Bayesian Final Project

Sunny Jing

6/6/2017

```
datapath="/Users/dongpingjing/UChicago/Spring 2017/Bayesian Methods/Course Project/Data"
dat<-read.csv(paste(datapath,"MScA_32014_BayesianMethods_CourseProjectData.csv",sep="/"))
head(dat)
```

```
##      sex    race   age       education state y
## 1 1.Male 1.White 18-24     1.NoCollege    GA 0
## 2 1.Male 1.White 25-34     1.NoCollege    AZ 0
## 3 1.Male 1.White 25-34   2.SomeCollege    SD 0
## 4 1.Male 1.White 18-24 3.CollegeOrMore    SC 0
## 5 1.Male 1.White 18-24 3.CollegeOrMore    SC 0
## 6 1.Male 1.White 18-24 3.CollegeOrMore    SC 0
```

```
unique(dat$sex)
```

```
## [1] 1.Male   2.Female
## Levels: 1.Male 2.Female
```

```
unique(dat$race)
```

```
## [1] 1.White    2.Black    3.Hispanic 4.Other
## Levels: 1.White 2.Black 3.Hispanic 4.Other
```

```
unique(dat$age)
```

```
## [1] 18-24 25-34 35-44 45-54 55+
## Levels: 18-24 25-34 35-44 45-54 55+
```

```
unique(dat$education)
```

```
## [1] 1.NoCollege     2.SomeCollege   3.CollegeOrMore
## Levels: 1.NoCollege 2.SomeCollege 3.CollegeOrMore
```

```
unique(dat$state)
```

```
##  [1] GA AZ SD SC AL VA KS TN IA ME AR WA CT OH PA MA NH MD WI NE MS CA NY
## [24] DE MN MI ND ID HI IN VT FL OK UT NM KY LA WY DC RI IL OR NJ MT MO CO
## [47] NV WV AK NC TX
## 51 Levels: AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA ... WY
```

```
dataList=list(N = length(dat$y),
              y = dat$y,
              sex = as.integer(dat$sex),
              NSex = nlevels(dat$sex),
              race = as.integer(dat$race),
              NRace = nlevels(dat$race),
              age = as.integer(dat$age),
              NAge = nlevels(dat$age),
              education = as.integer(dat$education),
              NEducation = nlevels(dat$education),
              state = as.integer(dat$state),
              NState = nlevels(dat$state))
```

After running MCMC with the these data and the model obtain a Markov chain posterior sample for 870 parameters including 2-way interactions. Each Markov chain of the stan object obama_fit has length 36000.

```
library (rstan)
```

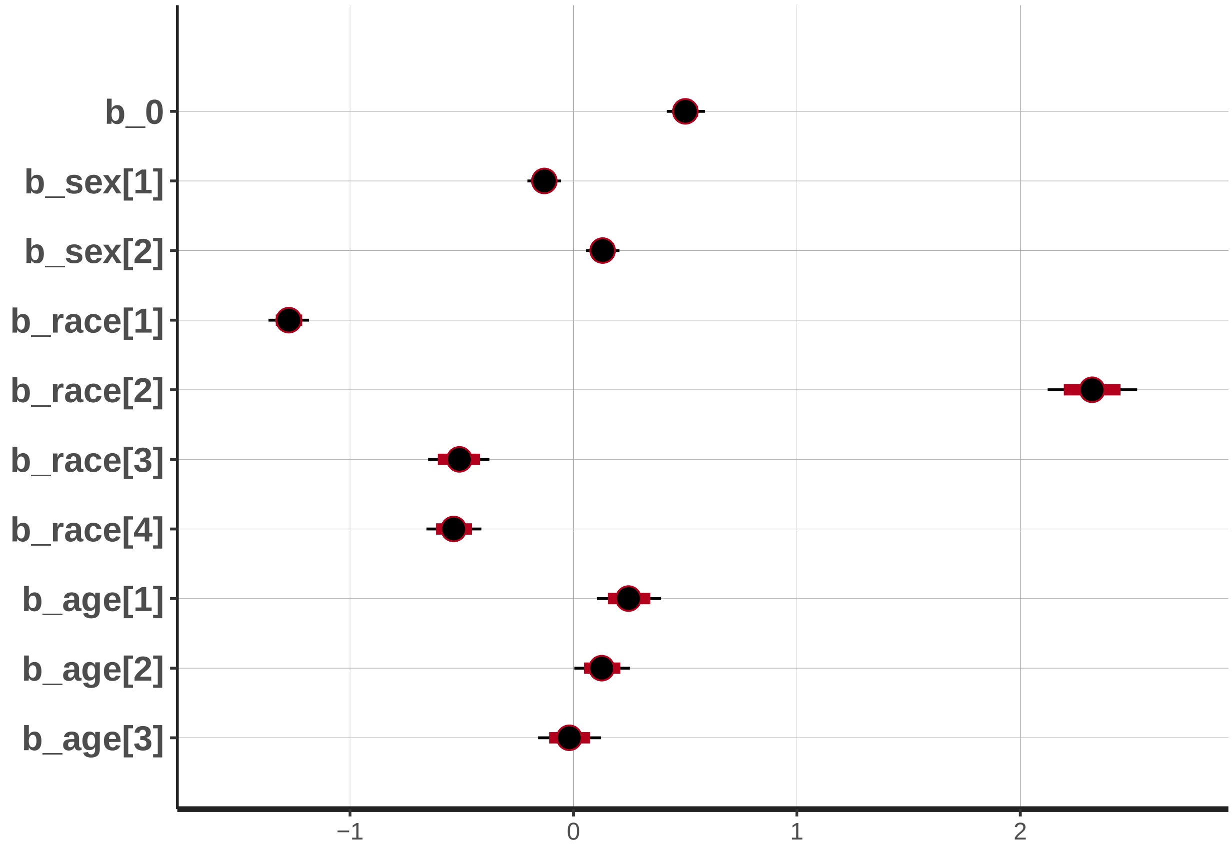
```
## Loading required package: ggplot2

## Loading required package: StanHeaders

## rstan (Version 2.15.1, packaged: 2017-04-19 05:03:57 UTC, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## rstan_options(auto_write = TRUE)
## options(mc.cores = parallel::detectCores())
```

```
# rstan_options(auto_write = TRUE)

# t1 <- Sys.time()
# message('Started fitting at ', t1)

# model <- stan_model(file="/Users/dongpingjing/UChicago/Spring 2017/Bayesian Methods/Course Project/R/
# t2 <- Sys.time()
# message('Model compiled at ', t2)

# obama_fit <- sampling(model,
#                       data=dataList,
#                       pars=c('b_0','b_sex', 'b_race', 'b_age', 'b_education', 'b_state',
#                              'b_sex_race', 'b_sex_age', 'b_sex_education', 'b_sex_state',
#                              'b_race_age', 'b_race_education', 'b_race_state',
#                              'b_age_education', 'b_age_state', 'b_education_state',
#                              'var_0', 'var_sex', 'var_race', 'var_age', 'var_education', 'var_state',
#                              'var_sex_race', 'var_sex_age', 'var_sex_education', 'var_sex_state',
#                              'var_race_age', 'var_race_education', 'var_race_state',
#                              'var_age_education', 'var_age_state', 'var_education_state',
#                              'nu', 'sigma'
#                       ),
#                       control=list(adapt_delta=0.99, max_treedepth=12),
#                       iter=1000, chains = 4, cores = 4,
#                       verbose = F)

# t3 <- Sys.time()
# message('Finished fitting at ', t3)
# message('Time elapsed: ', difftime(t3,t1, units = 'h'), ' hours')

# save fitted model:
# file_suffix <- strftime(t2, format='%Y%m%d_%H%M%S')
# fn <- paste0('./fit_ext_', file_suffix, '.Rdata')
# save(obama_fit, file=fn)
# message('Saved model to ', fn, '. Goodbye!')
```

```
load("/Users/dongpingjing/UChicago/Spring 2017/Bayesian Methods/Course Project/R/Run05202016-fit_201605

stan_plot (obama_fit)
```

```
## 'pars' not specified. Showing first 10 parameters by default.

## ci_level: 0.8 (80% intervals)
```

```
## outer_level: 0.95 (95% intervals)
```




```
sum.obama_fit<-summary(obama_fit)$summary

MCMC<- rstan::extract(obama_fit)
```

Shinystan

```
#library(shinystan)
#launch_shinystan(obama_fit)
```

Explore sum.obama_fit - summary of obama_fit.

```
# Find parameters which are significantly different from zero: zero does not belong to 95% HDI. Show se
noselection=sum.obama_fit[,4]<0 & sum.obama_fit[,8]>0
selection=!noselection

rownames(sum.obama_fit[selection,c(4,8)])
```

```
##  [1] "b_0"                      "b_sex[1]"
##  [3] "b_sex[2]"                 "b_race[1]"
##  [5] "b_race[2]"                "b_race[3]"
##  [7] "b_race[4]"                "b_age[1]"
##  [9] "b_age[2]"                 "b_age[5]"
## [11] "b_state[2]"               "b_state[3]"
```

```
## [13] "b_state[8]"           "b_state[11]"
## [15] "b_state[12]"          "b_state[15]"
## [17] "b_state[16]"          "b_state[17]"
## [19] "b_state[19]"          "b_state[20]"
## [21] "b_state[24]"          "b_state[26]"
## [23] "b_state[32]"          "b_state[37]"
## [25] "b_state[41]"          "b_state[43]"
## [27] "b_state[44]"          "b_state[47]"
## [29] "b_state[48]"          "b_sex_age[1,2]"
## [31] "b_sex_age[1,5]"       "b_sex_age[2,2]"
## [33] "b_sex_age[2,5]"       "b_race_age[1,5]"
## [35] "b_race_age[2,1]"      "b_race_age[4,1]"
## [37] "b_race_age[4,2]"      "b_race_age[4,5]"
## [39] "b_race_education[1,3]" "b_race_state[1,2]"
## [41] "b_race_state[1,11]"   "b_race_state[1,16]"
## [43] "b_race_state[1,19]"   "b_race_state[1,26]"
## [45] "b_race_state[1,38]"   "b_race_state[1,41]"
## [47] "b_race_state[2,41]"   "b_race_state[4,48]"
## [49] "b_age_education[5,3]" "sd_sex"
## [51] "sd_race"              "sd_age"
## [53] "sd_education"         "sd_state"
## [55] "sd_sex_race"          "sd_sex_age"
## [57] "sd_sex_education"     "sd_sex_state"
## [59] "sd_race_age"          "sd_race_education"
## [61] "sd_race_state"        "sd_age_education"
## [63] "sd_age_state"         "sd_education_state"
## [65] "lp__"
```

Questions of the project:

1. Find groups from which the main support for Obama came in 2012

Obama gained main support from Group of race black.

Female, race of lack, age of 18-24, with education of college and more tended to support Obama more.

```
odds<-exp(sum.obama_fit[,1])
baseline=odds[1]
length(odds)
```

```
## [1] 867
```

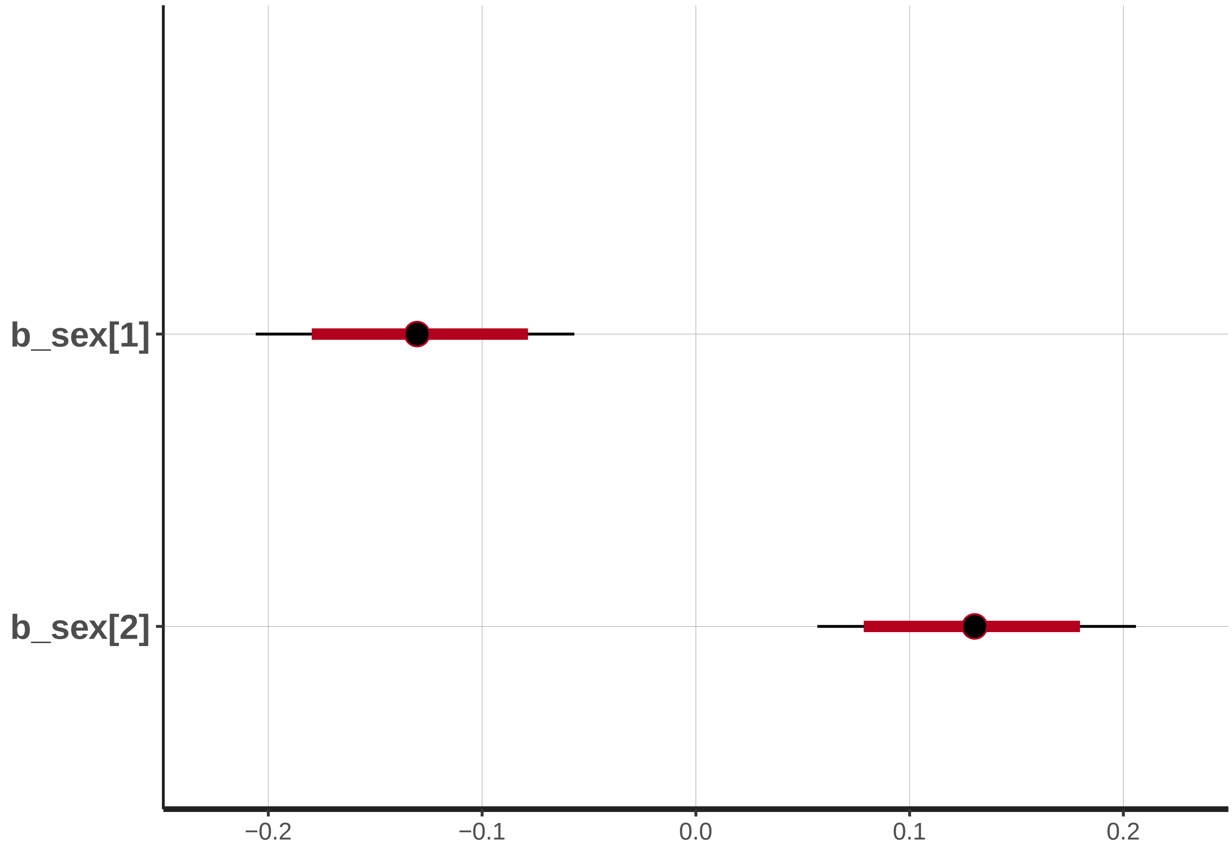
```
oddsratio<-sapply(2:851,function(x) prod(baseline,odds[x]))
names(oddsratio)<-names(odds)[2:851]

highestodds<-oddsratio[order(oddsratio,decreasing=TRUE)][1:5]
names(highestodds)<-c("Black","TN","WA","NV","Black in IL")
highestodds
```

```
##       Black          TN          WA          NV Black in IL
##   16.813310    3.755811    3.245583    3.195265    3.093406
```

```
# By group
stan_plot(obama_fit,c("b_sex"))
```

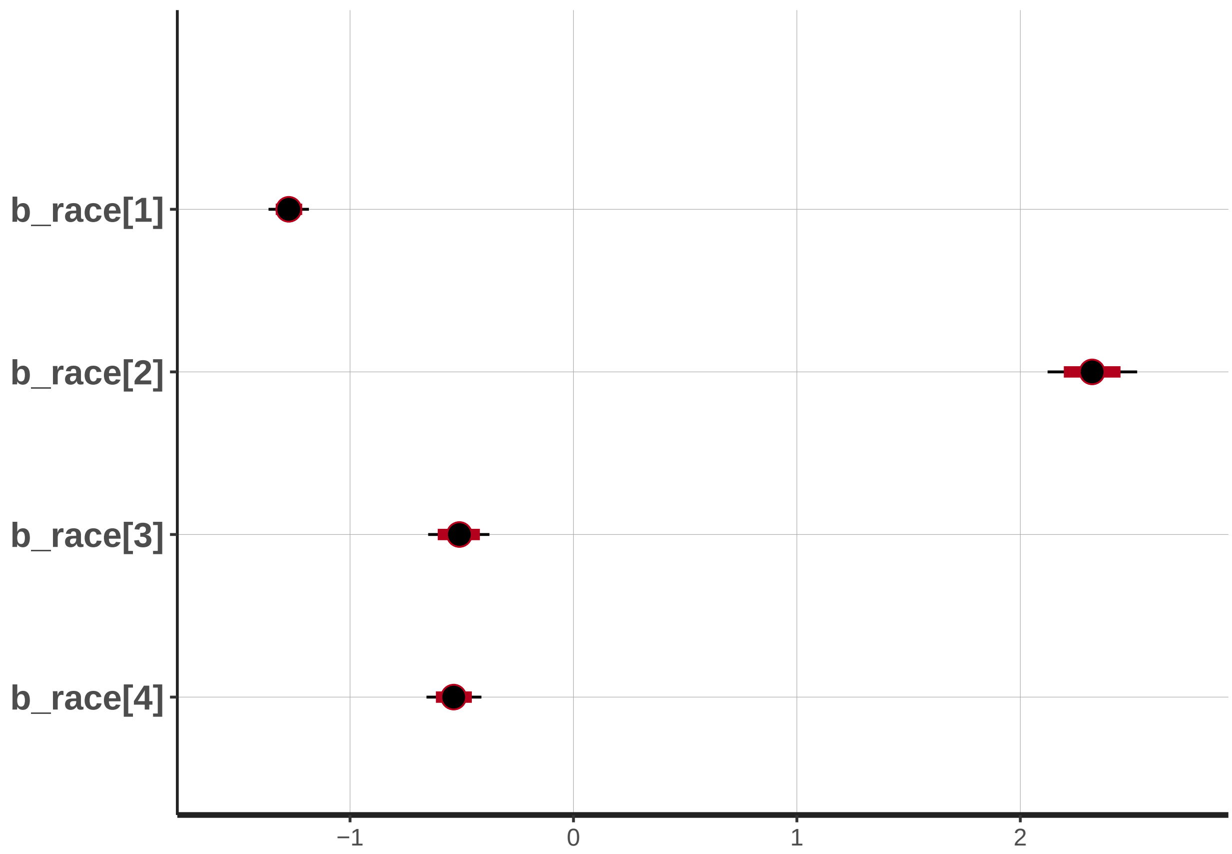
```
## ci_level: 0.8 (80% intervals)
```

```
## outer_level: 0.95 (95% intervals)
```



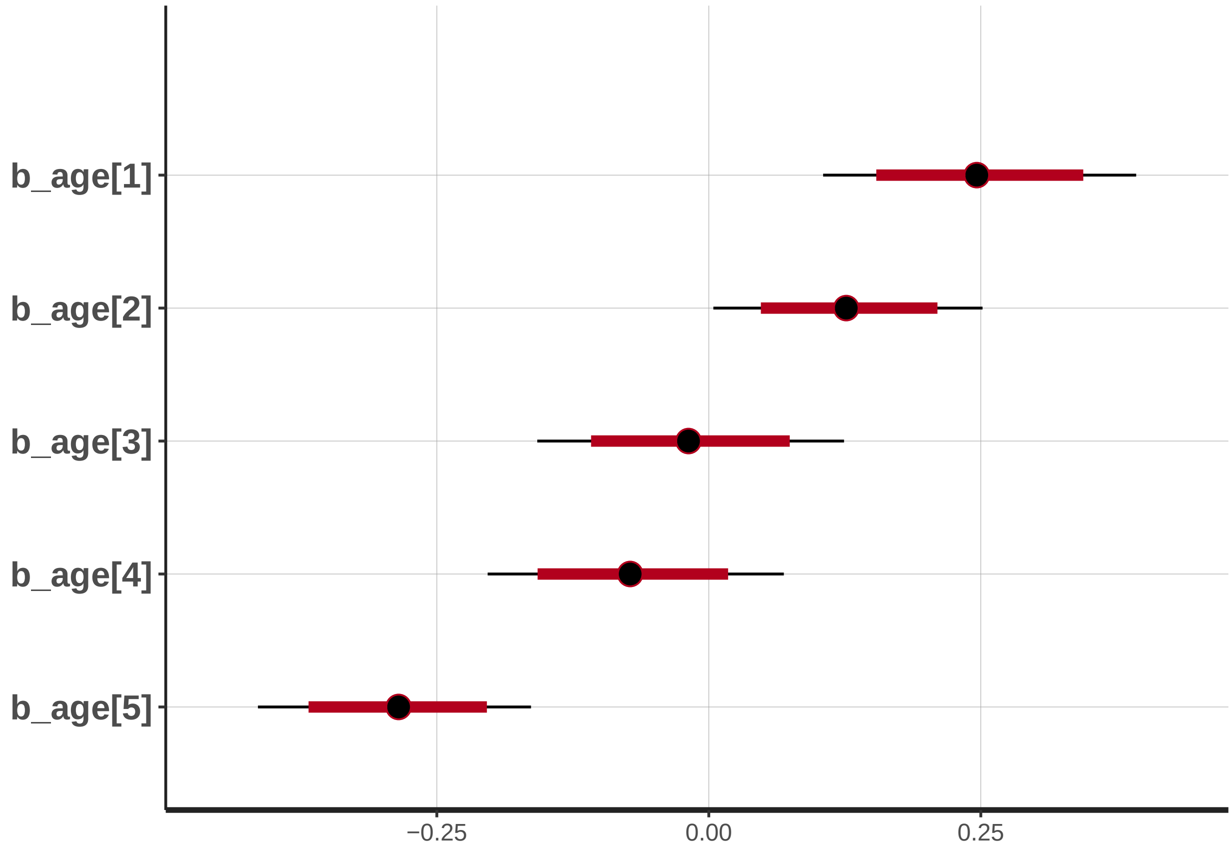

```r
stan_plot(obama_fit,c("b_race"))
```

```
## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)
```



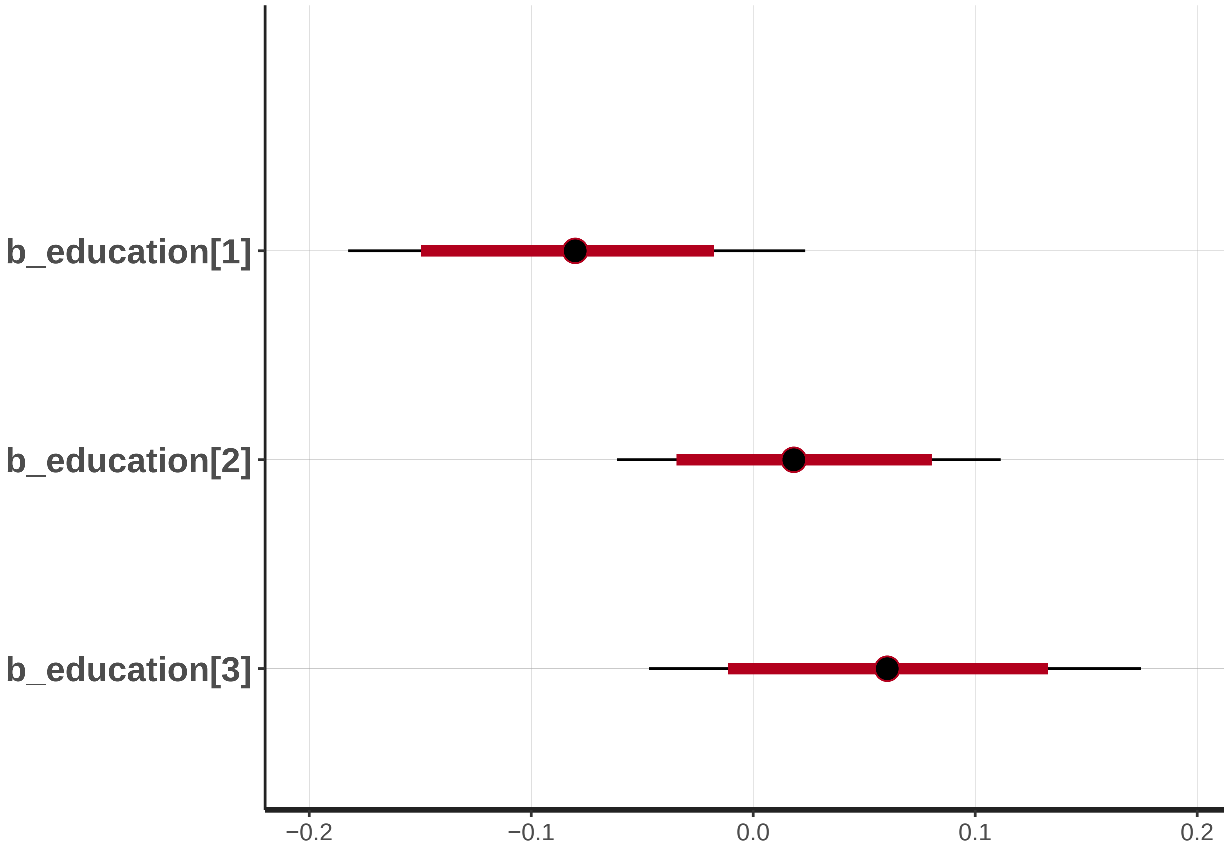

```
stan_plot(obama_fit,c("b_age"))
```

```
## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)
```




```r
stan_plot(obama_fit,c("b_education"))
```

```
## ci_level: 0.8 (80% intervals)
## outer_level: 0.95 (95% intervals)
```



2. Find groups of the lowest odds of approval

Group of the lowest odds of approval is the "White-Race" group, followed by the group of "White-Race" in WI, then the state "MI", then the group of white people in Michigan, and finally the "Other" race.

```
lowestodds<-oddsratio[order(oddsratio,decreasing=FALSE)][1:5]
names(lowestodds)<-c("White","White in WI","MI","White in MI","Other")
lowestodds
```

```
##       White White in WI          MI White in MI       Other
##   0.4618717   0.8460250   0.9493148   0.9510878   0.9584477
```

```
prod(odds[c(1,2,4,6)])
```

```
## [1] 0.2432751
```

3. Search for information on main support and no support for Hillary Clinton in 2016 and try to identify the dynamics between 2012 and 2016

"Women supported Clinton over Trump by 54% to 42%. This is about the same as the Democratic advantage among women in 2012 (55% Obama vs. 44% Romney) and 2008 (56% Obama vs. 43% McCain)." – Behind Trump's victory: Divisions by race, gender, education

"White non-Hispanic voters preferred Trump over Clinton by 21 percentage points (58% to 37%)" –Edison Research for the National Election Pool.

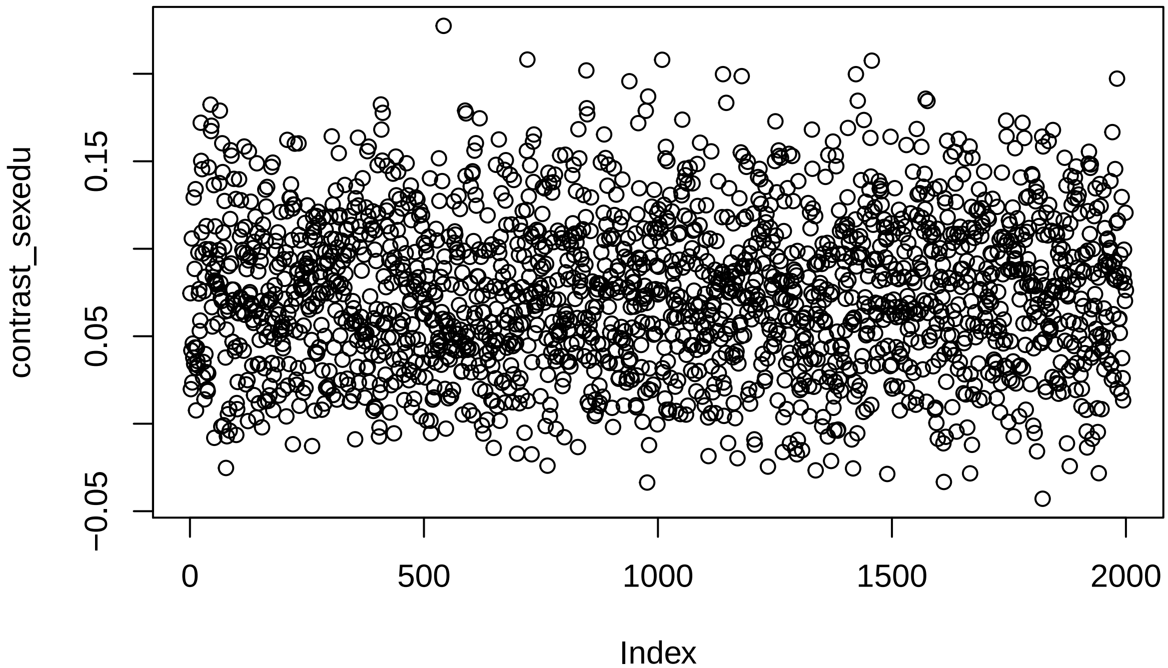
"In the 2016 election, a wide gap in presidential preferences emerged between those with and without a college degree. College graduates backed Clinton by a 9-point margin (52%-43%), while those without a college degree backed Trump 52%-44%." – Behind Trump's victory: Divisions by race, gender, education

The biggest dynamics change from the 2012 campaign to the 2016 campaign is that Hillary Clinton did not run as strongly among black voters as Obama did in 2012.

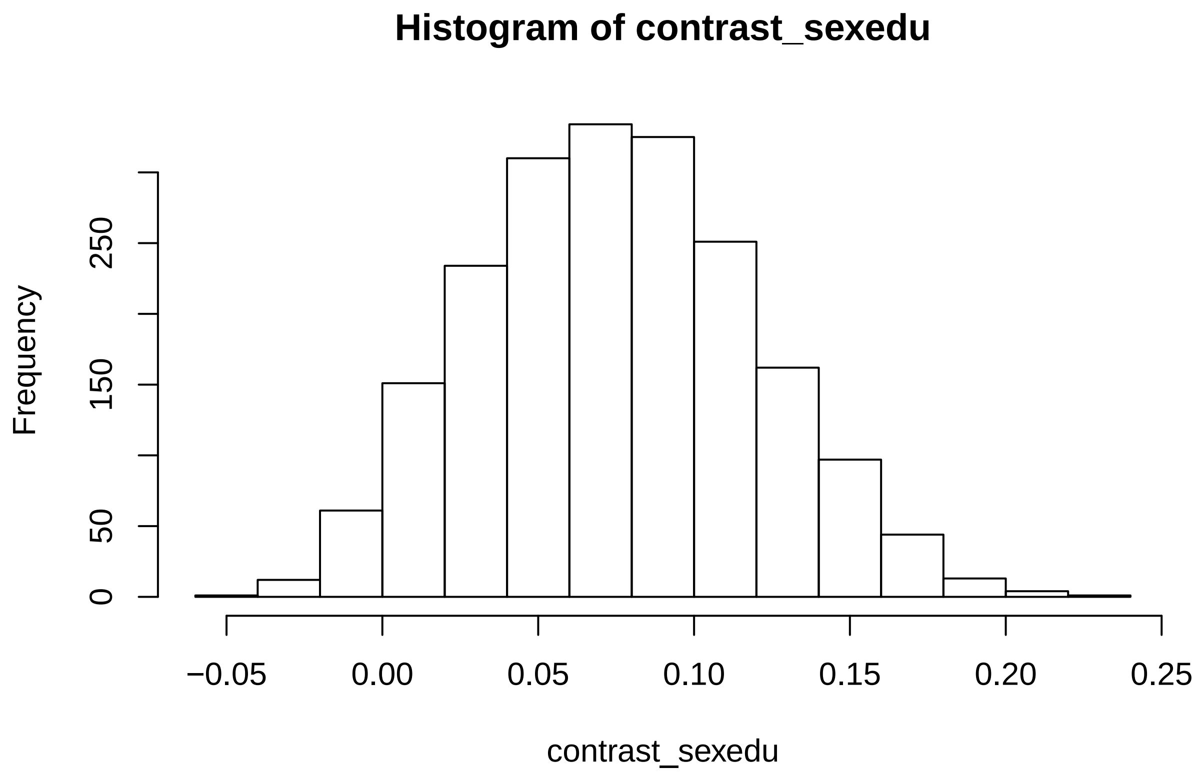
4.What else you find interesting in the results?

Among the highly educated group of college and more, more women support Obama than men.

```
# Contrast between high educated female and  high educated male:
contrast_sexedu<-MCMC$b_sex_education[,2,3]-MCMC$b_sex_education[,1,3]
plot(contrast_sexedu)
```




```
hist(contrast_sexedu)
```

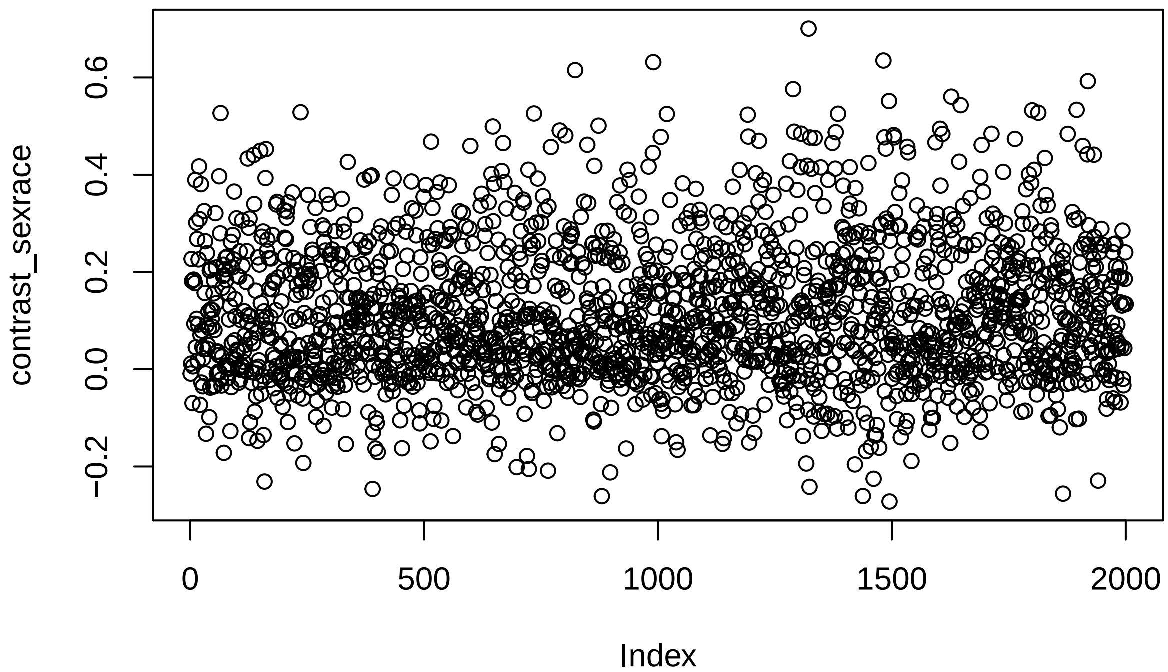



```
suppressWarnings(library(HDInterval))
(hdiContrast_sexedu<-hdi(contrast_sexedu))
```

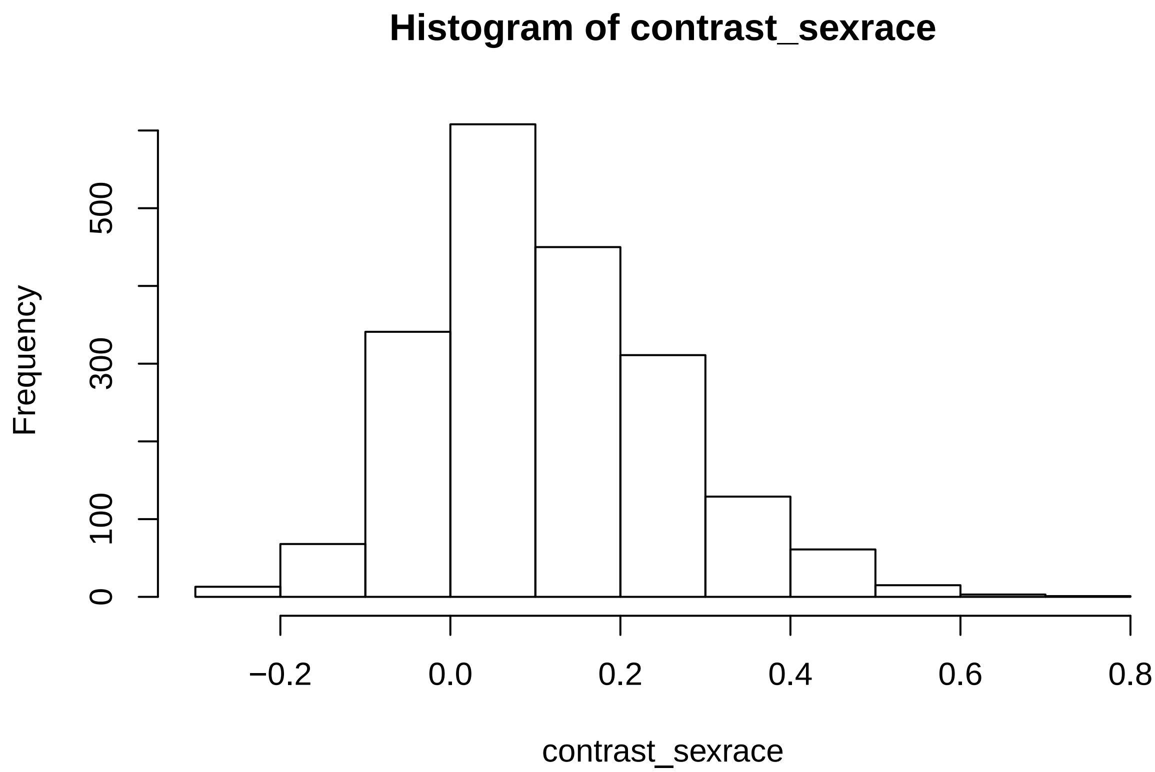
```
##        lower        upper
## -0.008939893  0.158482600
## attr(,"credMass")
## [1] 0.95
```

Among the Black race, women are men equally support Obama.

```
contrast_sexrace<-MCMC$b_sex_race[,2,2]-MCMC$b_sex_race[,1,2]
plot(contrast_sexrace)
```




```
hist(contrast_sexrace)
```

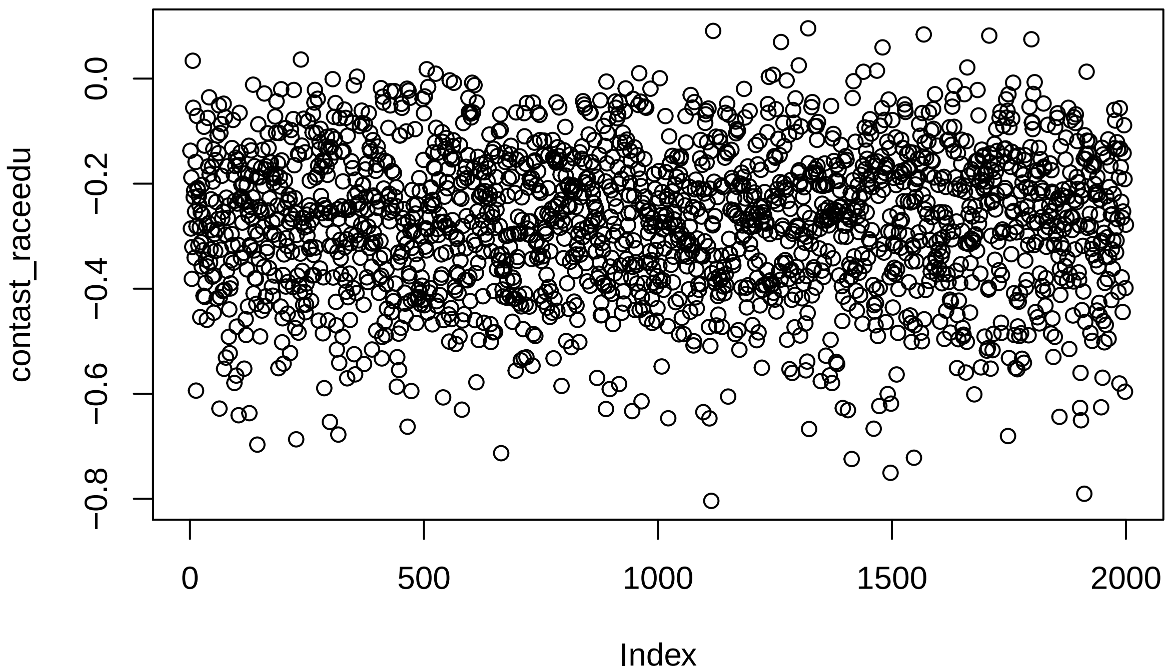



```
(hdiContrast_sexrace<-hdi(contrast_sexrace))
```

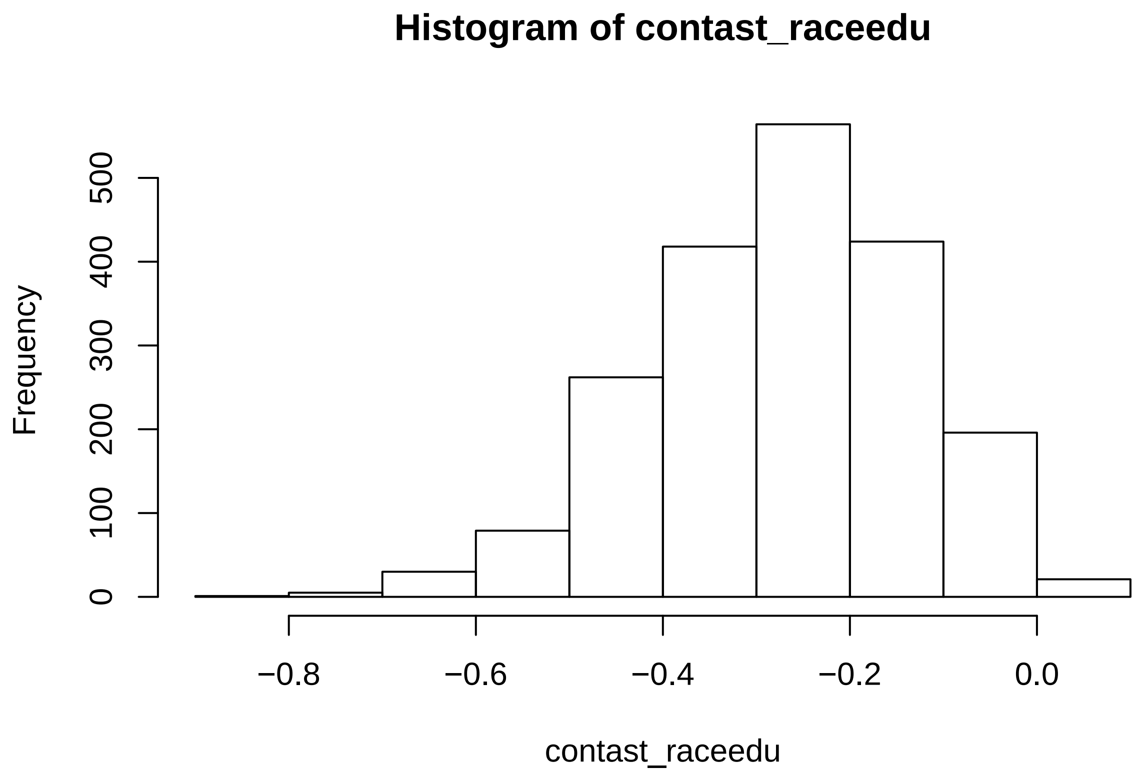
```
##      lower      upper
## -0.1418124  0.4185095
## attr(,"credMass")
## [1] 0.95
```

Interestingly, among the highly educated group, white voters support Obama more than black voters.

```
contast_raceedu<-MCMC$b_race_education[,2,3]-MCMC$b_race_education[,1,3]
plot(contast_raceedu)
```




```r
hist(contast_raceedu)
```




```r
(hdiContrast_raceduc<-hdi(contast_raceedu))
```

```
##      lower      upper
## -0.5533191 -0.0192069
## attr(,"credMass")
## [1] 0.95
```