Bayesian Project

James Lee

12/5/2019

# Bayesian Project

**Hierarchical Model** The data of the project contains 5 factors: Sex (2 levels) Race (4 levels) Age (5 levels) Education (3 levels) State (51 levels) The total number of combinations of levels is 6120.

dataPath <- 'C:/Users/jalee/Desktop/Bayesian/'  
dat<-read.csv(paste(dataPath,"MScA\_32014\_BayesianMethods\_CourseProjectData.csv",sep="/"))  
str(dat)

## 'data.frame': 23223 obs. of 6 variables:  
## $ sex : Factor w/ 2 levels "1.Male","2.Female": 1 1 1 1 1 1 1 1 1 1 ...  
## $ race : Factor w/ 4 levels "1.White","2.Black",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ age : Factor w/ 5 levels "18-24","25-34",..: 1 2 2 1 1 1 1 2 2 2 ...  
## $ education: Factor w/ 3 levels "1.NoCollege",..: 1 1 2 3 3 3 1 3 3 3 ...  
## $ state : Factor w/ 51 levels "AK","AL","AR",..: 11 4 42 41 41 41 2 41 41 41 ...  
## $ y : int 0 0 0 0 0 0 1 0 1 0 ...

**2 Model** 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.

load(paste(dataPath,"fit\_ext\_20160527\_103144.Rdata",sep="/"))  
library(shinystan)

## Warning: package 'shinystan' was built under R version 3.6.1

## Loading required package: shiny

## Registered S3 method overwritten by 'xts':  
## method from  
## as.zoo.xts zoo

##   
## This is shinystan version 2.5.0

**3 Analysis**

library(HDInterval)

## Warning: package 'HDInterval' was built under R version 3.6.1

library(rstan)

## Warning: package 'rstan' was built under R version 3.6.1

## Loading required package: StanHeaders

## Warning: package 'StanHeaders' was built under R version 3.6.1

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.6.1

## rstan (Version 2.19.2, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling  
## options(mc.cores = parallel::detectCores()).  
## To avoid recompilation of unchanged Stan programs, we recommend calling  
## rstan\_options(auto\_write = TRUE)

## For improved execution time, we recommend calling  
## Sys.setenv(LOCAL\_CPPFLAGS = '-march=native')  
## although this causes Stan to throw an error on a few processors.

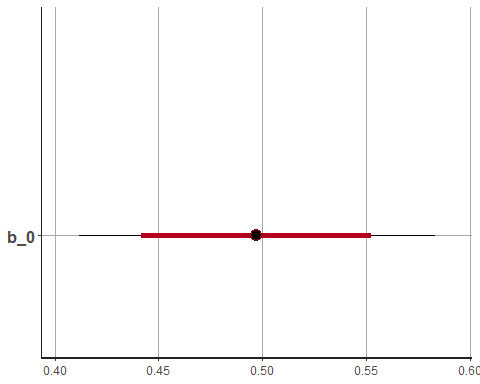
MCMC<-rstan::extract(obama\_fit)  
sum.obama\_fit<-rstan::summary(obama\_fit)[[1]]  
HDI95<-sum.obama\_fit[,c(4,8)]  
selection<-apply(HDI95,1,function(z) findInterval(0,z)!=1)  
sum.obama\_fit[selection,c(4,8)]

## 2.5% 97.5%  
## b\_0 4.116903e-01 5.832719e-01  
## b\_sex[1] -1.905134e-01 -5.765717e-02  
## b\_sex[2] 5.765717e-02 1.905134e-01  
## b\_race[1] -1.359407e+00 -1.178978e+00  
## b\_race[2] 2.111867e+00 2.506630e+00  
## b\_race[3] -6.431486e-01 -3.649229e-01  
## b\_race[4] -6.578031e-01 -4.089638e-01  
## b\_age[1] 1.098589e-01 3.798599e-01  
## b\_age[2] 6.147626e-03 2.474967e-01  
## b\_age[5] -3.951658e-01 -1.537930e-01  
## b\_state[2] -9.094052e-01 -1.456489e-01  
## b\_state[3] -7.231898e-01 -1.139872e-02  
## b\_state[8] 3.422349e-01 1.258903e+00  
## b\_state[12] 3.068338e-01 9.842034e-01  
## b\_state[15] 1.477354e-01 8.018328e-01  
## b\_state[16] -8.116920e-01 -9.055859e-02  
## b\_state[17] -7.760941e-01 -4.964744e-02  
## b\_state[19] -8.609764e-01 -1.909234e-01  
## b\_state[20] 5.869325e-02 7.604090e-01  
## b\_state[24] 9.907271e-02 8.626887e-01  
## b\_state[26] -9.254027e-01 -1.464358e-01  
## b\_state[32] 1.472328e-02 6.595229e-01  
## b\_state[37] -7.065583e-01 -2.847404e-02  
## b\_state[41] -5.295257e-01 -4.392333e-02  
## b\_state[43] -8.047343e-01 -9.874344e-02  
## b\_state[44] -6.246807e-01 -3.772476e-02  
## b\_state[47] 1.295356e-01 1.120061e+00  
## b\_state[48] 7.427813e-02 7.620203e-01  
## b\_sex\_race[1,1] 3.575282e-03 1.438617e-01  
## b\_sex\_race[2,1] -1.438617e-01 -3.575282e-03  
## b\_sex\_age[1,2] 2.491784e-02 1.613331e-01  
## b\_sex\_age[1,5] -1.600067e-01 -5.825384e-02  
## b\_sex\_age[2,2] -1.613331e-01 -2.491784e-02  
## b\_sex\_age[2,5] 5.825384e-02 1.600067e-01  
## b\_sex\_education[1,3] -8.389410e-02 -2.137467e-03  
## b\_sex\_education[2,3] 2.137467e-03 8.389410e-02  
## b\_race\_age[1,5] 1.367724e-02 2.607192e-01  
## b\_race\_age[2,1] -6.760829e-01 -3.227627e-02  
## b\_race\_age[4,1] 1.306369e-02 4.460711e-01  
## b\_race\_age[4,2] 2.618185e-03 3.772187e-01  
## b\_race\_age[4,5] -4.268007e-01 -4.351891e-02  
## b\_race\_education[1,3] 6.107326e-02 2.603283e-01  
## b\_race\_state[1,2] -8.475173e-01 -5.381702e-02  
## b\_race\_state[1,8] 4.587682e-03 9.590370e-01  
## b\_race\_state[1,11] -8.035621e-01 -1.436249e-01  
## b\_race\_state[1,16] 2.451781e-02 7.651911e-01  
## b\_race\_state[1,19] -1.048246e+00 -3.017050e-01  
## b\_race\_state[1,26] -9.712715e-01 -1.473189e-01  
## b\_race\_state[1,38] 8.870777e-02 8.227156e-01  
## b\_race\_state[1,41] -5.512252e-01 -4.569208e-02  
## b\_race\_state[2,41] 2.039491e-01 1.109687e+00  
## b\_race\_state[3,44] 1.143239e-03 7.638489e-01  
## b\_race\_state[4,48] 2.711633e-02 9.730746e-01  
## b\_age\_education[5,3] 1.771619e-02 1.543087e-01  
## var\_0 2.305762e-03 5.611773e-01  
## var\_sex 2.380204e-03 2.642263e-01  
## var\_race 5.287224e-01 7.635608e+00  
## var\_age 2.391913e-03 1.320765e-01  
## var\_education 1.939184e-03 1.220115e-01  
## var\_state 1.843261e-02 2.010558e-01  
## var\_sex\_race 2.323496e-03 9.326720e-02  
## var\_sex\_age 4.432531e-03 6.974446e-02  
## var\_sex\_education 1.884873e-03 5.751763e-02  
## var\_sex\_state 2.258017e-03 2.437896e-02  
## var\_race\_age 9.982349e-03 1.585065e-01  
## var\_race\_education 4.695783e-03 7.270962e-02  
## var\_race\_state 6.954932e-02 2.338833e-01  
## var\_age\_education 2.243573e-03 3.019117e-02  
## var\_age\_state 1.053481e-02 4.701068e-02  
## var\_education\_state 2.194384e-03 2.158250e-02  
## nu 7.911521e-01 3.673505e+00  
## sigma 6.658656e-02 1.728506e-01  
## lp\_\_ -1.215310e+04 -1.186593e+04

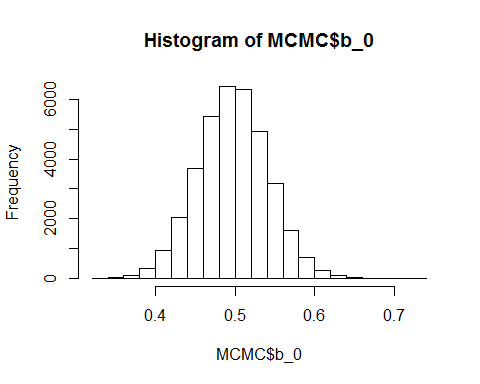
plot(obama\_fit,pars=c("b\_0"))

## ci\_level: 0.8 (80% intervals)

## outer\_level: 0.95 (95% intervals)

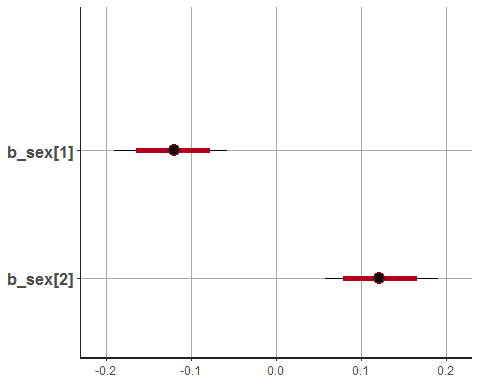


hist(MCMC$b\_0)



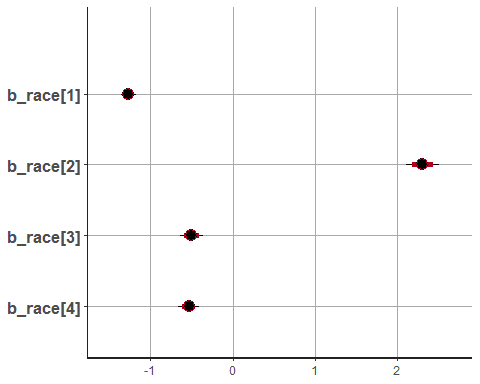
plot(obama\_fit,pars=c("b\_sex"))

## ci\_level: 0.8 (80% intervals)  
## outer\_level: 0.95 (95% intervals)



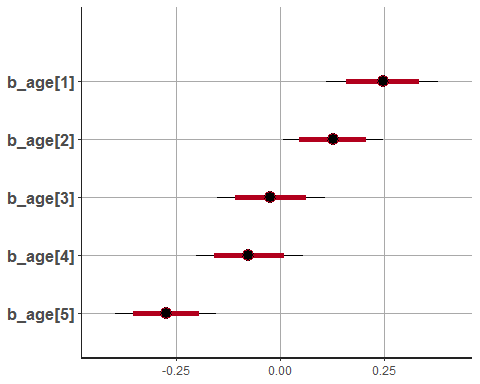
plot(obama\_fit,pars=c("b\_race"))

## ci\_level: 0.8 (80% intervals)  
## outer\_level: 0.95 (95% intervals)



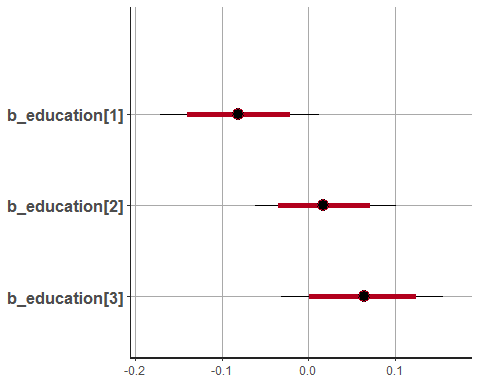
plot(obama\_fit,pars=c("b\_age"))

## ci\_level: 0.8 (80% intervals)  
## outer\_level: 0.95 (95% intervals)



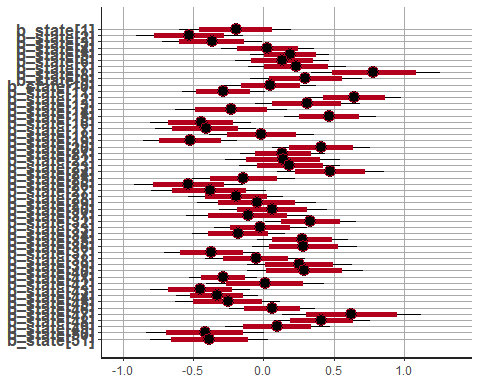
plot(obama\_fit,pars=c("b\_education"))

## ci\_level: 0.8 (80% intervals)  
## outer\_level: 0.95 (95% intervals)



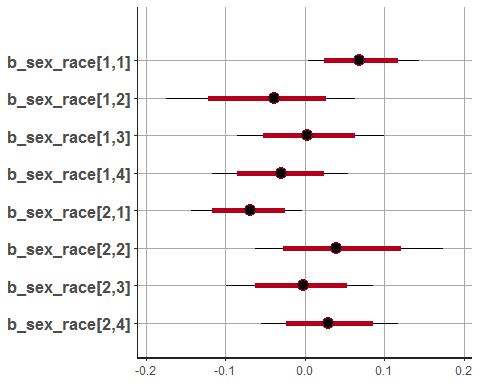
plot(obama\_fit,pars=c("b\_state"))

## ci\_level: 0.8 (80% intervals)  
## outer\_level: 0.95 (95% intervals)



plot(obama\_fit,pars=c("b\_sex\_race"))

## ci\_level: 0.8 (80% intervals)  
## outer\_level: 0.95 (95% intervals)



## Questions

1. Find the most supportive groups corresponding to each of the main effects for Barack Obama in 2012
2. Find the least supportive group corresponding to each of the main effects
3. Compare odds of approval by males grouped by race and education. Which of such subgroups shows the highest and the lowest support for the candidate
4. Answer the same question, but for females

**Main Effects** Sex (2 levels) Race (4 levels) Age (5 levels) Education (3 levels) State (51 levels)

#Sex  
sex<-MCMC$b\_sex  
colnames(sex)<-levels(dat$sex)  
#Race  
race<-MCMC$b\_race  
colnames(race)<-levels(dat$race)  
#Age  
age<-MCMC$b\_age  
colnames(age)<-levels(dat$age)  
#Education  
education<-MCMC$b\_education  
colnames(education)<-levels(dat$education)  
#State  
state<-MCMC$b\_state  
colnames(state)<-levels(dat$state)  
#race\_education (Q3,Q4)  
race\_education<-MCMC$b\_race\_education  
dimnames(race\_education)[[2]]<-levels(dat$race)  
dimnames(race\_education)[[3]]<-levels(dat$education)  
#b\_0  
intercept<-MCMC$b\_0

*Sex*

sex\_means<-apply(sex,2,function(z) mean(z+intercept))  
sex\_HDI<-apply(sex,2,function(z) hdi(z+intercept))  
sexOdds<-rbind(mean=exp(sex\_means),exp(sex\_HDI))  
sexOdds

## 1.Male 2.Female  
## mean 1.455949 1.856286  
## lower 1.294588 1.679686  
## upper 1.630791 2.050351

Females were more supportive than males.

*Race*

race\_means<-apply(race,2,function(z) mean(z+intercept))  
race\_HDI<-apply(race,2,function(z) hdi(z+intercept))  
raceOdds<-rbind(mean=exp(race\_means),exp(race\_HDI))  
raceOdds

## 1.White 2.Black 3.Hispanic 4.Other  
## mean 0.4620976 16.50647 0.9927362 0.9646288  
## lower 0.4430622 12.84272 0.8443836 0.8432129  
## upper 0.4824002 21.37102 1.1649948 1.1014265

which.max(raceOdds[1,])

## 2.Black   
## 2

Black population was the most supportive

*Age*

age\_means<-apply(age,2,function(z) mean(z+intercept))  
age\_HDI<-apply(age,2,function(z) hdi(z+intercept))  
ageOdds<-rbind(mean=exp(age\_means),exp(age\_HDI))  
ageOdds

## 18-24 25-34 35-44 45-54 55+  
## mean 2.103702 1.866141 1.605320 1.523512 1.250665  
## lower 1.786283 1.616481 1.366515 1.308444 1.081427  
## upper 2.474256 2.164906 1.871676 1.775774 1.437299

which.max(ageOdds[1,])

## 18-24   
## 1

which.min(ageOdds[1,])

## 55+   
## 5

18-24 was the most supportive group. 55+ was the least supportive group.

*Education*

education\_means<-apply(education,2,function(z) mean(z+intercept))  
education\_HDI<-apply(education,2,function(z) hdi(z+intercept))  
educationOdds<-rbind(mean=exp(education\_means),exp(education\_HDI))  
educationOdds

## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## mean 1.515840 1.673248 1.751752  
## lower 1.333525 1.504955 1.546969  
## upper 1.732132 1.871543 1.995538

which.max(educationOdds[1,])

## 3.CollegeOrMore   
## 3

which.min(educationOdds[1,])

## 1.NoCollege   
## 1

College or More education level group was the most supportive group. No College education level group was the least supportive group.

*State*

state\_means<-apply(state,2,function(z) mean(z+intercept))  
state\_HDI<-apply(state,2,function(z) hdi(z+intercept))  
stateOdds<-rbind(mean=exp(state\_means),exp(state\_HDI))  
stateOdds

## AK AL AR AZ CA CO CT  
## mean 1.3476856 0.9696427 1.1393769 1.690446 1.985228 1.874901 2.077820  
## lower 0.8827718 0.6552689 0.7871296 1.194617 1.484320 1.325678 1.459179  
## upper 2.0098714 1.4158702 1.6215941 2.352537 2.638478 2.652787 3.003641  
## DC DE FL GA HI IA ID  
## mean 3.604544 2.216578 1.725519 1.2345273 3.129431 2.238279 1.3083928  
## lower 2.248595 1.473026 1.224880 0.9215636 2.197365 1.516694 0.8760142  
## upper 5.771946 3.329536 2.365356 1.6550056 4.438253 3.290229 1.9734203  
## IL IN KS KY LA MA MD  
## mean 2.628237 1.0519490 1.0881749 1.619371 0.9720502 2.474350 1.884324  
## lower 1.900043 0.7188661 0.7387039 1.095936 0.6975811 1.729469 1.377507  
## upper 3.733680 1.5067773 1.5541253 2.358757 1.3587647 3.572645 2.570738  
## ME MI MN MO MS MT NC  
## mean 1.889644 1.978596 2.643456 1.4260513 0.9628397 1.1186730 1.3487614  
## lower 1.238782 1.358984 1.773180 0.9808641 0.6519408 0.7306527 0.9542182  
## upper 2.893011 2.850413 3.910435 2.0746917 1.4208505 1.6911993 1.8963021  
## ND NE NH NJ NM NV NY  
## mean 1.567775 1.751783 1.4694228 2.291579 1.604123 1.3754591 2.165378  
## lower 1.011873 1.178856 0.9331125 1.669673 1.144317 0.9754387 1.548402  
## upper 2.428760 2.603631 2.2804704 3.239776 2.254516 1.9150176 3.005885  
## OH OK OR PA RI SC SD  
## mean 2.181995 1.136575 1.554238 2.120849 2.194671 1.233403 1.660243  
## lower 1.472890 0.805980 1.080299 1.446736 1.432160 0.974229 1.065935  
## upper 3.180823 1.603605 2.262743 3.112789 3.378892 1.571722 2.547570  
## TN TX UT VA VT WA WI  
## mean 1.0479435 1.1781476 1.2767905 1.743564 3.071641 2.487922 1.809932  
## lower 0.7280519 0.8737884 0.8734067 1.284176 1.821589 1.767373 1.216294  
## upper 1.4820015 1.5814508 1.8921464 2.392000 5.118829 3.582052 2.645143  
## WV WY  
## mean 1.0838929 1.1162077  
## lower 0.7094964 0.7155686  
## upper 1.6773978 1.6997658

which.max(stateOdds[1,])

## DC   
## 8

which.min(stateOdds[1,])

## MS   
## 26

District of Columbia was the most supportive state. Mississippi(MS) was the least supportive state.

*Male and Education Interaction*

odds\_male<-rep(intercept+sex[,1],12)  
dim(odds\_male)<-dim(race\_education)  
odds\_male<-odds\_male+race\_education  
odds\_male[2,,]

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White 0.3381952 0.3022762 0.4509062  
## 2.Black 0.3922210 0.3500470 0.3491095  
## 3.Hispanic 0.4277703 0.3401593 0.3234479  
## 4.Other 0.2969835 0.4626876 0.3317064

#race main effects  
race\_matrix<-cbind(race,race,race)  
dim(race\_matrix)<-dim(race\_education)  
dimnames(race\_matrix)<-dimnames(race\_education)  
race\_matrix[2,,]

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White -1.2576649 -1.2576649 -1.2576649  
## 2.Black 2.2629216 2.2629216 2.2629216  
## 3.Hispanic -0.4689053 -0.4689053 -0.4689053  
## 4.Other -0.5363515 -0.5363515 -0.5363515

#education main effects  
education\_matrix<-cbind(education,education,education,education)  
dim(education\_matrix)<-dim(race\_education)[c(1,3,2)]  
education\_matrix<-aperm(education\_matrix,c(1,3,2))  
dimnames(education\_matrix)<-dimnames(race\_education)  
education\_matrix[2,,]

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White -0.1245603 0.007175821 0.1173845  
## 2.Black -0.1245603 0.007175821 0.1173845  
## 3.Hispanic -0.1245603 0.007175821 0.1173845  
## 4.Other -0.1245603 0.007175821 0.1173845

#odds (final)  
odds\_male<-odds\_male+race\_matrix+education\_matrix  
odds\_male[2,,]

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White -1.0440300 -0.94821289 -0.68937421  
## 2.Black 2.5305824 2.62014442 2.72941561  
## 3.Hispanic -0.1656953 -0.12157024 -0.02807290  
## 4.Other -0.3639282 -0.06648802 -0.08726062

(odds\_male\_mean<-apply(odds\_male,c(2,3),mean))

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White -1.0621128 -0.94585601 -0.67234793  
## 2.Black 2.6136840 2.77211123 2.66108165  
## 3.Hispanic -0.1194900 -0.17415100 -0.09260948  
## 4.Other -0.2540419 -0.07887639 -0.13949712

(odds\_male\_hdi<-apply(odds\_male,c(2,3),hdi))

## , , = 1.NoCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower -1.165419 2.282065 -0.3831312 -0.50168211  
## upper -0.957524 2.971747 0.1451652 -0.01521954  
##   
## , , = 2.SomeCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower -1.040187 2.451264 -0.39000004 -0.2794081  
## upper -0.857496 3.090033 0.04273974 0.1241406  
##   
## , , = 3.CollegeOrMore  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower -0.7677193 2.307093 -0.3423469 -0.33622700  
## upper -0.5767772 3.025455 0.1529047 0.05152855

exp(odds\_male\_mean)

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White 0.3457246 0.3883470 0.5105085  
## 2.Black 13.6492428 15.9923620 14.3117611  
## 3.Hispanic 0.8873729 0.8401700 0.9115494  
## 4.Other 0.7756593 0.9241542 0.8697955

exp(odds\_male\_hdi)

## , , = 1.NoCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower 0.3117919 9.796894 0.6817234 0.6055113  
## upper 0.3838421 19.526008 1.1562306 0.9848957  
##   
## , , = 2.SomeCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower 0.3533885 11.60301 0.6770568 0.7562313  
## upper 0.4242230 21.97780 1.0436662 1.1321750  
##   
## , , = 3.CollegeOrMore  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower 0.4640703 10.04518 0.7101018 0.7144609  
## upper 0.5617057 20.60337 1.1652140 1.0528793

Among males, Black population with some college degree was the most supportive group. Among males, White population with no college degree was the least supportive group.

*Female and Education Interaction*

odds\_female<-rep(intercept+sex[,2],12)  
dim(odds\_female)<-dim(race\_education)  
odds\_female<-odds\_female+race\_education+race\_matrix+education\_matrix  
(odds\_female\_mean<-apply(odds\_female,c(2,3),mean))

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White -0.81919311 -0.70293637 -0.4294283  
## 2.Black 2.85660369 3.01503088 2.9040013  
## 3.Hispanic 0.12342969 0.06876865 0.1503102  
## 4.Other -0.01112221 0.16404326 0.1034225

(odds\_female\_hdi<-apply(odds\_female,c(2,3),hdi))

## , , = 1.NoCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower -0.9179918 2.526634 -0.1289625 -0.2446230  
## upper -0.7184181 3.200256 0.3764959 0.2291783  
##   
## , , = 2.SomeCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower -0.7917329 2.712904 -0.1339767 -0.02970443  
## upper -0.6142555 3.325144 0.2762705 0.35551907  
##   
## , , = 3.CollegeOrMore  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower -0.5191612 2.550978 -0.0934398 -0.08688782  
## upper -0.3341005 3.252281 0.3890330 0.29729359

exp(odds\_female\_mean)

##   
## 1.NoCollege 2.SomeCollege 3.CollegeOrMore  
## 1.White 0.4407872 0.4951293 0.6508811  
## 2.Black 17.4023228 20.3897205 18.2470112  
## 3.Hispanic 1.1313705 1.0711884 1.1621947  
## 4.Other 0.9889394 1.1782653 1.1089599

exp(odds\_female\_hdi)

## , , = 1.NoCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower 0.3993202 12.51132 0.8790069 0.7829997  
## upper 0.4875229 24.53882 1.4571695 1.2575662  
##   
## , , = 2.SomeCollege  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower 0.4530590 15.07299 0.8746105 0.9707324  
## upper 0.5410436 27.80300 1.3182044 1.4269211  
##   
## , , = 3.CollegeOrMore  
##   
##   
## 1.White 2.Black 3.Hispanic 4.Other  
## lower 0.5950195 12.81963 0.9107928 0.9167799  
## upper 0.7159818 25.84923 1.4755532 1.3462105

Among females, black population with some college degree was the most supportive group. Among females, white population with no college degree was the least supportive group.