542Regression2

Group 4

This is a document for our second hypothesis

```
# import our merged data by using the raw link and named it merge

link="https://raw.githubusercontent.com/Public-Policy-
COVID/students_merge/main/Merged_data.csv"
fromPy=read.csv(link, header = T)
row.names(fromPy)=NULL

# verifying data structure

str(fromPy,width = 50,strict.width='cut')

## 'data.frame':    133 obs. of  19 variables:
##  $ Number_of_beds       : num  3667 0 52 553 25 ..
##  $ Number_of_hospitals  : num  22 0 1 6 1 1 10 1..
##  $ Location             : chr  "Alameda_CA" "Al"..
##  $ Urban_Rural_Code     : chr  "Large central m"..
##  $ Deaths_COVID         : int  573 0 31 101 12 1..
##  $ Deaths_total         : int  10908 0 415 2313 ..
##  $ never                : num  0.019 0.025 0.045..
##  $ rarely               : num  0.008 0.085 0.013..
##  $ sometimes            : num  0.055 0.088 0.099..
##  $ frequently           : num  0.123 0.19 0.188 ..
##  $ always               : num  0.795 0.612 0.655..
##  $ mask_score           : num  3.67 3.28 3.4 3.3..
##  $ total_population     : num  1671329 1129 3975..
##  $ white_total_pct      : num  49.3 67.9 89.7 85..
##  $ black_total_pct      : num  11.03 0.35 2.68 1..
##  $ aian_total_pct       : num  1.06 25.69 2.33 2..
##  $ asian_total_pct      : num  32.33 1.59 1.67 5..
##  $ nhopi_total_pct      : num  0.94 0 0.29 0.29 ..
##  $ multiracial_total_pct: num  5.35 4.43 3.38 4...

# covert integer variables to numeric

fromPy$Deaths_COVID <- as.numeric(fromPy$Deaths_COVID)
fromPy$Deaths_total <- as.numeric(fromPy$Deaths_total)
str(fromPy,width = 50,strict.width='cut')

## 'data.frame':    133 obs. of  19 variables:
##  $ Number_of_beds       : num  3667 0 52 553 25 ..
##  $ Number_of_hospitals  : num  22 0 1 6 1 1 10 1..
##  $ Location             : chr  "Alameda_CA" "Al"..
```

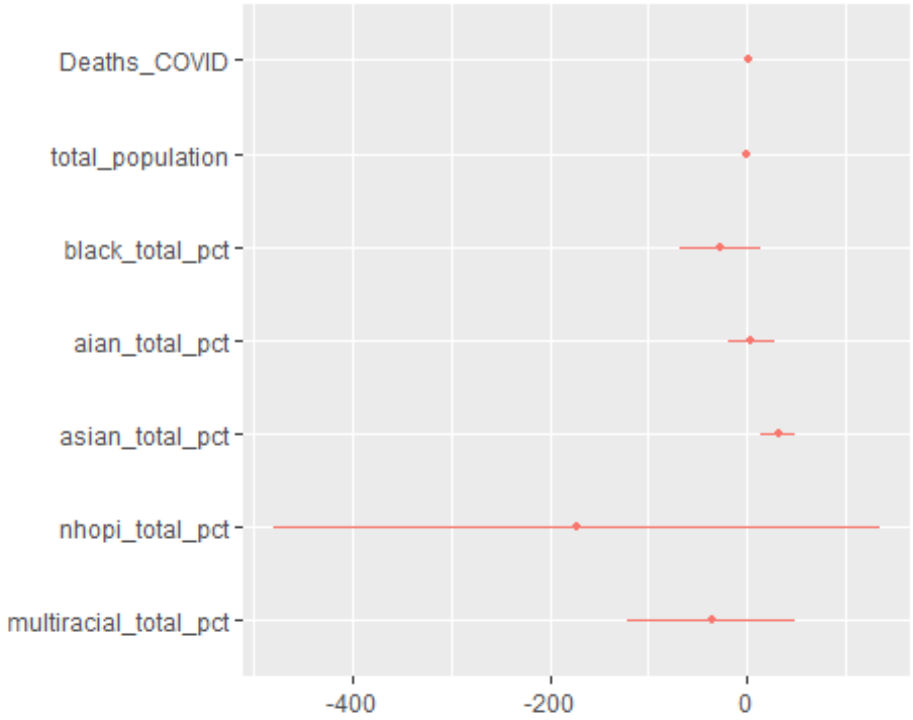
```
##  $ Urban_Rural_Code     : chr  "Large central m"..
##  $ Deaths_COVID         : num  573 0 31 101 12 1..
##  $ Deaths_total         : num  10908 0 415 2313 ..
##  $ never                : num  0.019 0.025 0.045..
##  $ rarely               : num  0.008 0.085 0.013..
##  $ sometimes            : num  0.055 0.088 0.099..
##  $ frequently           : num  0.123 0.19 0.188 ..
##  $ always               : num  0.795 0.612 0.655..
##  $ mask_score           : num  3.67 3.28 3.4 3.3..
##  $ total_population     : num  1671329 1129 3975..
##  $ white_total_pct      : num  49.3 67.9 89.7 85..
##  $ black_total_pct      : num  11.03 0.35 2.68 1..
##  $ aian_total_pct       : num  1.06 25.69 2.33 2..
##  $ asian_total_pct      : num  32.33 1.59 1.67 5..
##  $ nhopi_total_pct      : num  0.94 0 0.29 0.29 ..
##  $ multiracial_total_pct: num  5.35 4.43 3.38 4...

# summary data
summary(fromPy)

##  Number_of_beds    Number_of_hospitals   Location         Urban_Rural_Code
##  Min.   :    0.0   Min.   :  0         Length:133         Length:133
##  1st Qu.:   25.0   1st Qu.:  1         Class :character   Class :character
##  Median :  131.0   Median :  2         Mode  :character   Mode  :character
##  Mean   :  885.4   Mean   :  5
##  3rd Qu.:  553.0   3rd Qu.:  4
##  Max.   :26672.0   Max.   :112
##   Deaths_COVID   Deaths_total       never             rarely
##  Min.   :   0   Min.   :    0   Min.   :0.00100   Min.   :0.00000
##  1st Qu.:   0   1st Qu.:    0   1st Qu.:0.01600   1st Qu.:0.01400
##  Median :  22   Median :  637   Median :0.02600   Median :0.02800
##  Mean   : 206   Mean   : 2896   Mean   :0.03513   Mean   :0.03806
##  3rd Qu.: 128   3rd Qu.: 2537   3rd Qu.:0.04500   3rd Qu.:0.05600
##  Max.   :8034   Max.   :75463   Max.   :0.14000   Max.   :0.20600
##    sometimes         frequently         always         mask_score
##  Min.   :0.00400   Min.   :0.0580   Min.   :0.3050   Min.   :2.470
##  1st Qu.:0.04800   1st Qu.:0.1410   1st Qu.:0.6160   1st Qu.:3.301
##  Median :0.06900   Median :0.1680   Median :0.6810   Median :3.464
##  Mean   :0.07167   Mean   :0.1736   Mean   :0.6814   Mean   :3.428
##  3rd Qu.:0.09100   3rd Qu.:0.2040   3rd Qu.:0.7540   3rd Qu.:3.591
##  Max.   :0.21300   Max.   :0.3320   Max.   :0.8890   Max.   :3.822
##  total_population   white_total_pct black_total_pct  aian_total_pct
##  Min.   :    1129   Min.   :49.28   Min.   : 0.000   Min.   : 0.590
##  1st Qu.:   24658   1st Qu.:82.16   1st Qu.: 0.770   1st Qu.: 1.430
##  Median :   79481   Median :88.64   Median : 1.260   Median : 2.010
##  Mean   :  385537   Mean   :85.50   Mean   : 2.318   Mean   : 2.985
##  3rd Qu.:  283111   3rd Qu.:91.84   3rd Qu.: 2.620   3rd Qu.: 3.070
##  Max.   :10039107   Max.   :96.13   Max.   :14.770   Max.   :25.690
##  asian_total_pct  nhopi_total_pct  multiracial_total_pct
##  Min.   : 0.500   Min.   :0.0000   Min.   :1.200
```

```
##  1st Qu.: 1.210   1st Qu.:0.2100   1st Qu.:3.160
##  Median : 1.870   Median :0.2800   Median :3.720
##  Mean   : 4.961   Mean   :0.3838   Mean   :3.856
##  3rd Qu.: 5.840   3rd Qu.:0.4500   3rd Qu.:4.440
##  Max.   :39.020   Max.   :1.7100   Max.   :7.800
```

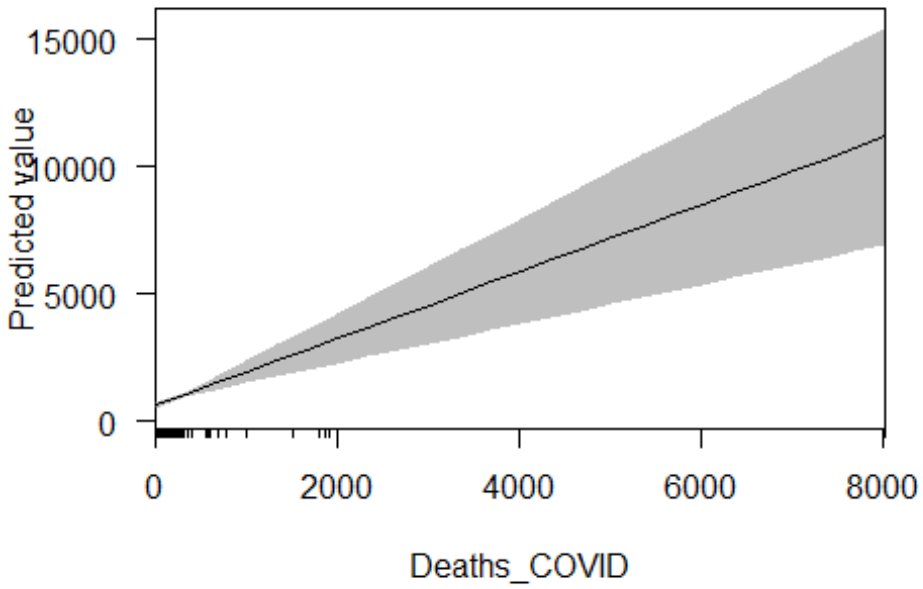
```
# state the hypothesis#, and name it hypo2
# hypo2 = hypothesis 2: state with higher Deaths_COVID number has more
Number_of_beds in hospitals

hypo2=formula(Number_of_beds~
Deaths_COVID+total_population+black_total_pct+aian_total_pct+asian_total_pct+
nhopi_total_pct+multiracial_total_pct)

# compute regression models

gauss2=glm(hypo2,
           data = fromPy,
           family = 'gaussian')

# check the result of our regression

summary(gauss2)
```

```
##
## Call:
## glm(formula = hypo2, family = "gaussian", data = fromPy)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2149.50    -94.91     -2.91     70.70   2337.16
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.155e+02  1.400e+02   0.825 0.411063
## Deaths_COVID           1.313e+00  2.778e-01   4.727 6.04e-06 ***
## total_population       1.551e-03  2.199e-04   7.054 1.05e-10 ***
## black_total_pct       -2.731e+01  2.062e+01  -1.324 0.187765
## aian_total_pct         4.344e+00  1.191e+01   0.365 0.716016
## asian_total_pct        3.176e+01  8.762e+00   3.625 0.000419 ***
## nhopi_total_pct       -1.723e+02  1.567e+02  -1.100 0.273487
## multiracial_total_pct -3.622e+01  4.310e+01  -0.841 0.402221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 169887.2)
##
##     Null deviance: 930593834  on 132  degrees of freedom
## Residual deviance:  21235896  on 125  degrees of freedom
## AIC: 1988.9
```

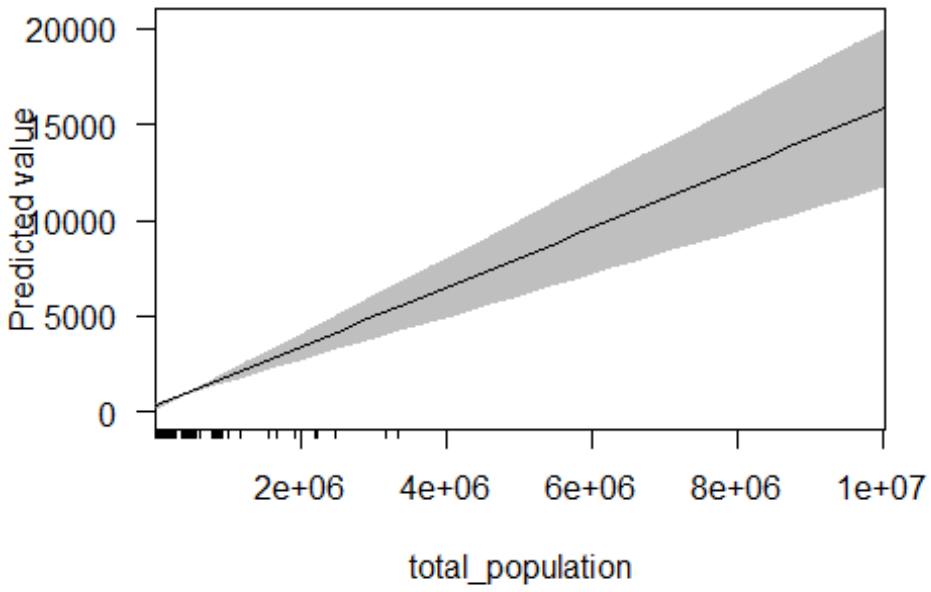
```
##
## Number of Fisher Scoring iterations: 2

# get R square of this regression

library(rsq)
rsq(gauss2, adj = T)

## [1] 0.9759024

# to summary plots
library(dotwhisker)

## Loading required package: ggplot2

## Registered S3 method overwritten by 'broom.mixed':
##   method      from
##   tidy.gamlss broom

dwplot(gauss2,by_2sd = F)
```



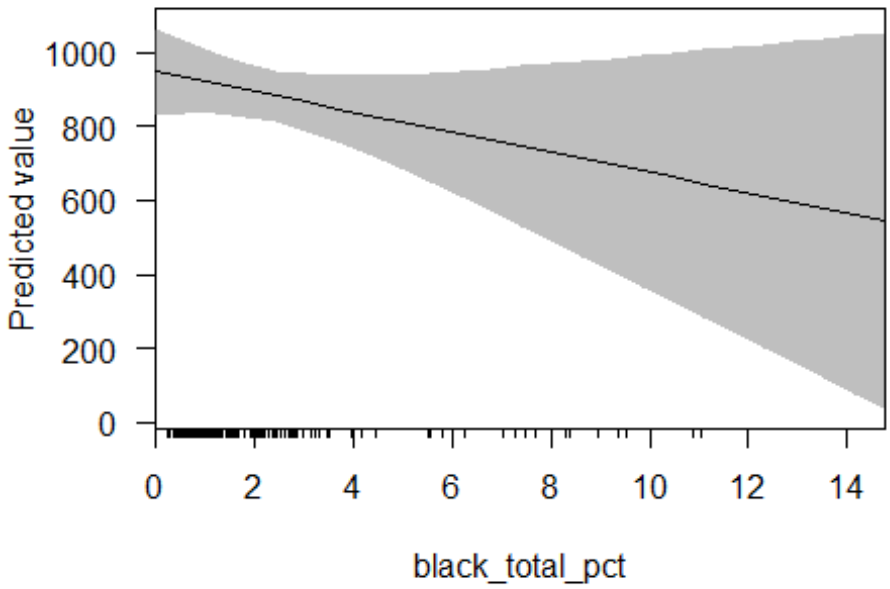

```
# check the margins

library(margins)

cplot(gauss2,'Deaths_COVID')
```



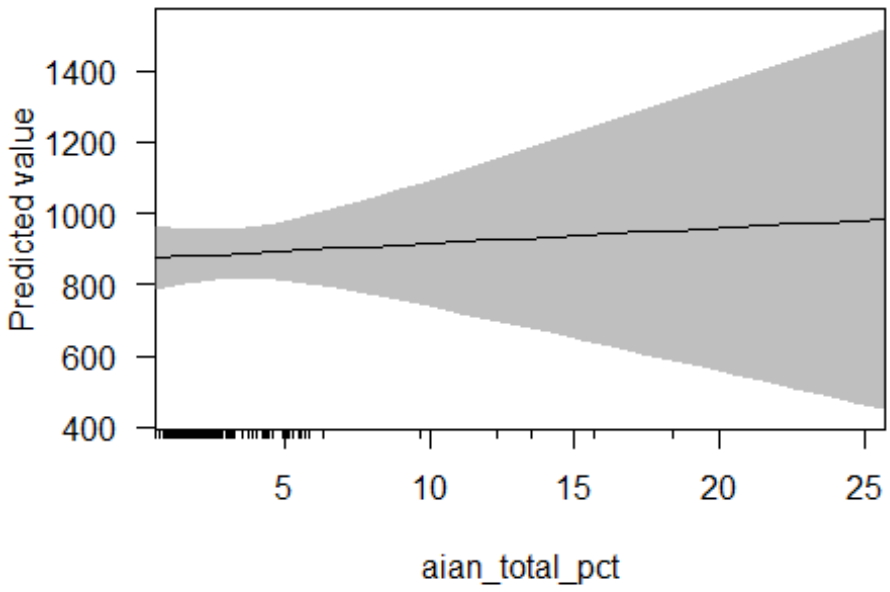

```
cplot(gauss2,'total_population')
```



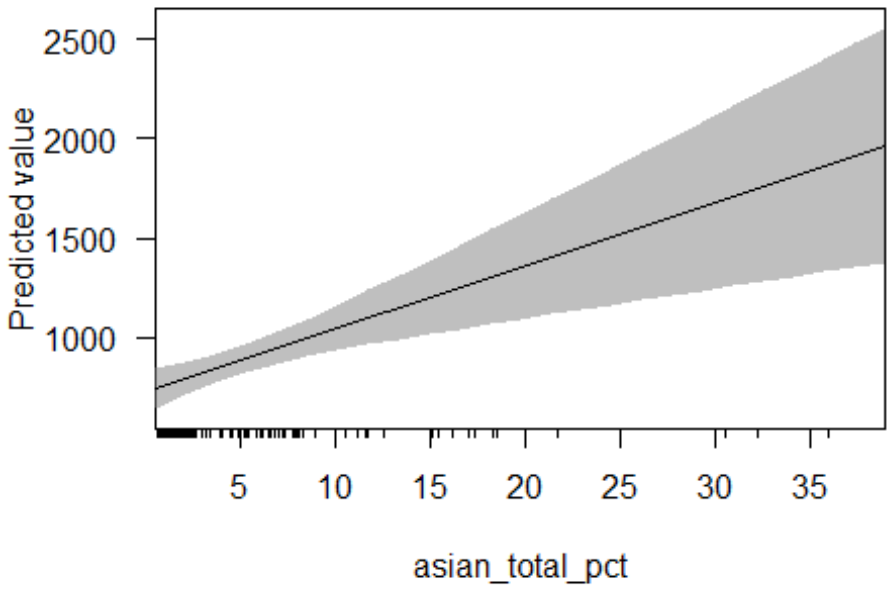

```
cplot(gauss2,'black_total_pct')
```



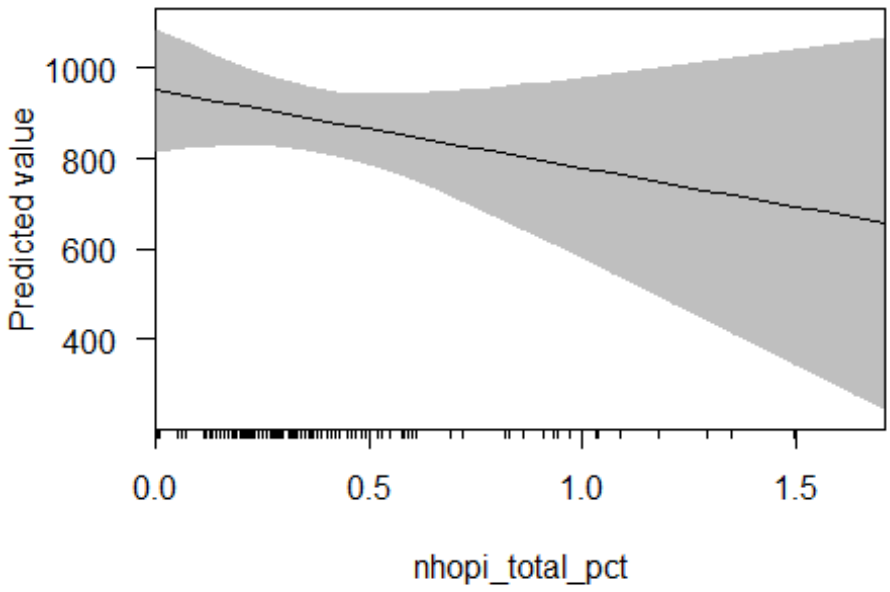

```
cplot(gauss2,'aian_total_pct')
```



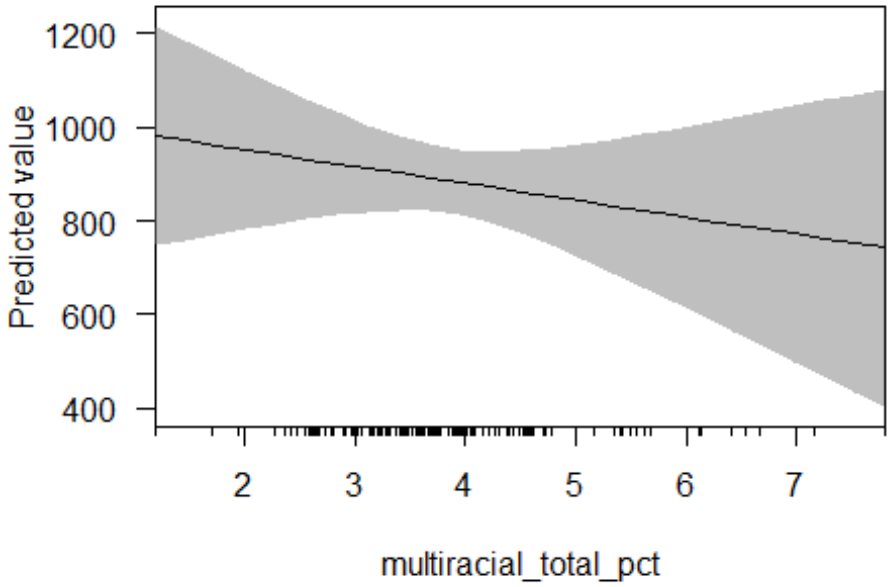

```
cplot(gauss2,'asian_total_pct')
```




```
cplot(gauss2,'nhopi_total_pct')
```




```
cplot(gauss2,'multiracial_total_pct')
```




```
# plot the interaction

persp(gauss2)
```