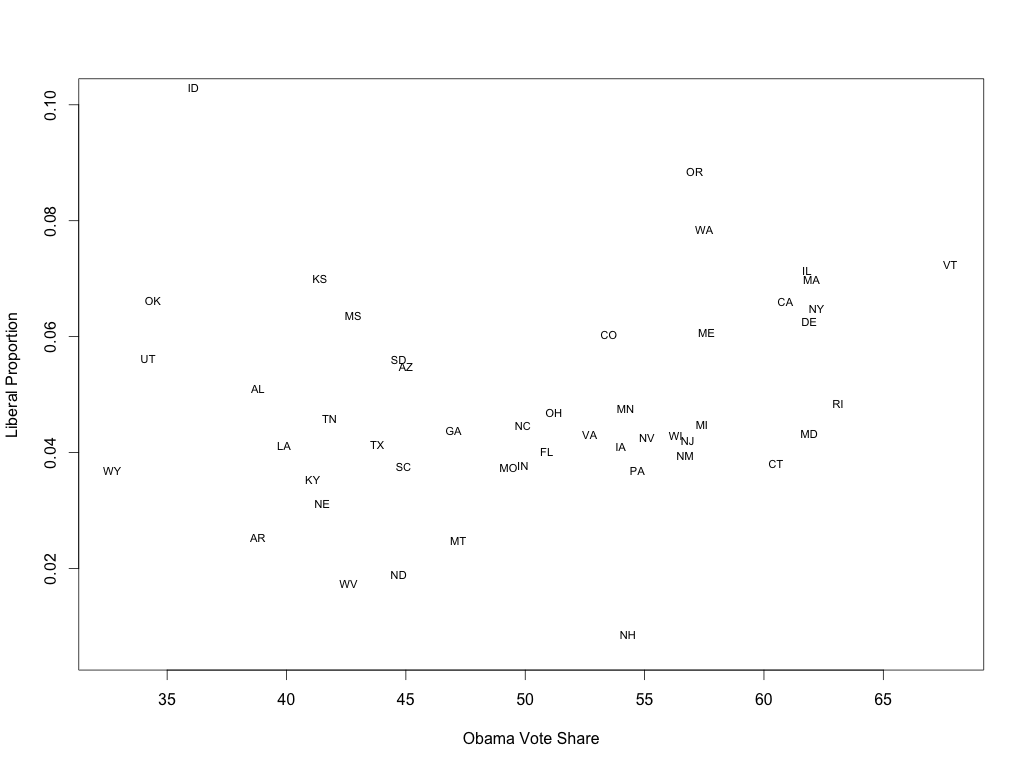
quantile(0.5)

# 95% central posterior interval

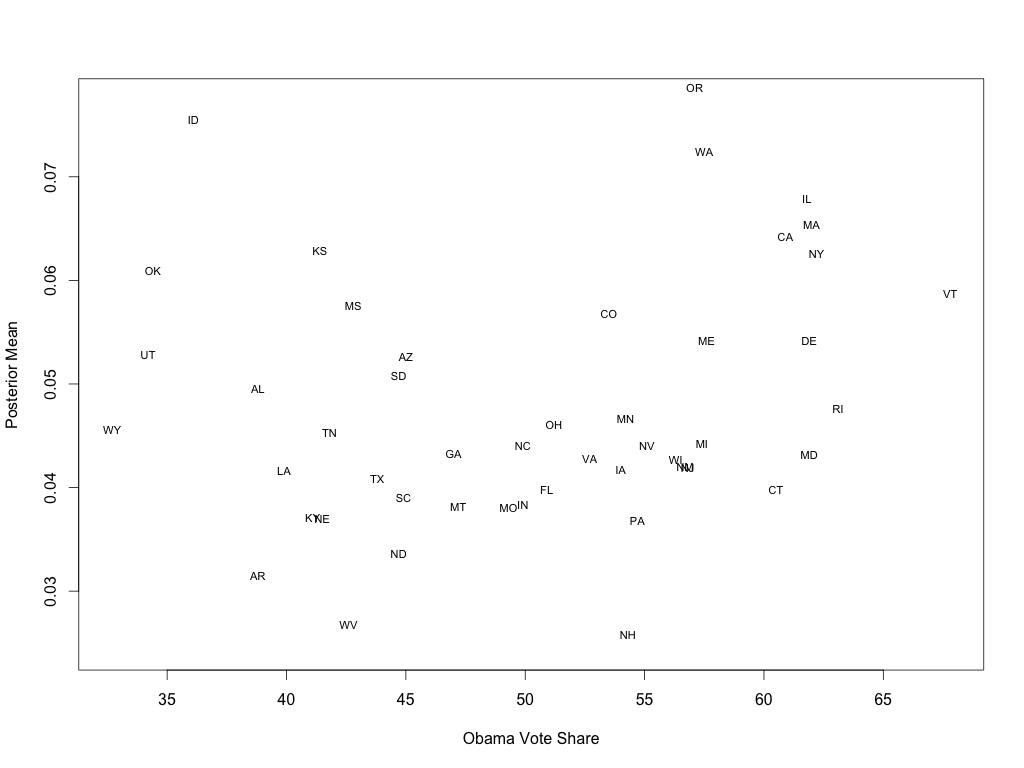
quantile(0.025)

quantile(0.975)

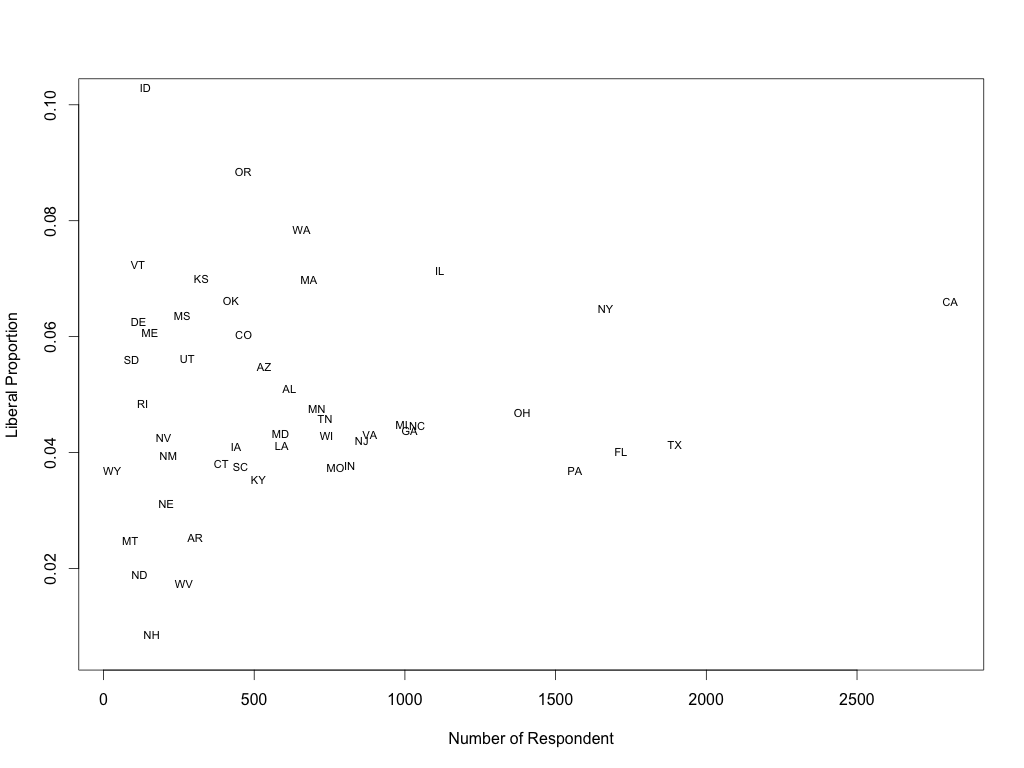
1.5(a)

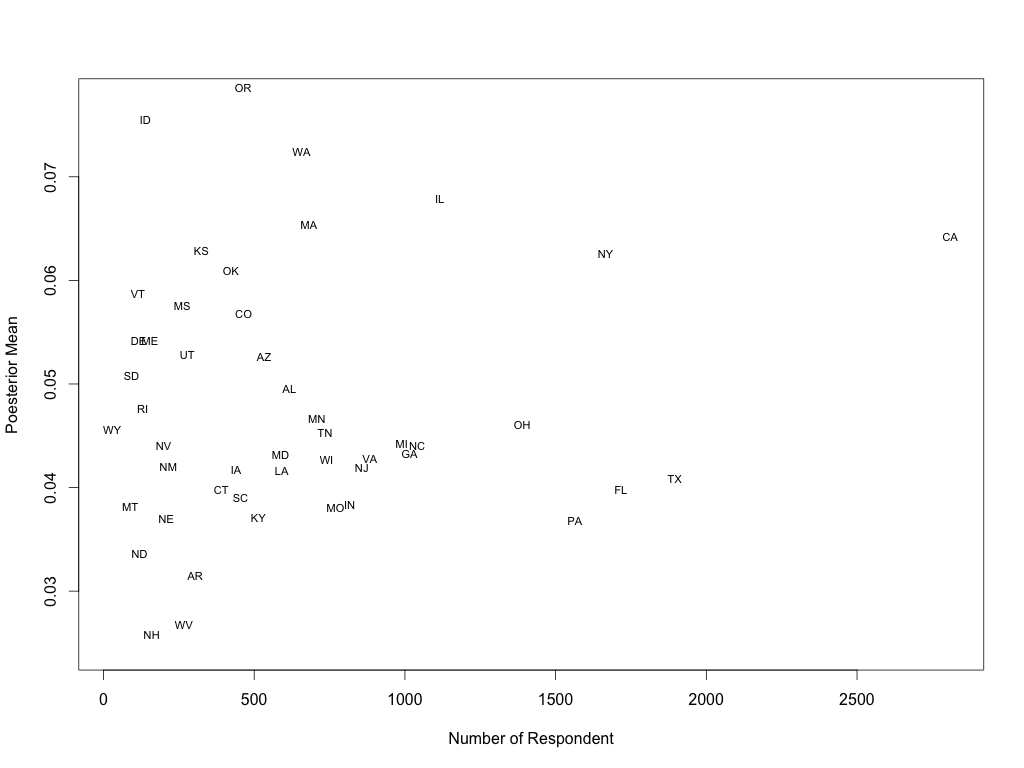


(b)



(c)





library(foreign)

library(plyr)

library(dplyr)

library(openintro)

# read data and select useful features

elect.data <- read.dta('pew\_research\_center\_june\_elect\_wknd\_data.dta')

Obama\_votes <- read.csv('2008ElectionResult.csv')

elect.data <- elect.data[,c('ideo', 'state')]

Obama\_votes <- Obama\_votes[, c('state', 'vote\_Obama\_pct')]

# Exclude Alaska, Hawaii and Washington DC and rows with NA

elect.data <- filter(elect.data, !(state %in% c('alaska', 'washington dc', 'hawaii')))

elect.data <- elect.data[complete.cases(elect.data),]

#Transfer state name to state abbreviation

elect.data$state <- state2abbr(elect.data$state)

Obama\_votes$state <- state2abbr(Obama\_votes$state)

# split the election data by state, calculate the proportion of very liberal people

# and combine together

liberal\_proportion <- ddply(elect.data, 'state', function(dataframe){

liberal\_people <- dataframe[dataframe$ideo == 'very liberal',]

proportion <- nrow(liberal\_people) / nrow(dataframe)

data.frame(liberal\_people = nrow(liberal\_people), total\_people = nrow(dataframe), state\_proportion = proportion)

})

# Combine liberal\_probortion dataset and Obama\_votes dataset by state abbreviation

liberal\_proportion\_vs\_Obama\_vote <- merge(liberal\_proportion, Obama\_votes, by = 'state')

# Plot the liberal proportion vs Obama vote

plot(liberal\_proportion\_vs\_Obama\_vote$vote\_Obama\_pct, liberal\_proportion\_vs\_Obama\_vote$state\_proportion,

ylab = 'Liberal Proportion', xlab = 'Obama Vote Share', type = 'n')

text(liberal\_proportion\_vs\_Obama\_vote$vote\_Obama\_pct, liberal\_proportion\_vs\_Obama\_vote$state\_proportion,

labels = liberal\_proportion\_vs\_Obama\_vote$state, cex = 0.7, pos = 3)

# Estimate the prior

prior\_mean <- mean(liberal\_proportion$state\_proportion)

prior\_var <- var(liberal\_proportion$state\_proportion)

alpha <- (prior\_mean ^ 2) \* (1 - prior\_mean) / prior\_var - prior\_mean

beta <- alpha \* (1 - prior\_mean) / prior\_mean

# Posterior Mean for each state

liberal\_proportion\_vs\_Obama\_vote$posterior\_mean <- (alpha + liberal\_proportion$liberal\_people) / (alpha + beta + liberal\_proportion$total\_people)

# Plot the posterior mean vs Obama vote

plot(liberal\_proportion\_vs\_Obama\_vote$vote\_Obama\_pct, liberal\_proportion\_vs\_Obama\_vote$posterior\_mean,

ylab = 'Posterior Mean', xlab = 'Obama Vote Share', type = 'n')

text(liberal\_proportion\_vs\_Obama\_vote$vote\_Obama\_pct, liberal\_proportion\_vs\_Obama\_vote$posterior\_mean,

labels = liberal\_proportion\_vs\_Obama\_vote$state, cex = 0.7, pos = 3)

# Plot the liberal Mean vs number of respondents

plot(liberal\_proportion\_vs\_Obama\_vote$total\_people, liberal\_proportion\_vs\_Obama\_vote$state\_proportion,

ylab = 'Liberal Proportion', xlab = 'Number of Respondent', type = 'n')

text(liberal\_proportion\_vs\_Obama\_vote$total\_people, liberal\_proportion\_vs\_Obama\_vote$state\_proportion,

labels = liberal\_proportion\_vs\_Obama\_vote$state, cex = 0.7, pos = 3)

# Plot the liberal Mean vs Number of Respondents

plot(liberal\_proportion\_vs\_Obama\_vote$total\_people, liberal\_proportion\_vs\_Obama\_vote$posterior\_mean,

ylab = 'Poesterior Mean', xlab = 'Number of Respondent', type = 'n')

text(liberal\_proportion\_vs\_Obama\_vote$total\_people, liberal\_proportion\_vs\_Obama\_vote$posterior\_mean,

labels = liberal\_proportion\_vs\_Obama\_vote$state, cex = 0.7, pos = 3)