```
link='https://github.com/ComputationalThinkingGroup5/Merge/raw/master/MergedData.csv'
# a RDS file from the web needs:
mergedFile=read.csv(link)
```

```
str(mergedFile,width = 50,strict.width='cut')
```

```
## 'data.frame':    162 obs. of  4 variables:
##  $ country            : chr  "Cambodia" "Niger""..
##  $ percentunemployment: num  0.3 0.3 0.7 0.78 0...
##  $ pct_GDP_exp        : num  2.2 3.5 2.9 4.8 4.8..
##  $ percentbirthrate   : num  1.34 3.65 1.46 0.75..
```

```
# hypothesis 1: percentunemployment increases as percentbirthrate advances:
hypo1=formula(percentunemployment~ percentbirthrate)
# hypothesis 2: percentunemployment increases as percentbirthrate and pct_GDP_exp advance:
hypo2=formula(percentunemployment~ percentbirthrate * pct_GDP_exp)
```

```
#
# results
gauss1=glm(hypo1,
           data = mergedFile,
           family = 'gaussian')

gauss2=glm(hypo2,
           data = mergedFile,
           family = 'gaussian')
```

#3.See results: · First Hypothesis

```
summary(gauss1)
```

```
##
## Call:
## glm(formula = hypo1, family = "gaussian", data = mergedFile)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -15.166   -5.531   -2.877    1.528   63.857
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.5423     1.0753   7.014 6.16e-11 ***
## percentbirthrate   2.1710     0.7683   2.826  0.00532 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 100.6676)
##
##     Null deviance: 16911  on 161  degrees of freedom
## Residual deviance: 16107  on 160  degrees of freedom
## AIC: 1210.8
##
## Number of Fisher Scoring iterations: 2
```

· Second Hypothesis

```
summary(gauss2)
```

```
##
## Call:
## glm(formula = hypo2, family = "gaussian", data = mergedFile)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -13.654   -5.713   -2.580    1.972   60.820
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    6.1840     2.8351   2.181   0.0306 *
## percentbirthrate              -0.7806     2.0912  -0.373   0.7095
## pct_GDP_exp                    0.2716     0.5519   0.492   0.6233
## percentbirthrate:pct_GDP_exp   0.7568     0.4563   1.658   0.0992 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 98.30933)
##
##     Null deviance: 16911  on 161  degrees of freedom
## Residual deviance: 15533  on 158  degrees of freedom
## AIC: 1209
##
## Number of Fisher Scoring iterations: 2
```

#4.Search for better model:

```
anova(gauss1,gauss2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: percentunemployment ~ percentbirthrate
## Model 2: percentunemployment ~ percentbirthrate * pct_GDP_exp
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       160      16107
## 2       158      15533  2   573.94  0.05398 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model for the first hypothesis is chosen. You can get the RSquared if needed:

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

```
## [1] 0.06403427
```

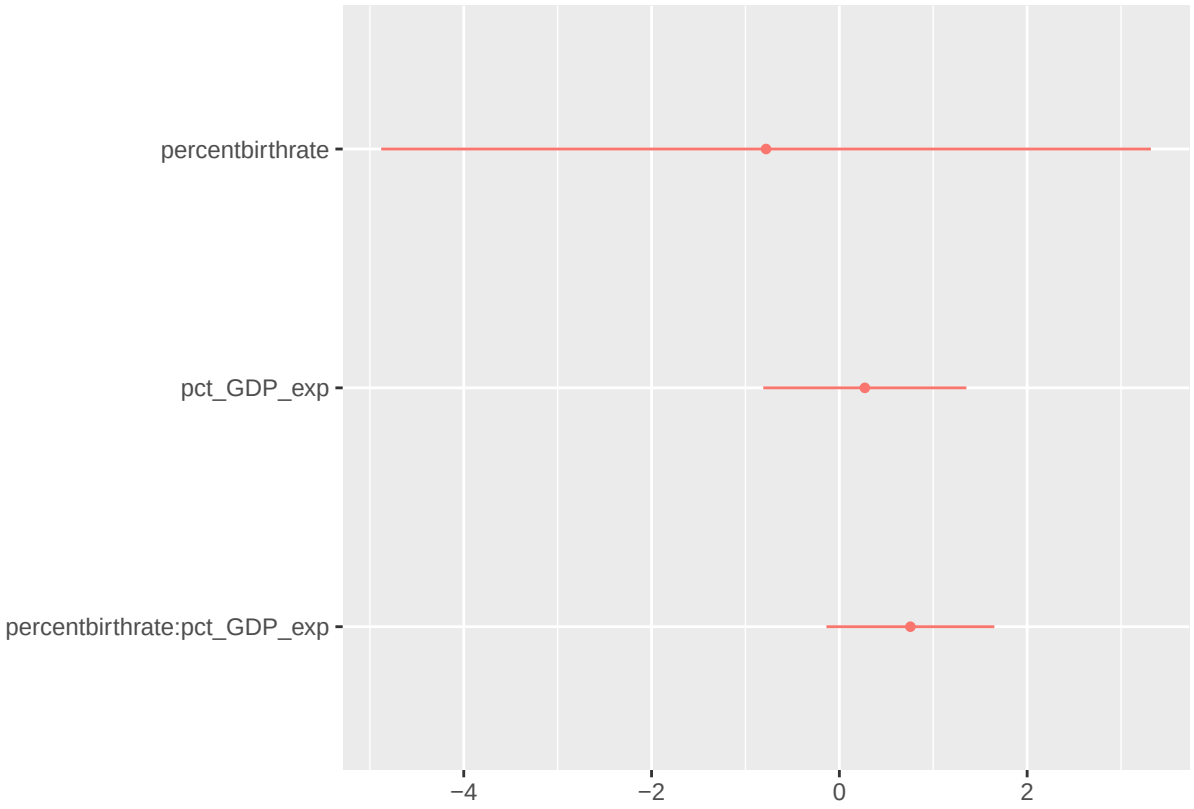
Finally, some nice summary plots: You can see the coefficient estimates like this:

```
library(dotwhisker)
```

```
## Loading required package: ggplot2
```

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

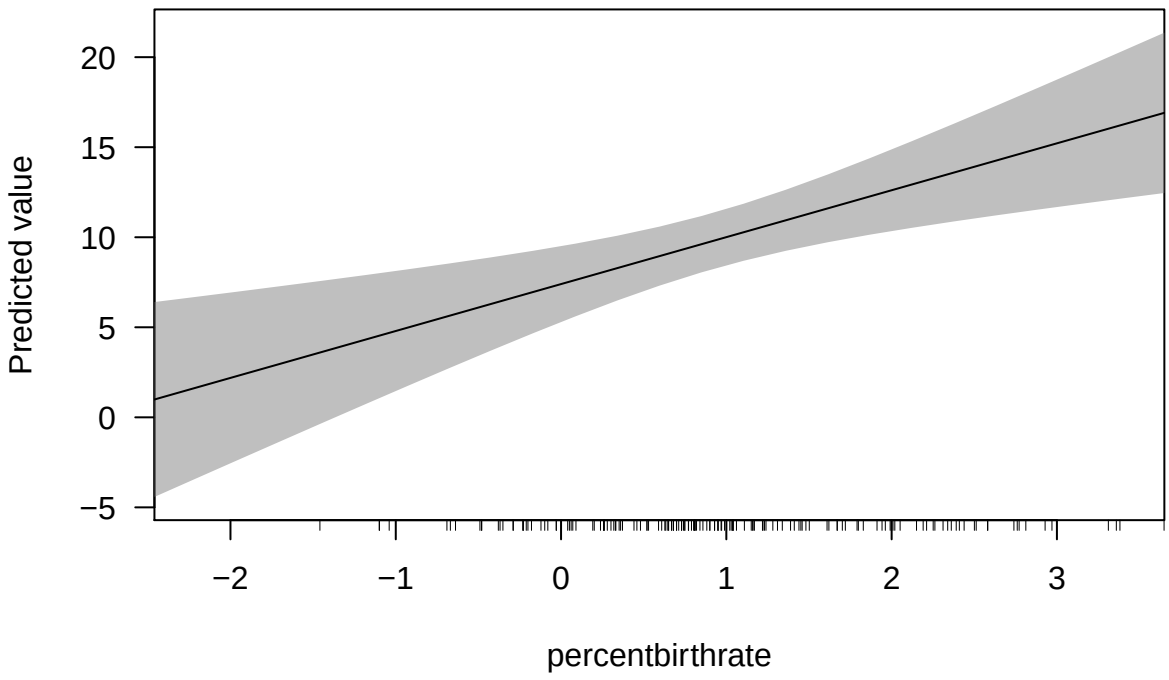
```
dwplot(gauss2,by_2sd = F)
```



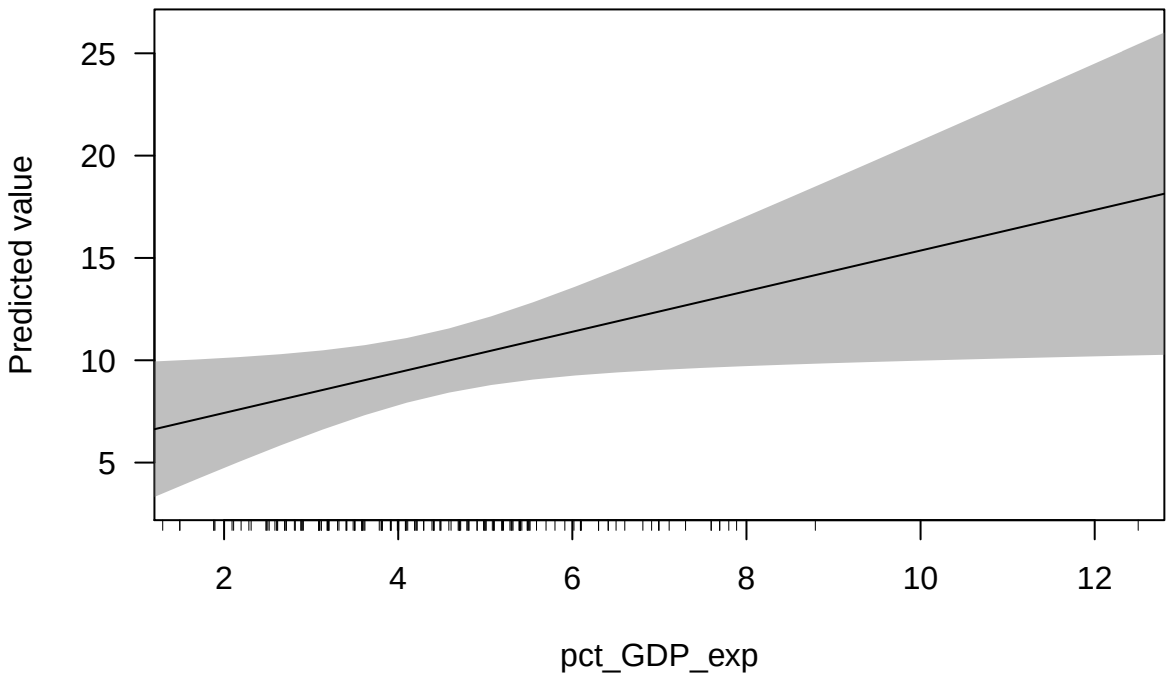

You can also use margins library:

```
library(margins)

cplot(gauss2,'percentbirthrate')
```

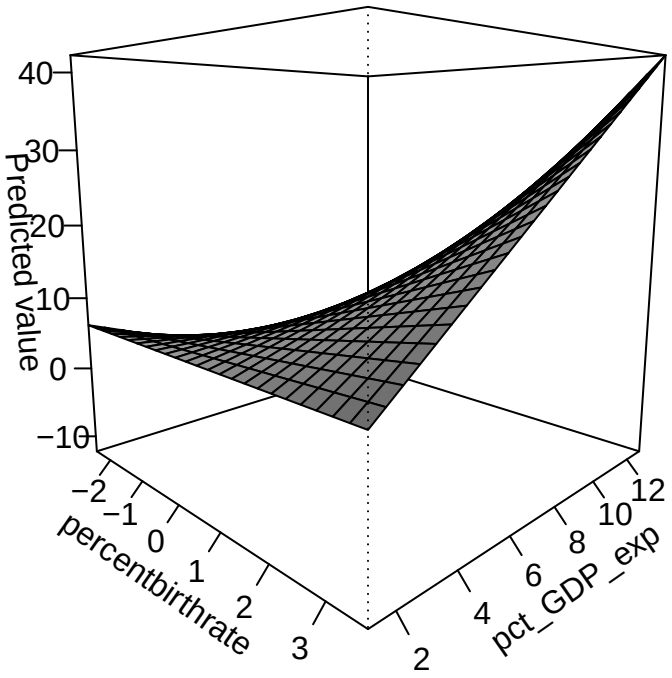



```
cplot(gauss2,'pct_GDP_exp')
```




And the interaction:

```
persp(gauss2)
```




Binary outcome In this situation you have a binary dependent variable, which we do not currently have:

```
mergedFile$percentunemploymentdico=ifelse(mergedFile$percentunemployment>median(mergedFile$percentunempl
                                          na.rm = T),
                       1,0)
```

Now we have it.

1.State hypothesis: Let's use the same ones:

```
hypoDico1=formula(percentunemploymentdico~ percentunemployment)
hypoDico2=formula(percentunemploymentdico~ percentunemployment * pct_GDP_exp)
```

2.Reformat

```
mergedFile$mergedFiledico=factor(mergedFile$percentunemploymentdico)
```

3.Compute regression models:

```
Logi1=glm(hypoDico1,data = mergedFile,
          family = "binomial")
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
Logi2=glm(hypoDico2,data = mergedFile,
          family = "binomial")
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

4.See results: · First Hypothesis:

```
summary(Logi1)
```

```
##
## Call:
## glm(formula = hypoDico1, family = "binomial", data = mergedFile)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1.830e-03  -2.000e-08   0.000e+00   2.000e-08   2.094e-03
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -3436.3    81284.3  -0.042    0.966
## percentunemployment    526.6    12450.9   0.042    0.966
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.2458e+02  on 161  degrees of freedom
## Residual deviance: 7.7324e-06  on 160  degrees of freedom
## AIC: 4
##
## Number of Fisher Scoring iterations: 25
```

3.Second Hypothesis:

```
summary(Logi2)
```

```
##
## Call:
## glm(formula = hypoDico2, family = "binomial", data = mergedFile)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -6.566e-04  -2.000e-08   0.000e+00   2.000e-08   6.917e-04
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -1934.37   93064.31  -0.021    0.983
## percentunemployment               290.89   13977.42   0.021    0.983
## pct_GDP_exp                       143.55    7241.72   0.020    0.984
## percentunemployment:pct_GDP_exp   -20.39    1060.78  -0.019    0.985
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.2458e+02  on 161  degrees of freedom
## Residual deviance: 1.3617e-06  on 158  degrees of freedom
## AIC: 8
##
## Number of Fisher Scoring iterations: 25
```

5.Search for better model:

```
lmtest::lrtest(Logi1,Logi2)
```

```
## Likelihood ratio test
##
## Model 1: percentunemploymentdico ~ percentunemployment
## Model 2: percentunemploymentdico ~ percentunemployment * pct_GDP_exp
##   #Df      LogLik Df Chisq Pr(>Chisq)
## 1   2 -3.8662e-06
## 2   4 -6.8090e-07  2     0          1
```

Model for the second hypothesis is chosen.

6.Logistics regression coefficients do not offer marginal effects on how much eah variable affects the probability of the '1' outcome. We can get it using margins library:

```
library(margins)
(marginsINFO = margins(Logi2))
```

```
## Average marginal effects

## glm(formula = hypoDico2, family = "binomial", data = mergedFile)

##  percentunemployment pct_GDP_exp
##            1.691e-05   4.696e-08
```

In this case, I could produce a plot for both coefficients. Take a look at the summary of

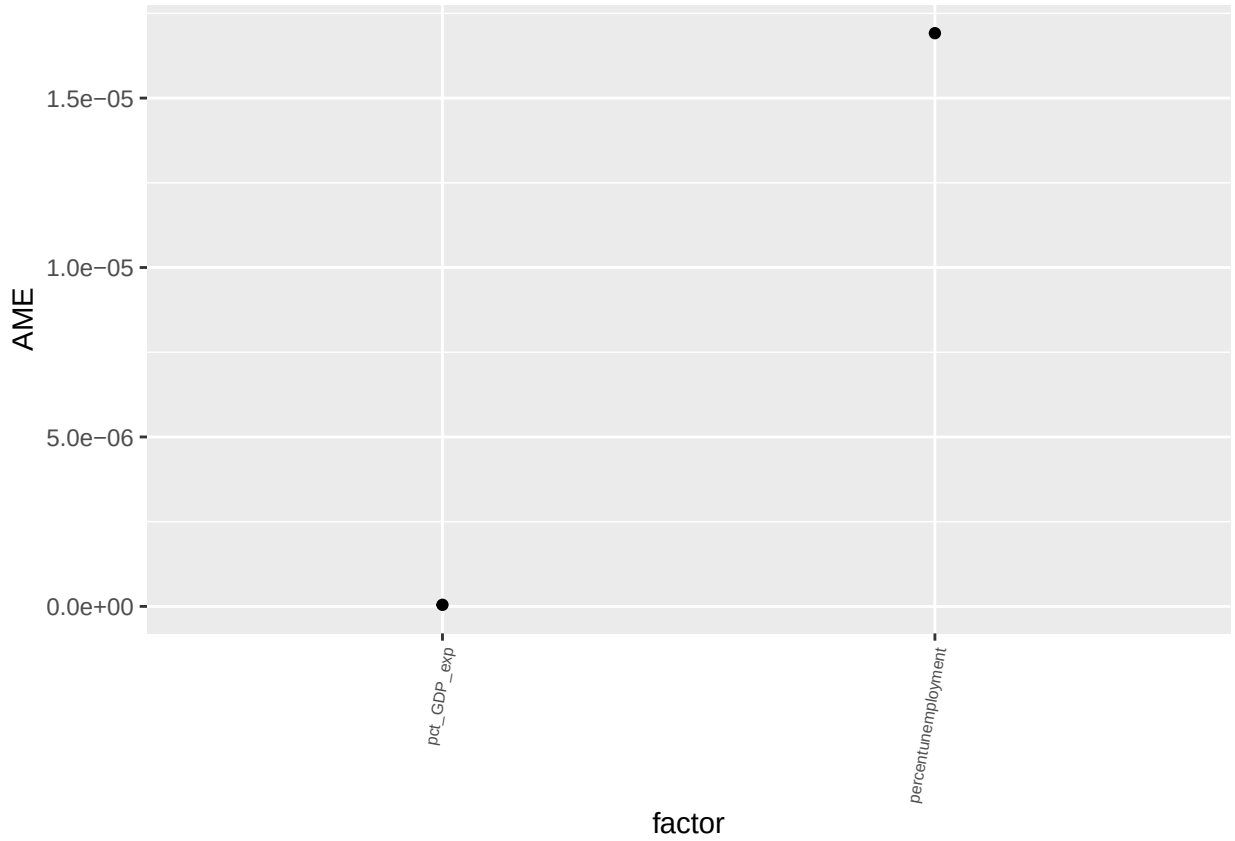
```
(marginsSUMM=summary(marginsINFO))
```

```
##               factor    AME     SE      z      p   lower  upper
##          pct_GDP_exp 0.0000 0.0000 0.0014 0.9989 -0.0001 0.0001
##  percentunemployment 0.0000 0.0089 0.0019 0.9985 -0.0175 0.0175
```

I can use that information like this:

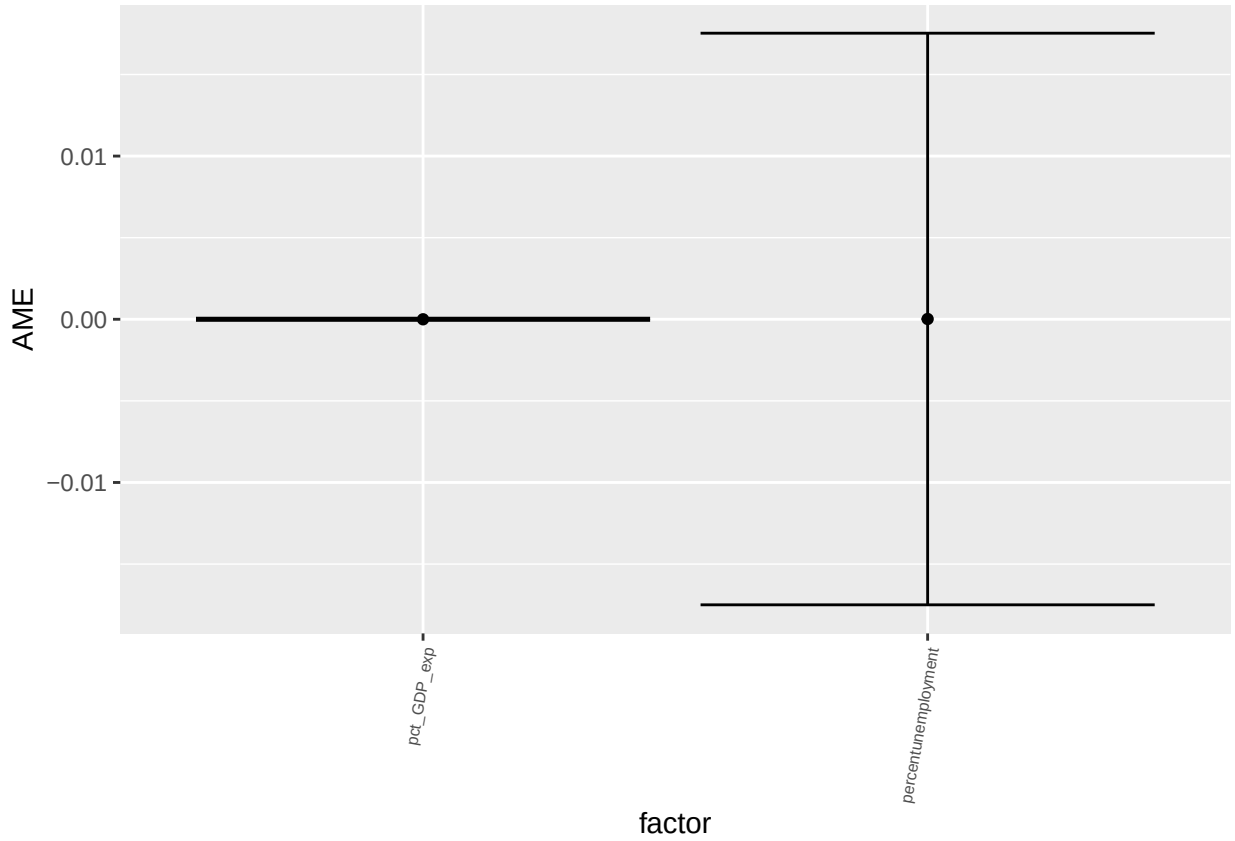
```
base= ggplot(marginsSUMM,
             aes(x=factor, y=AME))
base= base + geom_point()

plotMargins = base + theme(axis.text.x = element_text(angle = 80,size = 6,hjust = 1))
plotMargins
```



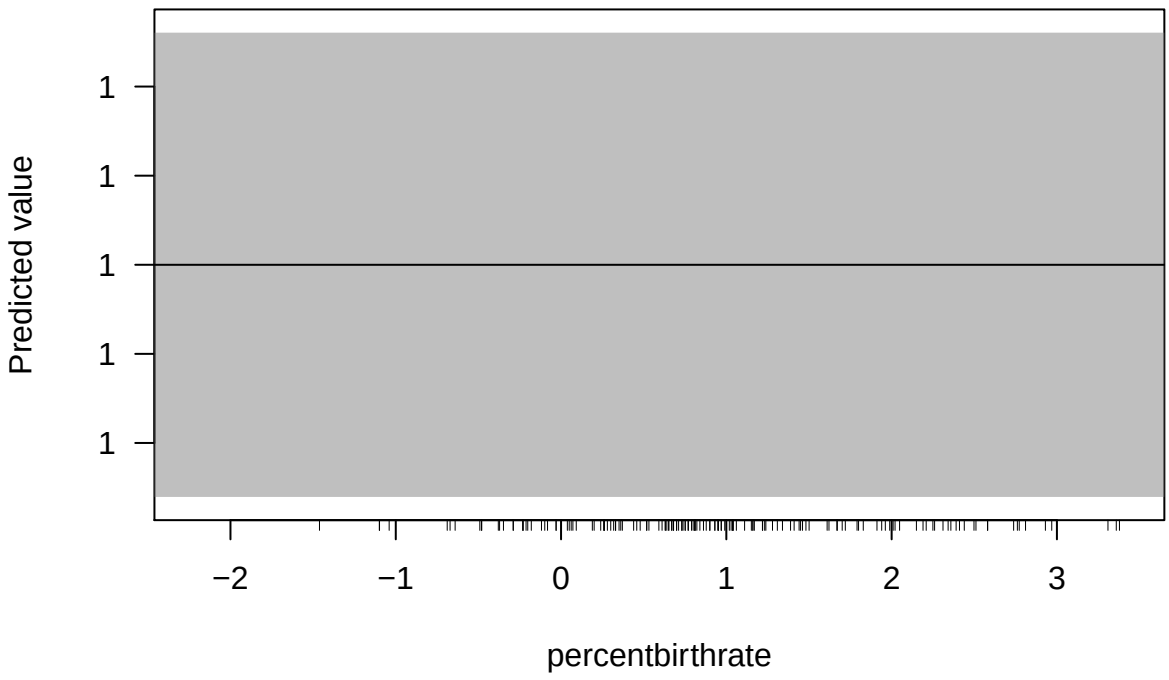

We can add the the confidence intervals:

```
plotMargins +  geom_errorbar(aes(ymin=lower,
                                 ymax=upper))
```

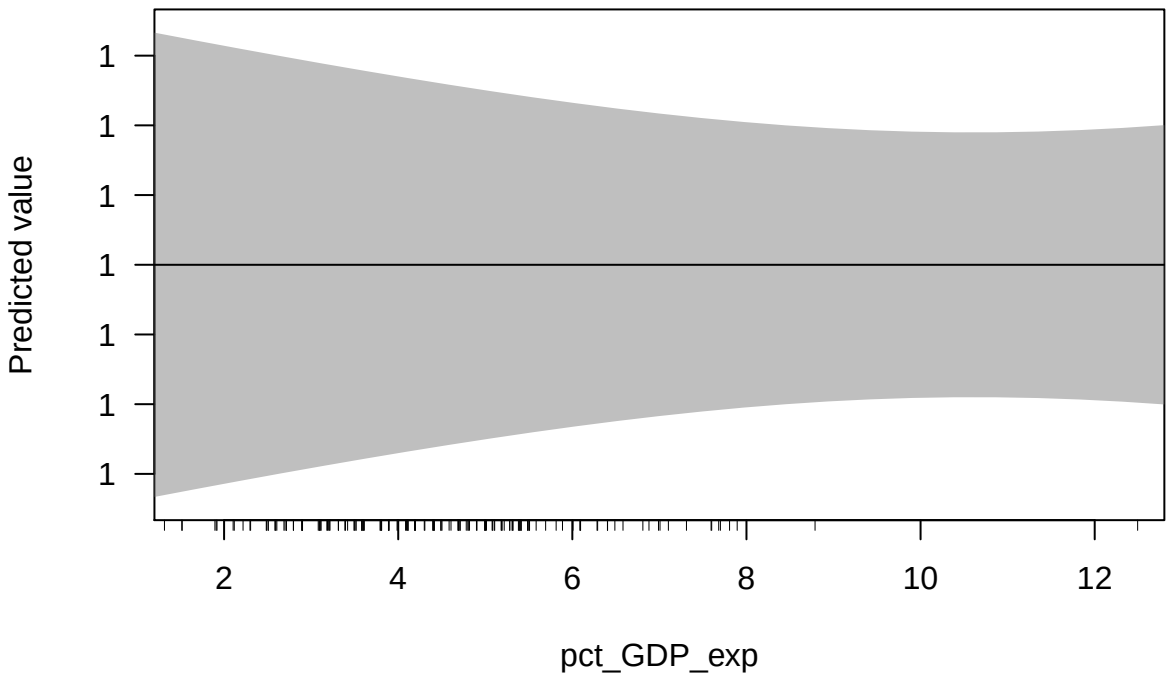



It is a good idea to use the indivual plots per variable:

```
cplot(Logi2, "percentbirthrate")
```

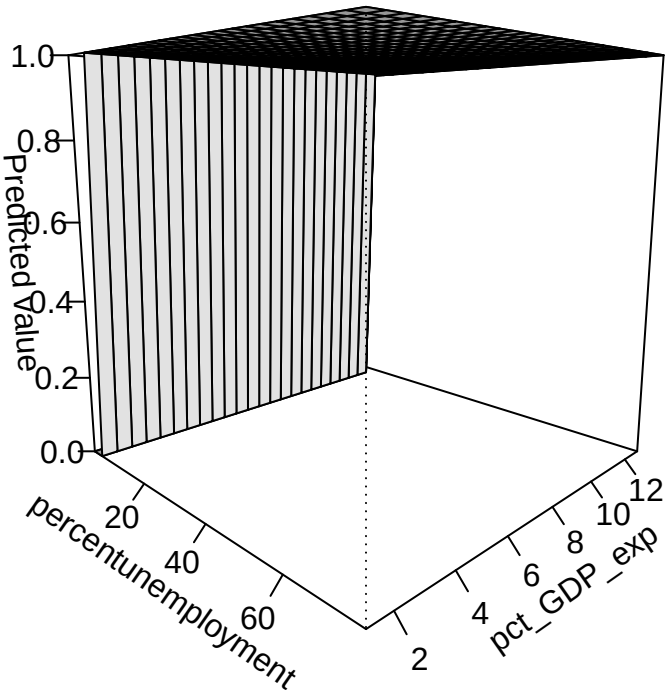



```
cplot(Logi2, "pct_GDP_exp")
```




And for the interaction:

```
persp(Logi2)
```




```
tinytex::install_tinytex()
```