Data Project

Iyanu Lamina

2023-01

```
View(pums_variables)
```

```
ND_pums<- get_pums(
variables = c("POVPIP", "SEX", "ESR","NP", "MIG","HINCP","AGEP"),
state = "ND",
survey = "acs1",
variables_filter = list(
    AGEP = 18:99,
    POVPIP = 1:501
  ),
year= 2021,
recode = TRUE

)%>%mutate(MIG= fct_recode(MIG,"2"="3"))  %>%
 mutate(poverty = factor(if_else(POVPIP >200,"above poverty","below poverty")))
```

```
## Getting data from the 2021 1-year ACS Public Use Microdata Sample
```

```
## Warning: • You have not set a Census API key. Users without a key are limited to 500
## queries per day and may experience performance limitations.
## ℹ For best results, get a Census API key at
## http://api.census.gov/data/key_signup.html and then supply the key to the
## `census_api_key()` function to use it throughout your tidycensus session.
## This warning is displayed once per session.
```

```
##
Downloading: 7.6 kB
Downloading: 7.6 kB
Downloading: 7.6 kB
Downloading: 7.6 kB
Downloading: 7.6 kB
Downloading: 7.6 kB
Downloading: 16 kB
Downloading: 16 kB
Downloading: 24 kB
Downloading: 24 kB
Downloading: 32 kB
Downloading: 32 kB
Downloading: 32 kB
Downloading: 32 kB
Downloading: 48 kB
Downloading: 48 kB
Downloading: 57 kB
Downloading: 57 kB
Downloading: 57 kB
Downloading: 57 kB
Downloading: 65 kB
Downloading: 65 kB
Downloading: 65 kB
Downloading: 65 kB
Downloading: 73 kB
Downloading: 73 kB
Downloading: 73 kB
Downloading: 73 kB
Downloading: 81 kB
Downloading: 81 kB
Downloading: 89 kB
Downloading: 89 kB
Downloading: 89 kB
Downloading: 89 kB
Downloading: 98 kB
Downloading: 98 kB
Downloading: 110 kB
Downloading: 110 kB
Downloading: 110 kB
Downloading: 110 kB
Downloading: 110 kB
Downloading: 110 kB
Downloading: 120 kB
Downloading: 120 kB
Downloading: 120 kB
Downloading: 120 kB
Downloading: 130 kB
Downloading: 130 kB
Downloading: 140 kB
Downloading: 140 kB
Downloading: 150 kB
```

```
Downloading: 150 kB
Downloading: 170 kB
Downloading: 170 kB
Downloading: 170 kB
Downloading: 170 kB
Downloading: 190 kB
Downloading: 190 kB
Downloading: 190 kB
Downloading: 190 kB
Downloading: 190 kB
Downloading: 190 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 200 kB
Downloading: 220 kB
Downloading: 220 kB
Downloading: 220 kB
Downloading: 220 kB
Downloading: 240 kB
Downloading: 240 kB
Downloading: 240 kB
Downloading: 240 kB
Downloading: 250 kB
Downloading: 250 kB
Downloading: 270 kB
Downloading: 270 kB
Downloading: 270 kB
Downloading: 270 kB
Downloading: 270 kB
Downloading: 270 kB
Downloading: 290 kB
Downloading: 290 kB
Downloading: 300 kB
Downloading: 300 kB
Downloading: 300 kB
Downloading: 300 kB
Downloading: 300 kB
Downloading: 300 kB
Downloading: 320 kB
Downloading: 320 kB
Downloading: 350 kB
Downloading: 350 kB
Downloading: 360 kB
Downloading: 360 kB
Downloading: 360 kB
Downloading: 360 kB
Downloading: 360 kB
```

```
Downloading: 360 kB
Downloading: 360 kB
Downloading: 360 kB
Downloading: 380 kB
Downloading: 380 kB
Downloading: 400 kB
Downloading: 400 kB
Downloading: 400 kB
Downloading: 400 kB
Downloading: 400 kB
Downloading: 400 kB
Downloading: 410 kB
Downloading: 410 kB
Downloading: 410 kB
Downloading: 410 kB
Downloading: 420 kB
Downloading: 420 kB
Downloading: 420 kB
Downloading: 420 kB
```

```
ND_pums
```

```
## # A tibble: 5,692 × 17
##    SERIALNO      SPORDER  WGTP PWGTP    NP  HINCP  AGEP POVPIP ST    MIG   SEX
##    <chr>           <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>  <dbl> <chr> <fct> <chr>
##  1 2021HU0423857       1   142   142     4 100300    38    176 38    1     2
##  2 2021HU0423857       2   142   187     4 100300    56    462 38    1     1
##  3 2021HU0424260       1   167   167     1  78300    78    501 38    1     2
##  4 2021HU0424681       1   368   368     6 102200    37    285 38    1     2
##  5 2021HU0424681       2   368   344     6 102200    31    285 38    1     1
##  6 2021HU0424869       1    77    77     1  96000    52    501 38    1     2
##  7 2021HU0424947       1    69    70     4 110000    43    409 38    1     1
##  8 2021HU0425266       1    45    45     2  60000    28    291 38    1     2
##  9 2021HU0425266       2    45   154     2  60000    29    145 38    1     1
## 10 2021HU0425305       1   237   237     2  96000    29    442 38    1     1
## # ℹ 5,682 more rows
## # ℹ 6 more variables: ESR <chr>, ST_label <ord>, MIG_label <ord>,
## #   SEX_label <ord>, ESR_label <ord>, poverty <fct>
```

```
summary(ND_pums)
```

```
##    SERIALNO            SPORDER            WGTP             PWGTP
##  Length:5692        Min.   : 1.000   Min.   :   0.00   Min.   :   4.00
##  Class :character   1st Qu.: 1.000   1st Qu.:  26.00   1st Qu.:  26.00
##  Mode  :character   Median : 1.000   Median :  56.00   Median :  57.00
##                     Mean   : 1.539   Mean   :  98.15   Mean   :  99.07
##                     3rd Qu.: 2.000   3rd Qu.: 119.00   3rd Qu.: 120.00
##                     Max.   :14.000   Max.   :2798.00   Max.   :2798.00
##
##        NP             HINCP             AGEP            POVPIP
##  Min.   : 1.000   Min.   :-60000   Min.   :18.00   Min.   :  1.0
##  1st Qu.: 2.000   1st Qu.: 43000   1st Qu.:36.00   1st Qu.:223.0
##  Median : 2.000   Median : 78105   Median :54.00   Median :383.0
##  Mean   : 2.614   Mean   : 95393   Mean   :51.97   Mean   :349.2
##  3rd Qu.: 3.000   3rd Qu.:122500   3rd Qu.:66.00   3rd Qu.:501.0
##  Max.   :14.000   Max.   :834000   Max.   :94.00   Max.   :501.0
##
##       ST            MIG          SEX                ESR
##  Length:5692        1:5118   Length:5692        Length:5692
##  Class :character   2: 574   Class :character   Class :character
##  Mode  :character            Mode  :character   Mode  :character
##
##
##
##
##             ST_label                                          MIG_label
##  North Dakota/ND:5692   N/A (less than 1 year old)              :   0
##  Alabama/AL     :   0   Yes, same house (nonmovers)              :5118
##  Alaska/AK      :   0   No, outside US and Puerto Rico           :   7
##  Arizona/AZ     :   0   No, different house in US or Puerto Rico: 567
##  Arkansas/AR    :   0
##  California/CA  :   0
##  (Other)        :   0
##   SEX_label                                          ESR_label
##  Male  :2897   N/A (less than 16 years old)                :   0
##  Female:2795   Civilian employed, at work                  :3539
##                Civilian employed, with a job but not at work:  52
##                Unemployed                                  :  86
##                Armed forces, at work                       :  52
##                Armed forces, with a job but not at work    :   1
##                Not in labor force                          :1962
##           poverty
##  above poverty:4459
##  below poverty:1233
##
##
##
##
##
```

```
#fitting the model
mod.1<- lm(POVPIP~ SEX + ESR + HINCP + NP + MIG + AGEP, data=ND_pums)
summary(mod.1)
```

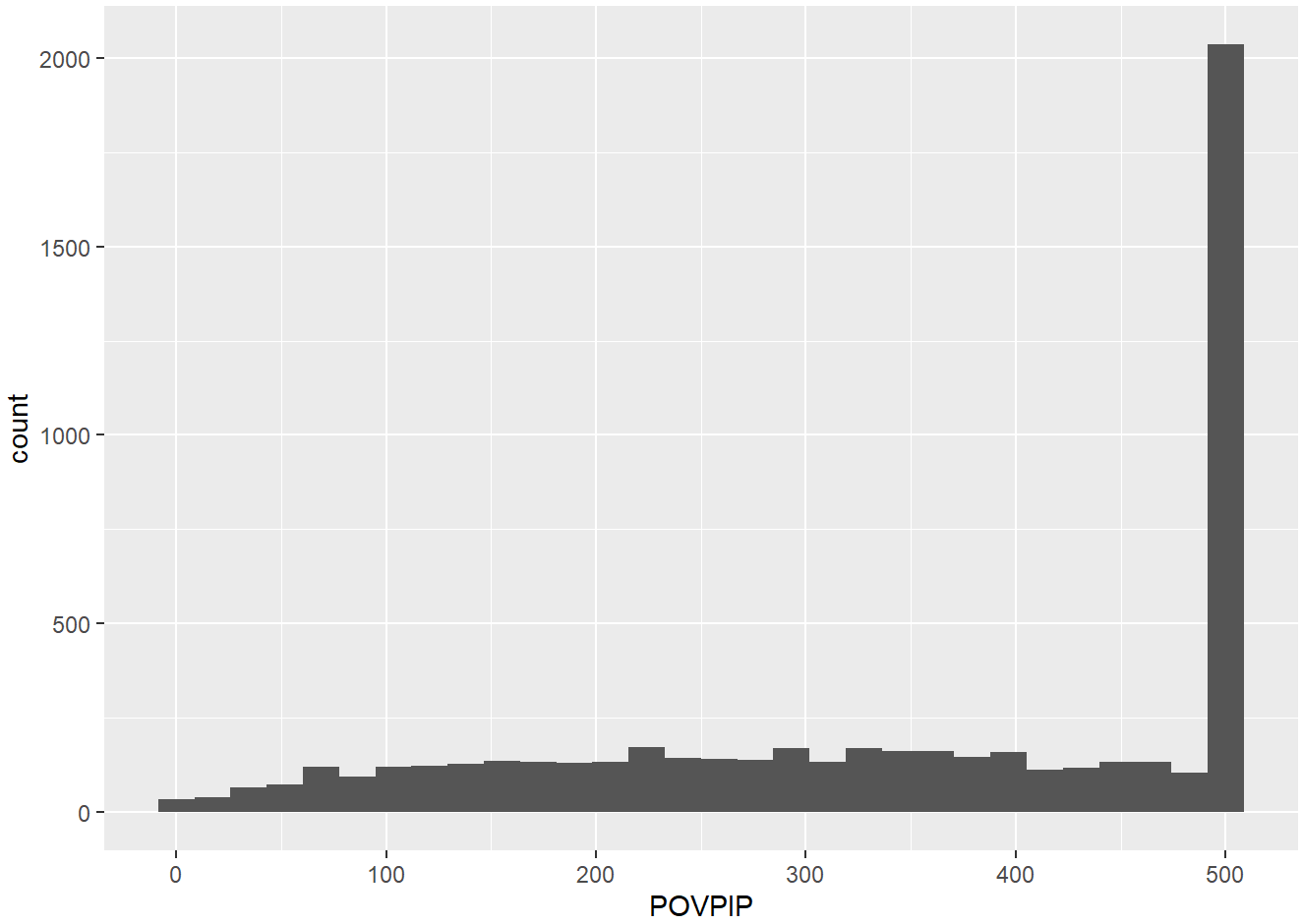
```
##
## Call:
## lm(formula = POVPIP ~ SEX + ESR + HINCP + NP + MIG + AGEP, data = ND_pums)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -680.89  -79.02   18.74   88.76  224.36
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.704e+02  7.287e+00  37.106  < 2e-16 ***
## SEX2        -4.606e+00  3.036e+00  -1.517 0.129270
## ESR2        -3.145e+01  1.590e+01  -1.978 0.047946 *
## ESR3        -4.764e+01  1.245e+01  -3.828 0.000131 ***
## ESR4         3.320e+00  1.599e+01   0.208 0.835513
## ESR5         4.969e-01  1.138e+02   0.004 0.996518
## ESR6        -7.034e+01  3.794e+00 -18.542  < 2e-16 ***
## HINCP        1.163e-03  1.959e-05  59.393  < 2e-16 ***
## NP          -1.686e+01  1.164e+00 -14.482  < 2e-16 ***
## MIG2        -3.426e+01  5.221e+00  -6.563 5.75e-11 ***
## AGEP         8.237e-01  1.050e-01   7.844 5.17e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 113.7 on 5681 degrees of freedom
## Multiple R-squared:  0.4507, Adjusted R-squared:  0.4498
## F-statistic: 466.2 on 10 and 5681 DF,  p-value: < 2.2e-16
```

```
coef(mod.1)
```

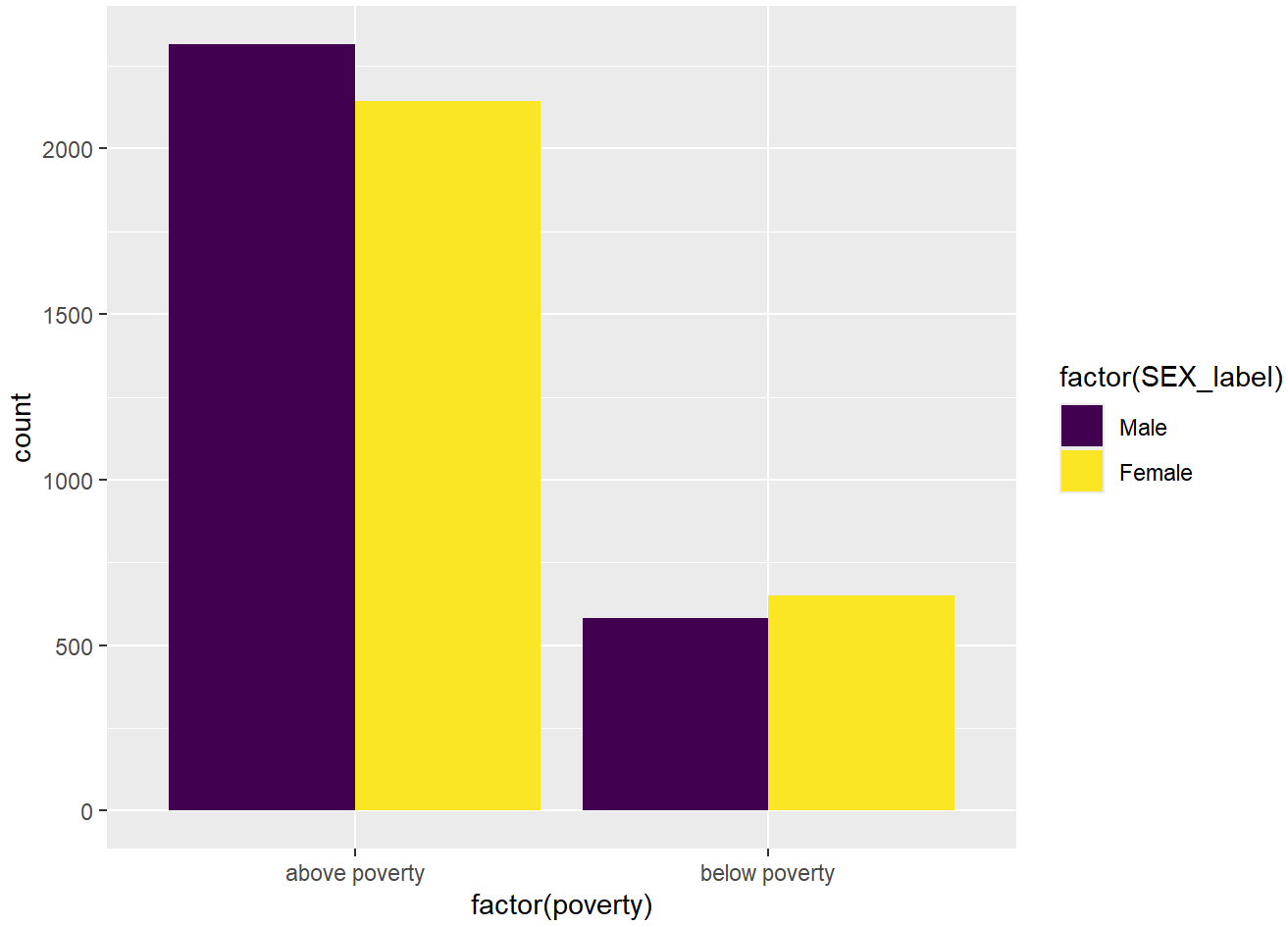
```
##   (Intercept)          SEX2          ESR2          ESR3          ESR4
## 270.392852180  -4.605655458 -31.445420580 -47.642175734   3.320220070
##          ESR5          ESR6         HINCP            NP          MIG2
##   0.496874749 -70.341743748   0.001163378 -16.858941151 -34.261618375
##          AGEP
##   0.823720384
```

```
ggplot(ND_pums)+
  geom_histogram(aes(x=POVPIP))
```

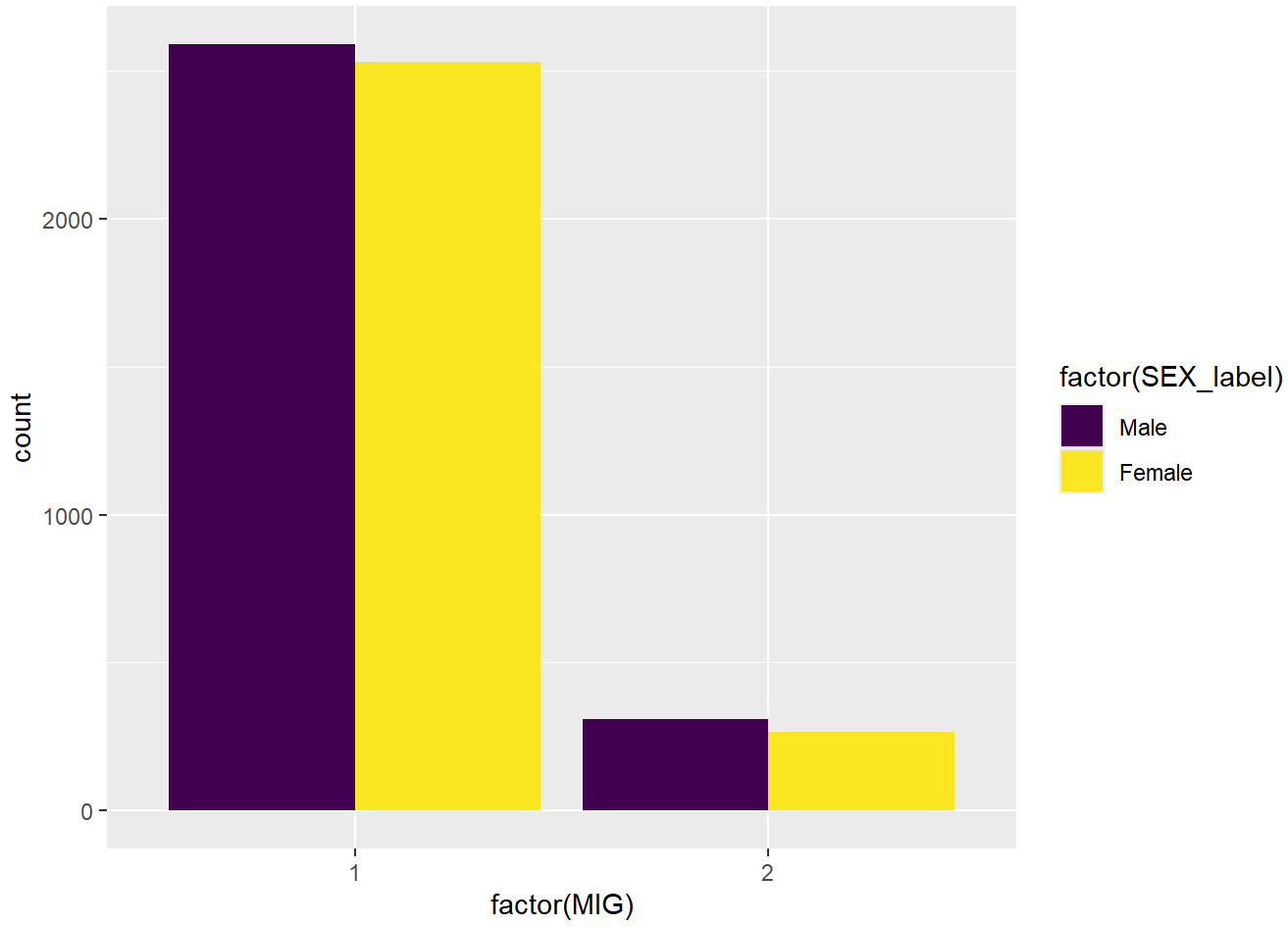
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(ND_pums) +
  geom_bar(aes(fill=factor(SEX_label), x= factor(poverty)), position = "dodge")
```



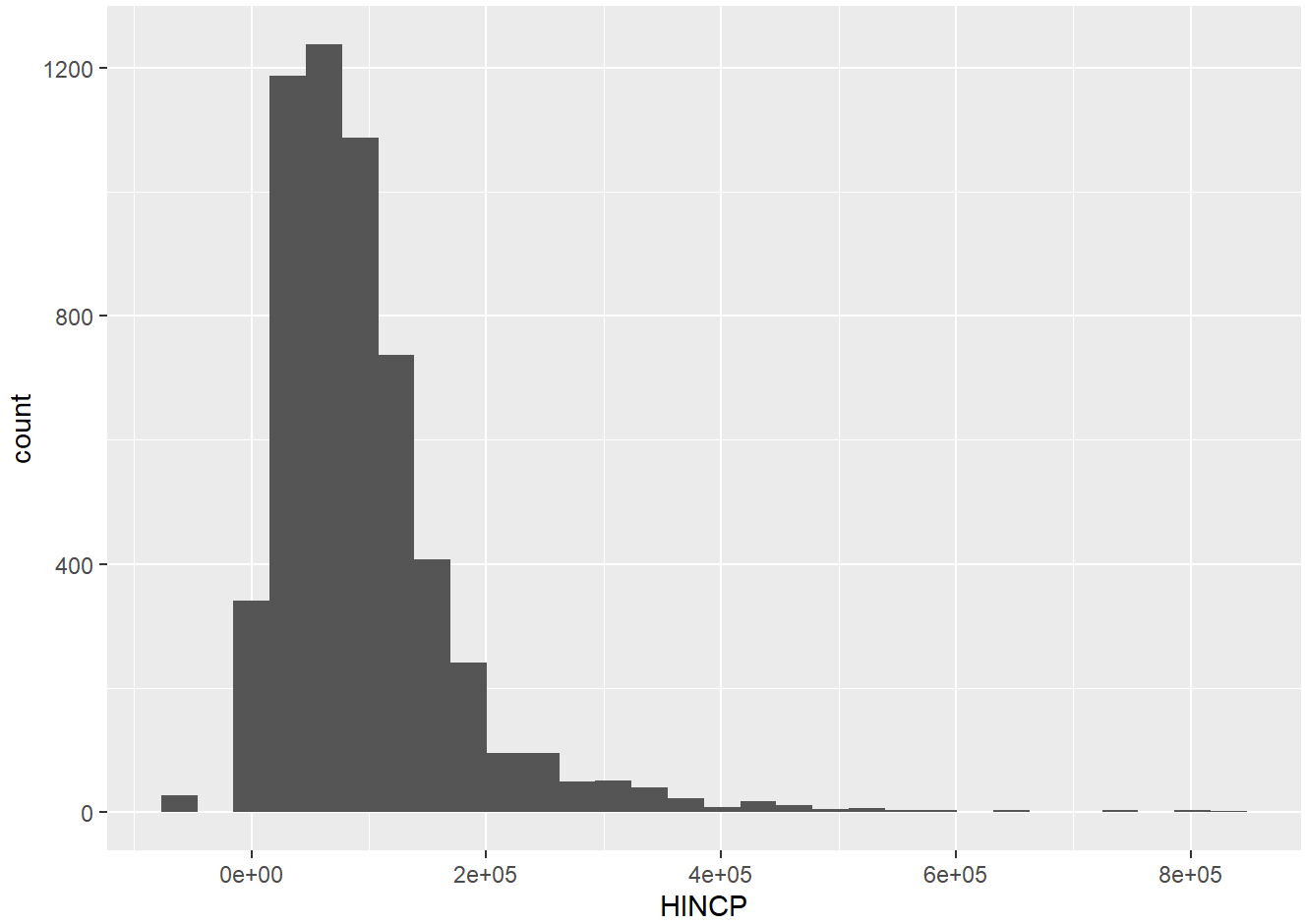

```
ggplot(ND_pums) +
  geom_bar(aes(fill=factor(SEX_label), x= factor(MIG)), position = "dodge")
```




```
ggplot(ND_pums)+
geom_histogram(aes(x=HINCP, bins=30))
```

```
## Warning in geom_histogram(aes(x = HINCP, bins = 30)): Ignoring unknown
## aesthetics: bins
```

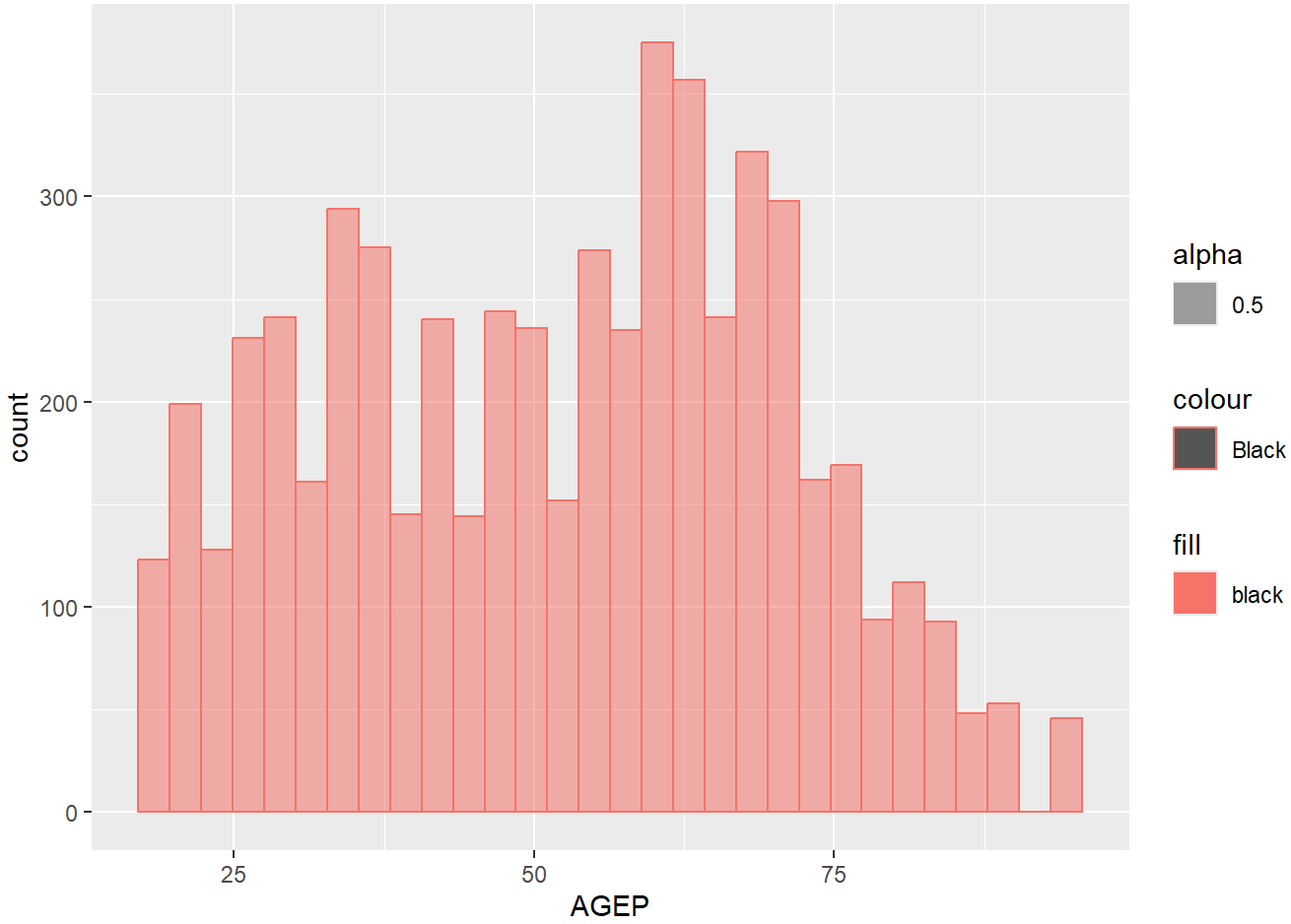
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



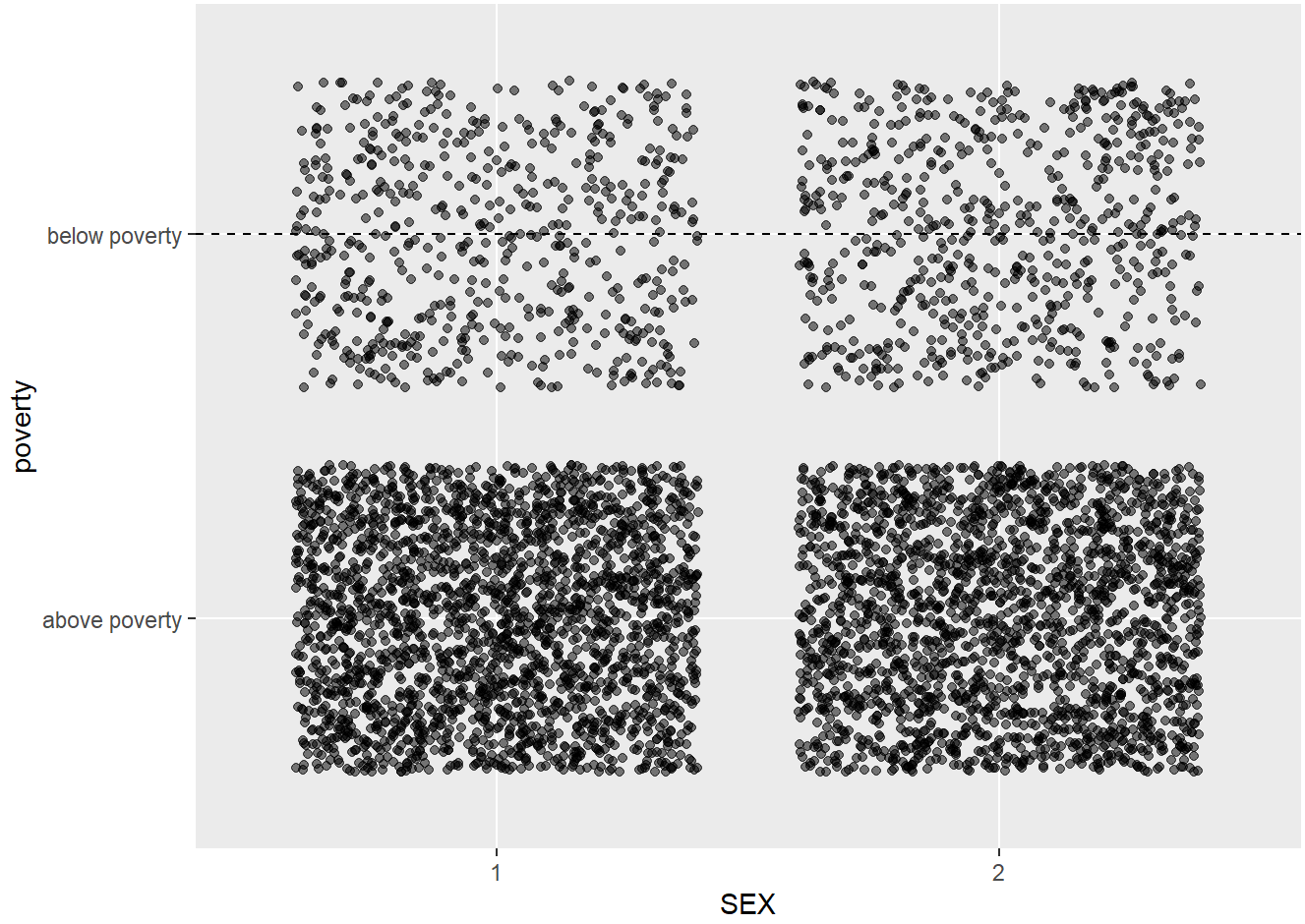

```
ggplot(ND_pums)+
geom_histogram(aes(x=AGEP,alpha= 0.5, bins=30, color= "Black", fill= "black"))
```

```
## Warning in geom_histogram(aes(x = AGEP, alpha = 0.5, bins = 30, color =
## "Black", : Ignoring unknown aesthetics: bins
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

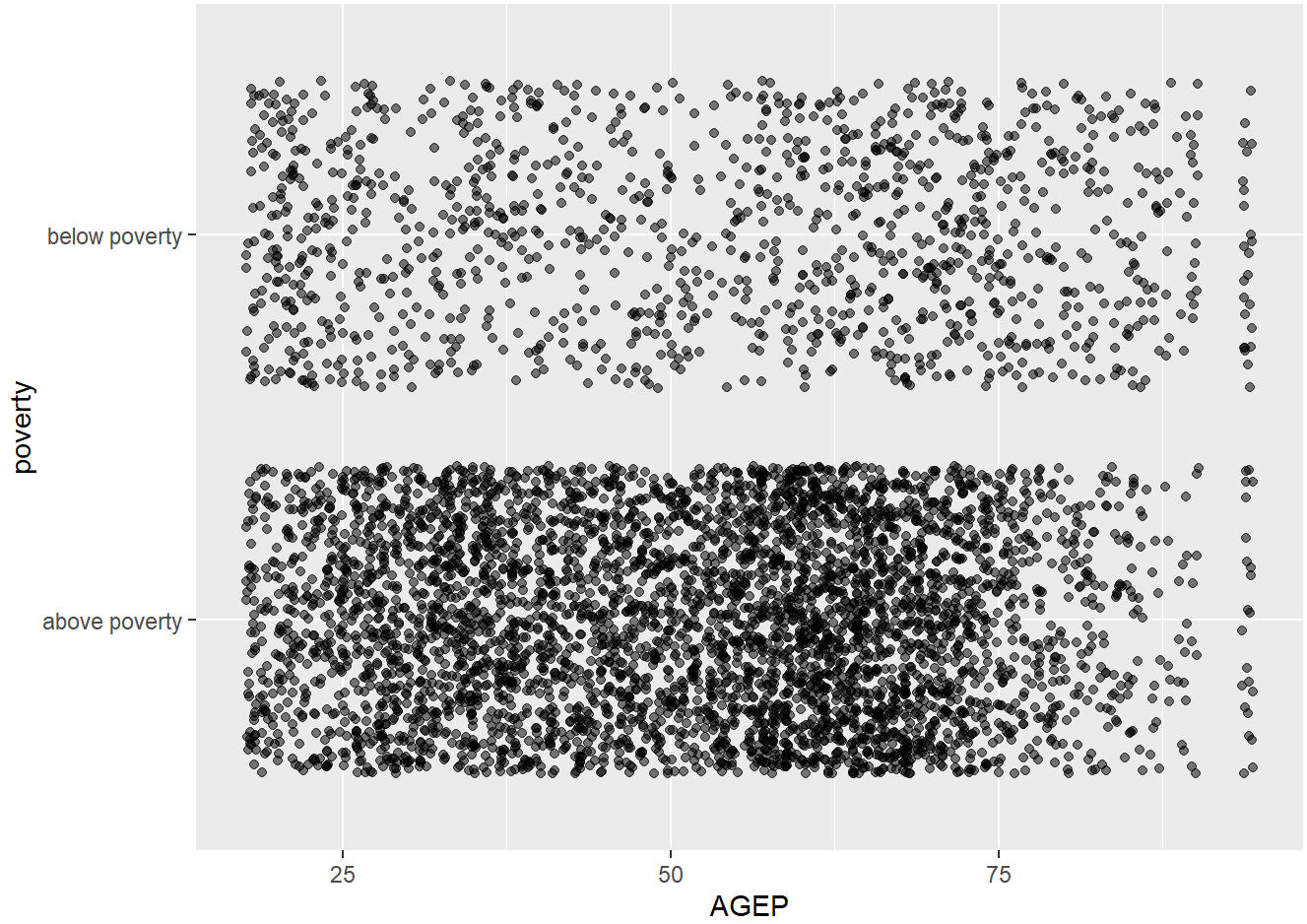



```
ggplot(ND_pums, aes(x= SEX, y = poverty)) +
  geom_jitter(alpha = 0.5) +
  geom_hline(yintercept = c(-2,2), lty=2)
```

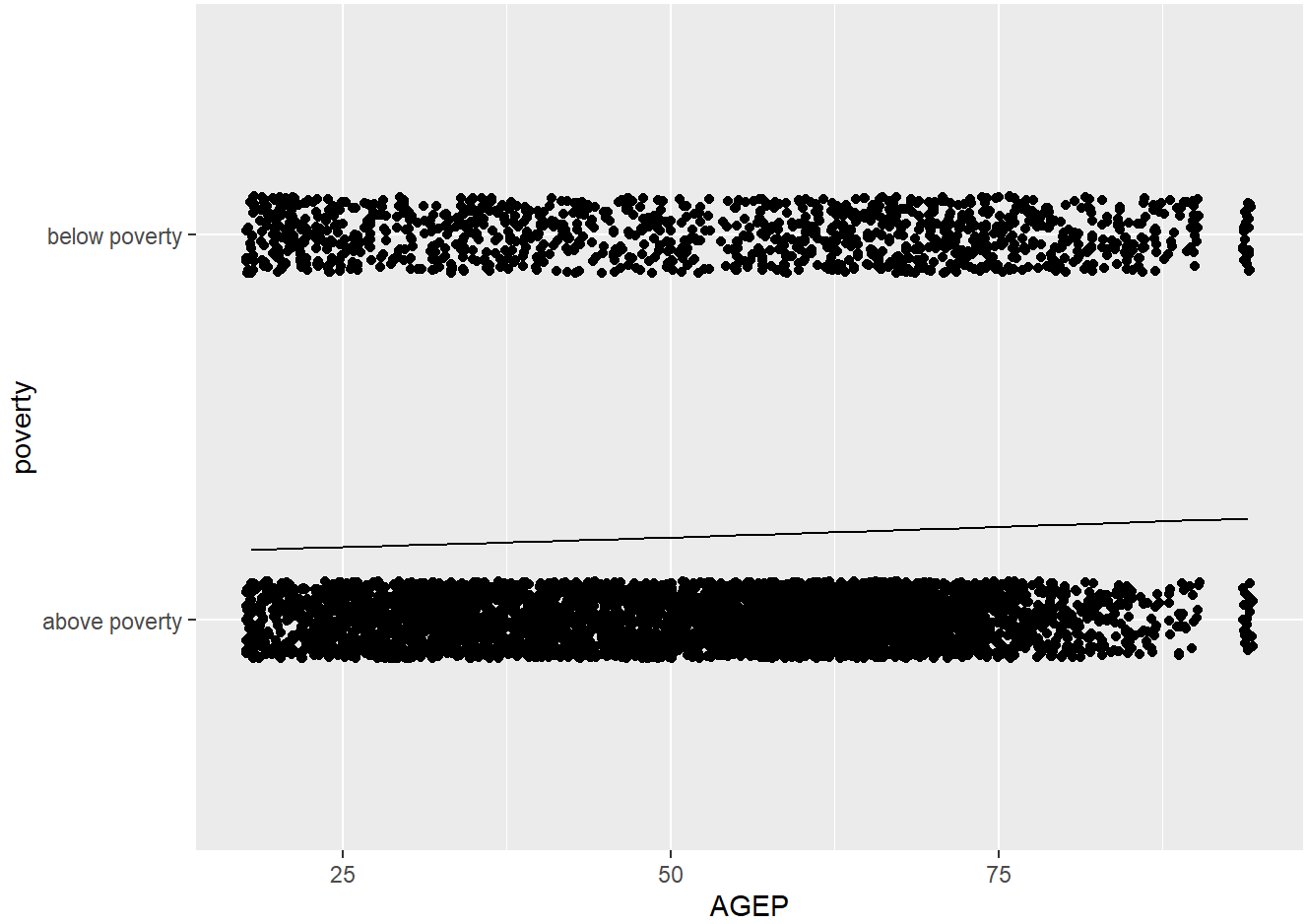



```
ggplot(ND_pums, aes(x= AGEP, y = poverty)) +
 geom_jitter(alpha = 0.5) +
  geom_abline(yintercept= 0)
```

```
## Warning in geom_abline(yintercept = 0): Ignoring unknown parameters:
## `yintercept`
```




```
#fitting all the explanatory variables aside HINCP
mod.2<-glm(poverty~ AGEP, data=ND_pums, family = "binomial")
ND_pums$pred.1<- predict(mod.2, type = "response")+1
ND_pums$res.1<- residuals(mod.2, type = "response")

ggplot(ND_pums, aes(x = AGEP)) +
  geom_jitter(aes(y = poverty), height = 0.1) +
  geom_line(aes(y = pred.1))
```




```
#deviance
display(mod.2)
```

```
## glm(formula = poverty ~ AGEP, family = "binomial", data = ND_pums)
##             coef.est coef.se
## (Intercept) -1.61     0.10
## AGEP         0.01     0.00
## ---
##   n = 5692, k = 2
##   residual deviance = 5936.3, null deviance = 5949.2 (difference = 12.9)
```

```
anova(mod.2, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: poverty
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                  5691     5949.2
## AGEP  1   12.929      5690     5936.3 0.0003236 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

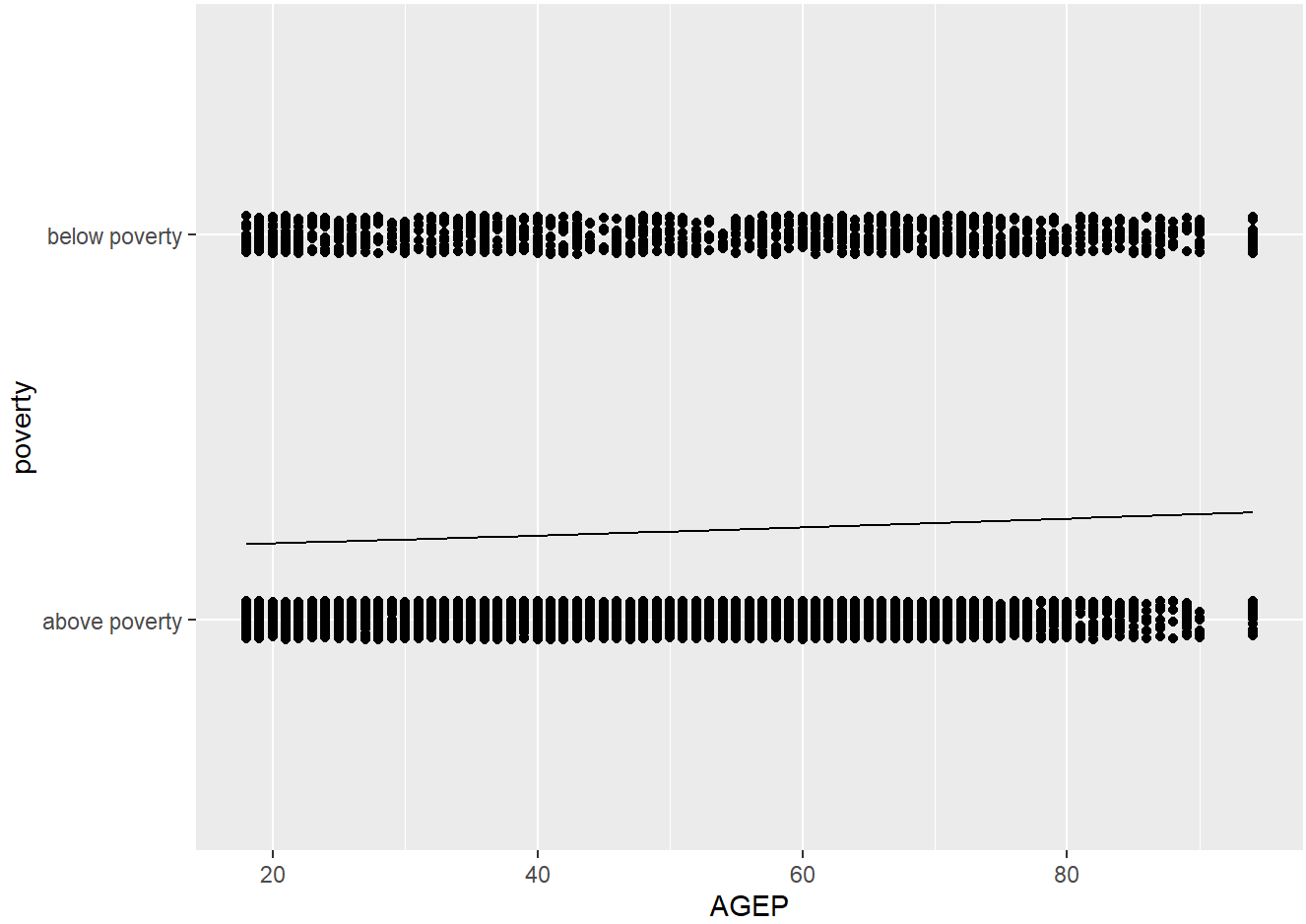
```
#AIC
summary(mod.2)
```

```
##
## Call:
## glm(formula = poverty ~ AGEP, family = "binomial", data = ND_pums)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.614261   0.098112 -16.453  < 2e-16 ***
## AGEP         0.006254   0.001743   3.589 0.000332 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5936.3  on 5690  degrees of freedom
## AIC: 5940.3
##
## Number of Fisher Scoring iterations: 4
```

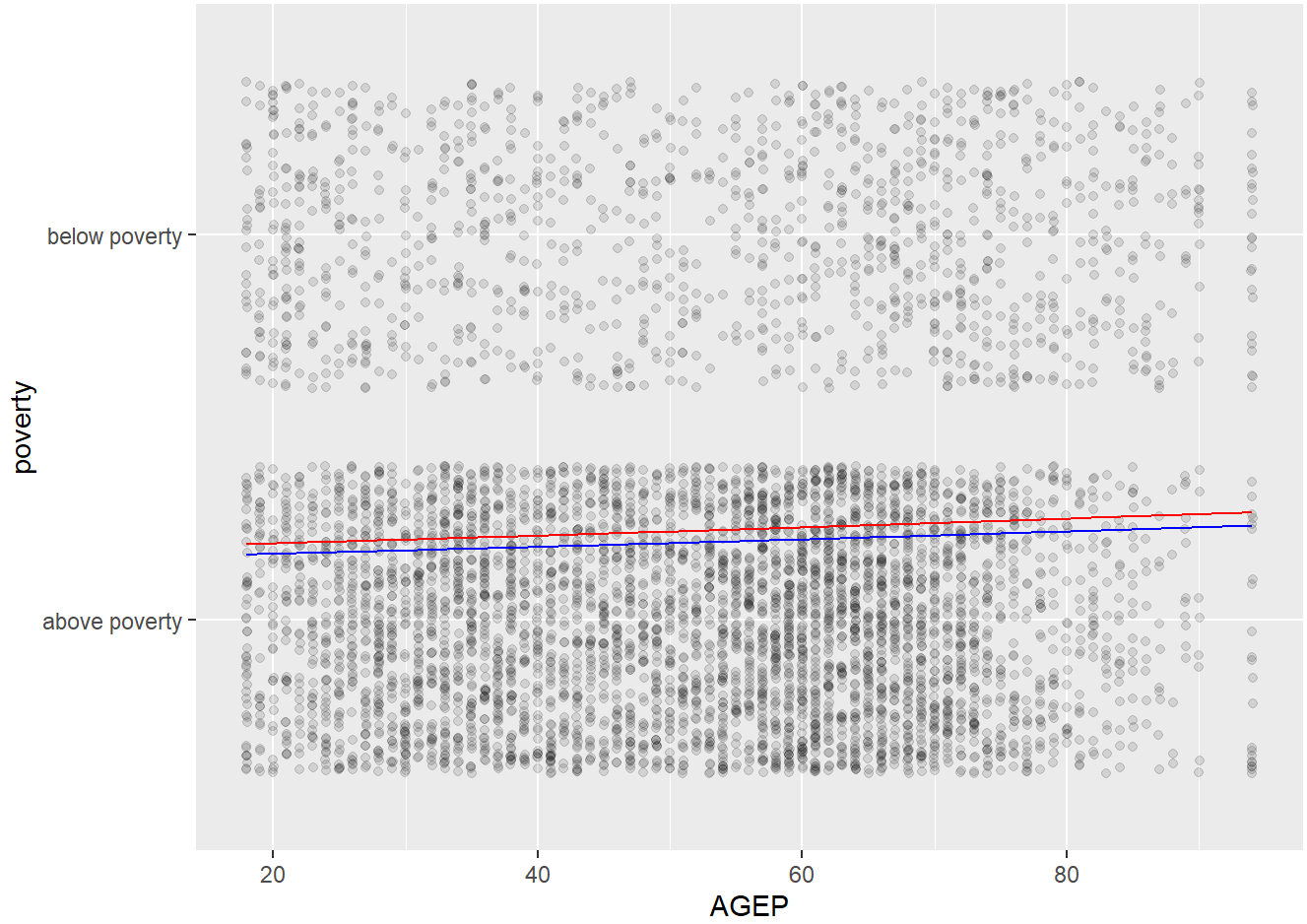
```
mod.3<-glm(poverty ~AGEP + SEX, data= ND_pums,family = "binomial")
theta.3 <- coef(mod.3)
theta.3
```

```
## (Intercept)        AGEP        SEX2
## -1.69445296  0.00605536  0.17965566
```

```
func.3.AGEP <- function(x){1+plogis(theta.3[[1]]+theta.3[[2]]*x+theta.3[[3]])}
ggplot(ND_pums, aes(x = AGEP)) +
  geom_jitter(aes(y=poverty), width = 0, height = 0.05) +
  geom_function(fun = func.3.AGEP)
```




```
func.3.male <- function(x){1+plogis(theta.3[[1]]+theta.3[[2]]*x)}
func.3.female <- function(x){1+plogis(theta.3[[1]]+theta.3[[2]]*x + theta.3[[3]])}

ggplot(ND_pums) +
  geom_jitter(aes(x=AGEP, y=poverty), alpha = 0.1, width = 0.05) +
  geom_function(aes(x=AGEP), fun = func.3.female, color = "red") +
  geom_function(aes(x=AGEP), fun = func.3.male, color = "blue")
```




```
#deviance
display(mod.3)
```

```
## glm(formula = poverty ~ AGEP + SEX, family = "binomial", data = ND_pums)
##             coef.est coef.se
## (Intercept) -1.69     0.10
## AGEP         0.01     0.00
## SEX2         0.18     0.06
## ---
##   n = 5692, k = 3
##   residual deviance = 5928.6, null deviance = 5949.2 (difference = 20.7)
```

```
anova(mod.2,mod.3, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: poverty ~ AGEP
## Model 2: poverty ~ AGEP + SEX
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      5690     5936.3
## 2      5689     5928.6  1   7.7546 0.005358 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#AIC
summary(mod.3)
```

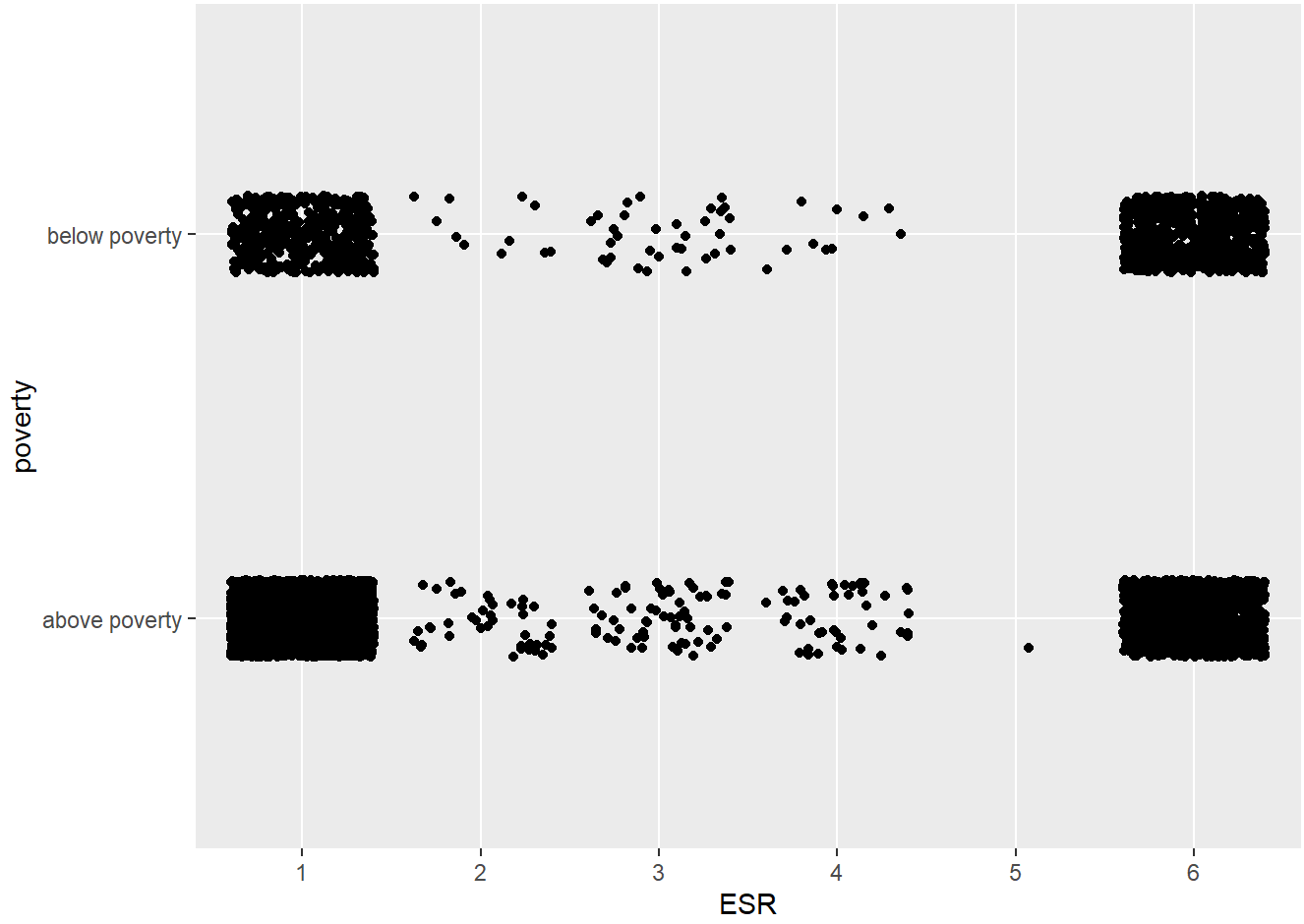
```
##
## Call:
## glm(formula = poverty ~ AGEP + SEX, family = "binomial", data = ND_pums)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.694453   0.102582 -16.518  < 2e-16 ***
## AGEP         0.006055   0.001745   3.471 0.000519 ***
## SEX2         0.179656   0.064559   2.783 0.005389 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5928.6  on 5689  degrees of freedom
## AIC: 5934.6
##
## Number of Fisher Scoring iterations: 4
```

```
person <- expand.grid(SEX=c("1","2"),NP=c(1,2,3,4,5), AGEP=c(20,30,40,50,60,70,80))

cbind(person,pred.1 = predict(mod.3,newdata=person,type="response"))
```

```
##    SEX NP AGEP    pred.1
## 1    1  1   20 0.1717399
## 2    2  1   20 0.1988193
## 3    1  2   20 0.1717399
## 4    2  2   20 0.1988193
## 5    1  3   20 0.1717399
## 6    2  3   20 0.1988193
## 7    1  4   20 0.1717399
## 8    2  4   20 0.1988193
## 9    1  5   20 0.1717399
## 10   2  5   20 0.1988193
## 11   1  1   30 0.1805254
## 12   2  1   30 0.2086410
## 13   1  2   30 0.1805254
## 14   2  2   30 0.2086410
## 15   1  3   30 0.1805254
## 16   2  3   30 0.2086410
## 17   1  4   30 0.1805254
## 18   2  4   30 0.2086410
## 19   1  5   30 0.1805254
## 20   2  5   30 0.2086410
## 21   1  1   40 0.1896573
## 22   2  1   40 0.2188154
## 23   1  2   40 0.1896573
## 24   2  2   40 0.2188154
## 25   1  3   40 0.1896573
## 26   2  3   40 0.2188154
## 27   1  4   40 0.1896573
## 28   2  4   40 0.2188154
## 29   1  5   40 0.1896573
## 30   2  5   40 0.2188154
## 31   1  1   50 0.1991389
## 32   2  1   50 0.2293422
## 33   1  2   50 0.1991389
## 34   2  2   50 0.2293422
## 35   1  3   50 0.1991389
## 36   2  3   50 0.2293422
## 37   1  4   50 0.1991389
## 38   2  4   50 0.2293422
## 39   1  5   50 0.1991389
## 40   2  5   50 0.2293422
## 41   1  1   60 0.2089723
## 42   2  1   60 0.2402196
## 43   1  2   60 0.2089723
## 44   2  2   60 0.2402196
## 45   1  3   60 0.2089723
## 46   2  3   60 0.2402196
## 47   1  4   60 0.2089723
## 48   2  4   60 0.2402196
## 49   1  5   60 0.2089723
## 50   2  5   60 0.2402196
## 51   1  1   70 0.2191584
```

```
## 52   2  1   70 0.2514447
## 53   1  2   70 0.2191584
## 54   2  2   70 0.2514447
## 55   1  3   70 0.2191584
## 56   2  3   70 0.2514447
## 57   1  4   70 0.2191584
## 58   2  4   70 0.2514447
## 59   1  5   70 0.2191584
## 60   2  5   70 0.2514447
## 61   1  1   80 0.2296968
## 62   2  1   80 0.2630127
## 63   1  2   80 0.2296968
## 64   2  2   80 0.2630127
## 65   1  3   80 0.2296968
## 66   2  3   80 0.2630127
## 67   1  4   80 0.2296968
## 68   2  4   80 0.2630127
## 69   1  5   80 0.2296968
## 70   2  5   80 0.2630127
```

```
mod.4<-glm(poverty~ ESR, data=ND_pums, family = "binomial")
ND_pums$pred.1<- predict(mod.4, type = "response")+1
ND_pums$res.1<- residuals(mod.4, type = "response")

ggplot(ND_pums, aes(x = ESR)) +
  geom_jitter(aes(y = poverty), height = 0.1) +
  geom_line(aes(y = pred.1))
```




```
#deviance
display(mod.4)
```

```
## glm(formula = poverty ~ ESR, family = "binomial", data = ND_pums)
##             coef.est coef.se
## (Intercept)  -1.84     0.05
## ESR2          0.52     0.34
## ESR3          1.27     0.23
## ESR4          0.40     0.36
## ESR5         -9.73   196.97
## ESR6          1.24     0.07
## ---
##   n = 5692, k = 6
##   residual deviance = 5597.0, null deviance = 5949.2 (difference = 352.2)
```

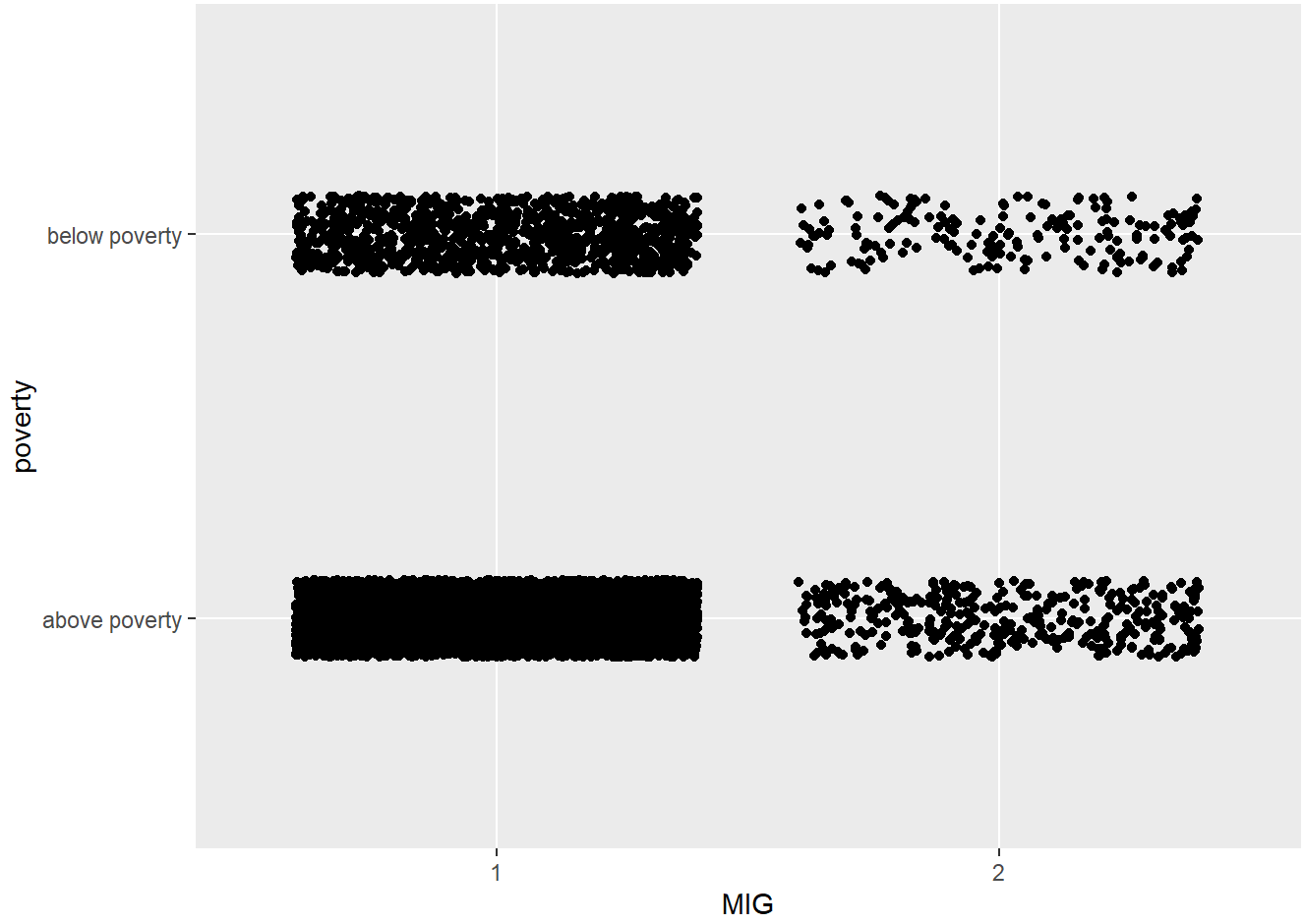
```
anova(mod.4, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: poverty
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                  5691     5949.2
## ESR   5   352.23      5686     5597.0 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod.4)
```

```
##
## Call:
## glm(formula = poverty ~ ESR, family = "binomial", data = ND_pums)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.84006    0.04888 -37.644  < 2e-16 ***
## ESR2          0.52438    0.34306   1.529    0.126
## ESR3          1.26671    0.22985   5.511 3.57e-08 ***
## ESR4          0.40497    0.35524   1.140    0.254
## ESR5         -9.72599  196.96769  -0.049    0.961
## ESR6          1.24179    0.06794  18.278  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5597.0  on 5686  degrees of freedom
## AIC: 5609
##
## Number of Fisher Scoring iterations: 10
```

```
mod.5<-glm(poverty~MIG, data=ND_pums, family = "binomial")
ND_pums$pred.1<- predict(mod.5, type = "response")+1
ND_pums$res.1<- residuals(mod.5, type = "response")

ggplot(ND_pums, aes(x = MIG)) +
  geom_jitter(aes(y = poverty), height = 0.1) +
  geom_line(aes(y = pred.1))
```




```
#deviance
display(mod.5)
```

```
## glm(formula = poverty ~ MIG, family = "binomial", data = ND_pums)
##             coef.est coef.se
## (Intercept) -1.36     0.03
## MIG2         0.61     0.10
## ---
##   n = 5692, k = 2
##   residual deviance = 5910.7, null deviance = 5949.2 (difference = 38.6)
```

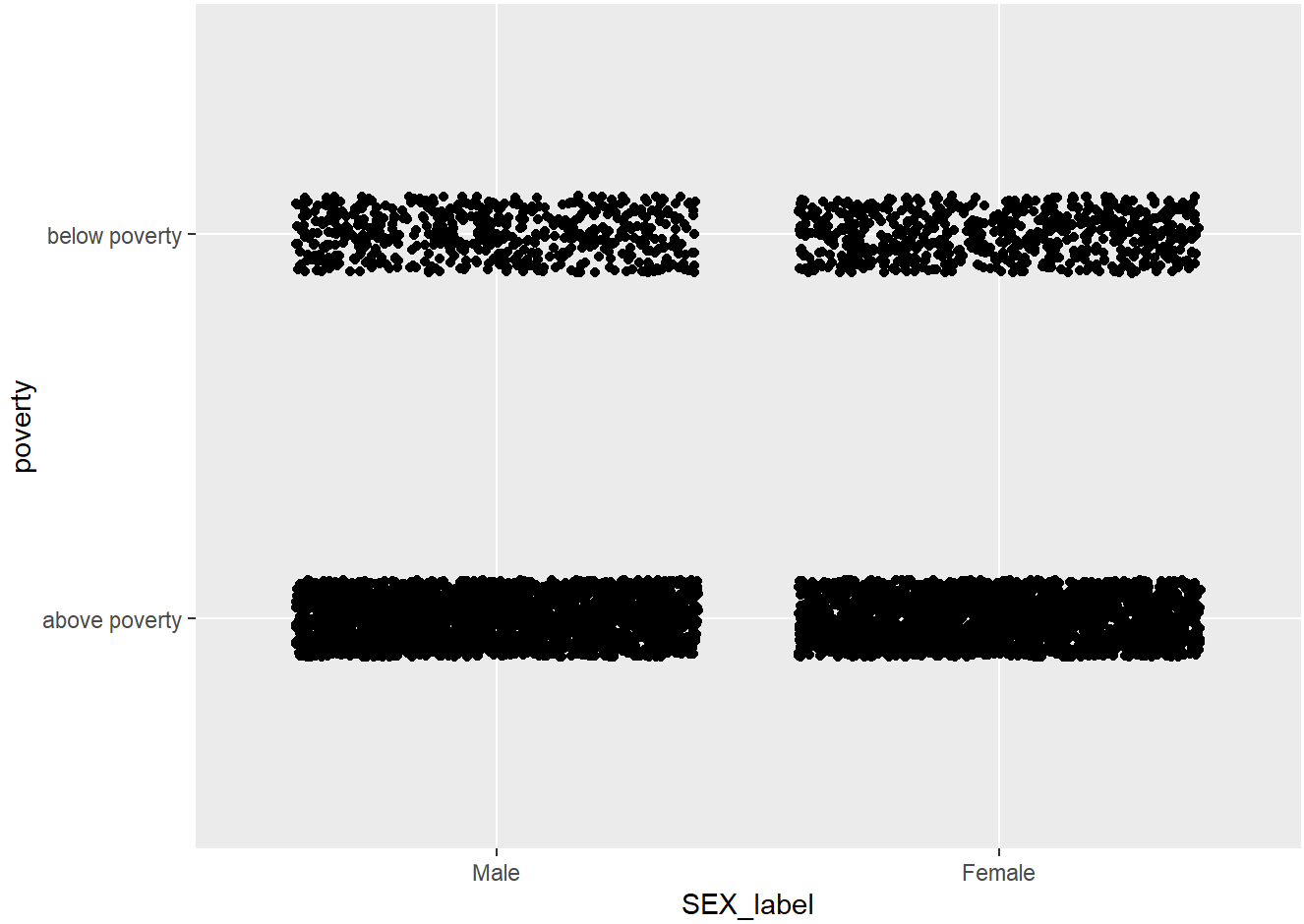
```
anova(mod.5, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: poverty
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                  5691     5949.2
## MIG   1   38.567      5690     5910.7 5.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod.5)
```

```
##
## Call:
## glm(formula = poverty ~ MIG, family = "binomial", data = ND_pums)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.35676    0.03464 -39.168  < 2e-16 ***
## MIG2         0.61354    0.09579   6.405  1.5e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5910.7  on 5690  degrees of freedom
## AIC: 5914.7
##
## Number of Fisher Scoring iterations: 4
```

```
mod.6<-glm(poverty ~ SEX_label, data= ND_pums,family = "binomial")
ND_pums$pred.1<- predict(mod.6, type = "response")+1
ND_pums$res.1<- residuals(mod.6, type = "response")

ggplot(ND_pums, aes(x = SEX_label)) +
  geom_jitter(aes(y = poverty), height = 0.1) +
  geom_line(aes(y = pred.1))
```




```
#deviance
display(mod.6)
```

```
## glm(formula = poverty ~ SEX_label, family = "binomial", data = ND_pums)
##             coef.est coef.se
## (Intercept) -1.29     0.03
## SEX_label.L  0.13     0.05
## ---
##   n = 5692, k = 2
##   residual deviance = 5940.6, null deviance = 5949.2 (difference = 8.6)
```

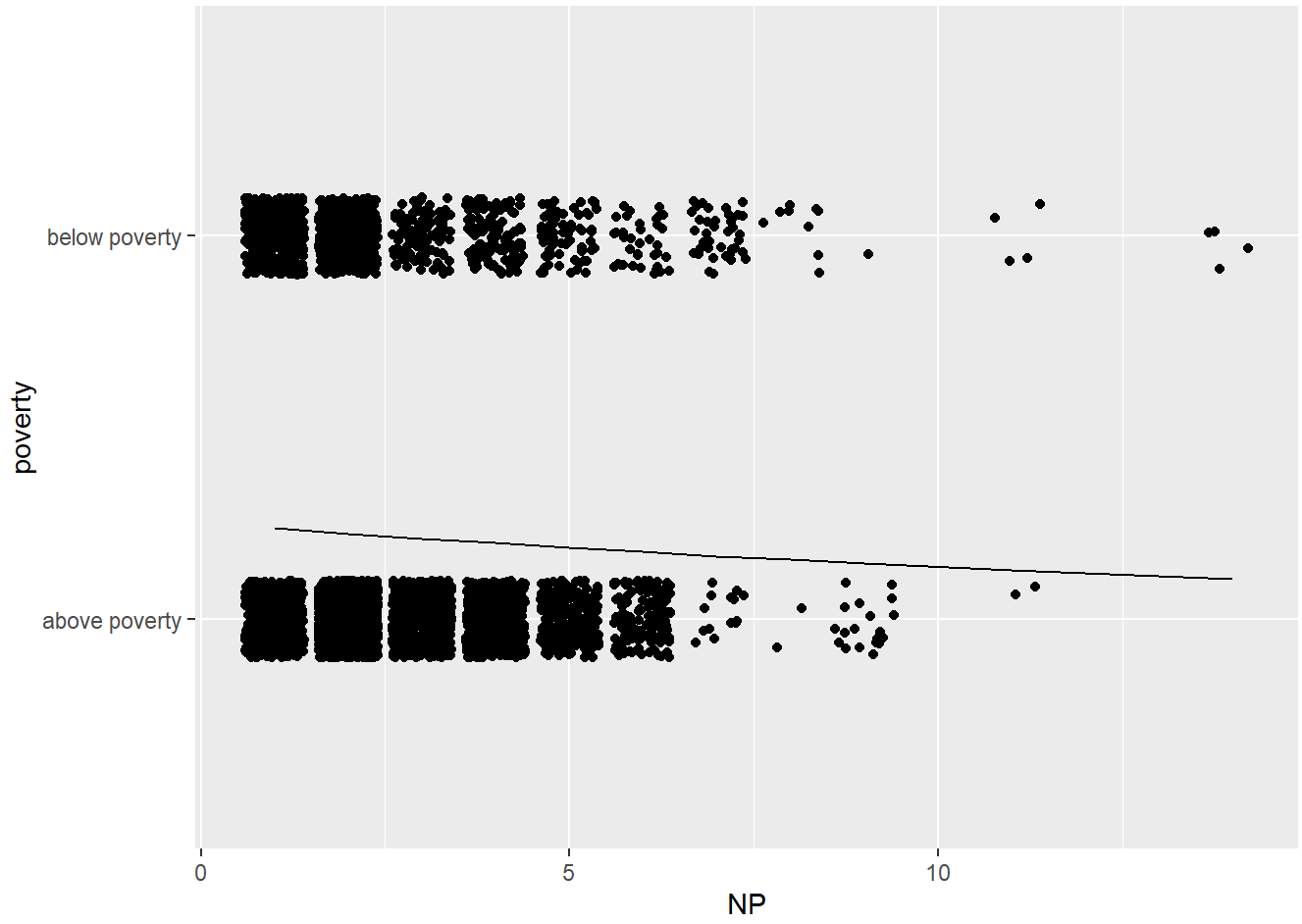
```
anova(mod.6, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: poverty
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                       5691     5949.2
## SEX_label  1   8.5948      5690     5940.6 0.003371 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#AIC
summary(mod.6)
```

```
##
## Call:
## glm(formula = poverty ~ SEX_label, family = "binomial", data = ND_pums)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.28631    0.03222 -39.922   <2e-16 ***
## SEX_label.L  0.13348    0.04557   2.929   0.0034 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5940.6  on 5690  degrees of freedom
## AIC: 5944.6
##
## Number of Fisher Scoring iterations: 4
```

```
mod.7<-glm(poverty ~ NP, data= ND_pums,family = "binomial")
ND_pums$pred.1<- predict(mod.7, type = "response")+1
ND_pums$res.1<- residuals(mod.7, type = "response")

ggplot(ND_pums, aes(x = NP)) +
  geom_jitter(aes(y = poverty), height = 0.1) +
  geom_line(aes(y = pred.1))
```




```
#deviance
display(mod.7)
```

```
## glm(formula = poverty ~ NP, family = "binomial", data = ND_pums)
##             coef.est coef.se
## (Intercept) -1.09     0.07
## NP          -0.08     0.02
## ---
##   n = 5692, k = 2
##   residual deviance = 5938.3, null deviance = 5949.2 (difference = 10.9)
```

```
anova(mod.7, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: poverty
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                  5691     5949.2
## NP    1    10.93      5690     5938.3 0.0009464 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

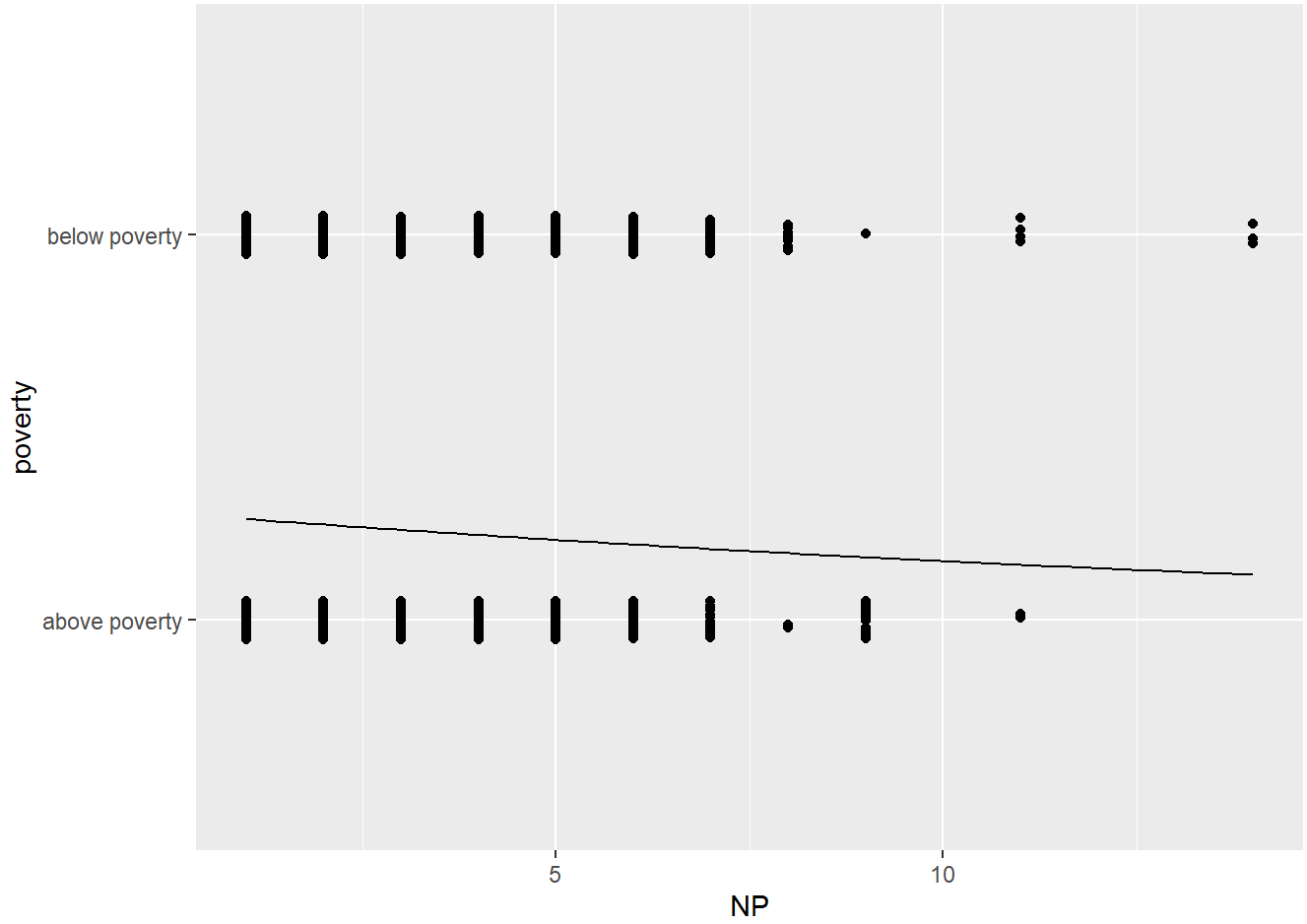
```
#AIC
summary(mod.7)
```

```
##
## Call:
## glm(formula = poverty ~ NP, family = "binomial", data = ND_pums)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.09204    0.06687 -16.330  < 2e-16 ***
## NP          -0.07532    0.02323  -3.243  0.00118 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5938.3  on 5690  degrees of freedom
## AIC: 5942.3
##
## Number of Fisher Scoring iterations: 4
```

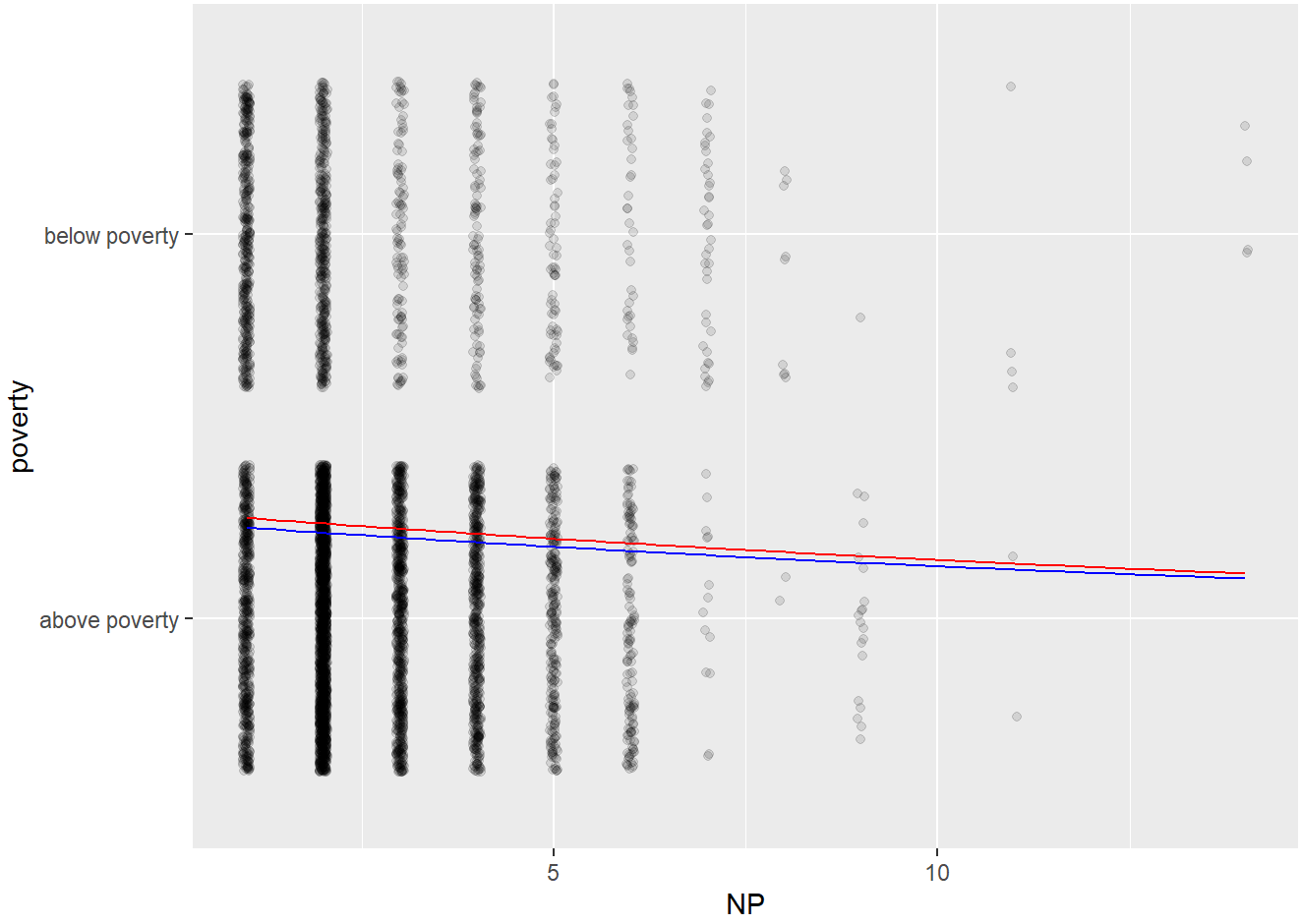
```
mod.8<-glm(poverty ~NP + SEX_label, data= ND_pums,family = "binomial")
theta.8 <- coef(mod.8)
theta.8
```

```
## (Intercept)          NP SEX_label.L
## -1.09252633 -0.07545555  0.13365792
```

```
func.8.NP <- function(x){1+plogis(theta.8[[1]]+theta.8[[2]]*x+theta.8[[3]])}
ggplot(ND_pums, aes(x = NP)) +
  geom_jitter(aes(y=poverty), width = 0, height = 0.05) +
  geom_function(fun = func.8.NP)
```




```
func.8.male <- function(x){1+plogis(theta.8[[1]]+theta.8[[2]]*x)}
func.8.female <- function(x){1+plogis(theta.8[[1]]+theta.8[[2]]*x + theta.8[[3]])}

ggplot(ND_pums) +
  geom_jitter(aes(x=NP, y=poverty), alpha = 0.1, width = 0.05) +
  geom_function(aes(x=NP), fun = func.8.female, color = "red") +
  geom_function(aes(x=NP), fun = func.8.male, color = "blue")
```




```
#deviance
display(mod.8)
```

```
## glm(formula = poverty ~ NP + SEX_label, family = "binomial",
##     data = ND_pums)
##             coef.est coef.se
## (Intercept) -1.09     0.07
## NP          -0.08     0.02
## SEX_label.L  0.13     0.05
## ---
##   n = 5692, k = 3
##   residual deviance = 5929.7, null deviance = 5949.2 (difference = 19.5)
```

```
anova(mod.7,mod.8, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: poverty ~ NP
## Model 2: poverty ~ NP + SEX_label
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      5690     5938.3
## 2      5689     5929.7  1    8.601  0.00336 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#AIC
summary(mod.8)
```

```
##
## Call:
## glm(formula = poverty ~ NP + SEX_label, family = "binomial",
##     data = ND_pums)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.09253    0.06697 -16.314  < 2e-16 ***
## NP          -0.07546    0.02326  -3.244  0.00118 **
## SEX_label.L  0.13366    0.04561   2.930  0.00338 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5949.2  on 5691  degrees of freedom
## Residual deviance: 5929.7  on 5689  degrees of freedom
## AIC: 5935.7
##
## Number of Fisher Scoring iterations: 4
```

```
ND_pums$pred.outcome.1 <- as.numeric(ND_pums$pred.1-1 > 0.5)

# Confusion matrix
addmargins(xtabs(~ pred.outcome.1 + poverty, data=ND_pums))
```

```
##               poverty
## pred.outcome.1 above poverty below poverty  Sum
##            0            4459          1233 5692
##            Sum          4459          1233 5692
```

```
cm <- xtabs(~ pred.outcome.1 + poverty, data=ND_pums)

# Overall %
addmargins(prop.table(cm))
```

```
##               poverty
## pred.outcome.1 above poverty below poverty       Sum
##            0       0.7833802     0.2166198 1.0000000
##            Sum     0.7833802     0.2166198 1.0000000
```

```
accuracy <- sum(diag(prop.table(cm)))
accuracy
```

```
## [1] 0.7833802
```

```
# % accuracy by prediction
addmargins(prop.table(cm, margin=1), margin = 2)
```

```
##               poverty
## pred.outcome.1 above poverty below poverty       Sum
##              0     0.7833802     0.2166198 1.0000000
```

```
# % accuracy by outcome
addmargins(prop.table(cm, margin=2), margin = 1)
```

```
##               poverty
## pred.outcome.1 above poverty below poverty
##            0               1             1
##            Sum             1             1
```

```
ND_pums%>%group_by(MIG)%>% summarize(n=n())%>%arrange(desc(n))
```

```
## # A tibble: 2 × 2
##   MIG       n
##   <fct> <int>
## 1 1      5118
## 2 2       574
```

```
ND_pums%>%group_by(MIG_label)%>% summarize(n=n())%>%arrange(desc(n))
```

```
## # A tibble: 3 × 2
##   MIG_label                                  n
##   <ord>                                <int>
## 1 Yes, same house (nonmovers)           5118
## 2 No, different house in US or Puerto Rico   567
## 3 No, outside US and Puerto Rico               7
```

```
ND_pums%>%group_by(SEX_label)%>% summarize(n=n())%>%arrange(desc(n))
```

```
## # A tibble: 2 × 2
##   SEX_label     n
##   <ord>     <int>
## 1 Male       2897
## 2 Female     2795
```

```
contrasts(ND_pums$SEX_label)
```

```
##              .L
## [1,] -0.7071068
## [2,]  0.7071068
```