```
In [1]:
library(foreign)
```

```
In [2]:
cps2015 = read.csv("cps2015.csv")
```

```
In [3]:
cps2015.include = subset(cps2015,cps2015$age>16)
```

```
In [4]:
cps2015.include.minWage = subset(cps2015.include,cps2015.include$hourwage <= 7.25)
summary(cps2015.include.minWage)
```

```
Out[4]:
     year          serial          hwtsupp             hhintype
 Min.   :2014   Min.   :  141   Min.   :  219.5   Interview:4320
 1st Qu.:2014   1st Qu.:29304   1st Qu.: 1484.3
 Median :2014   Median :57994   Median : 2579.3
 Mean   :2014   Mean   :54587   Mean   : 3114.0
 3rd Qu.:2014   3rd Qu.:79639   3rd Qu.: 4037.4
 Max.   :2014   Max.   :97898   Max.   :21319.6

                              region          statefip          statecensus
 Pacific Division              :776   California  : 476   California  : 476
 South Atlantic Division       :715   Texas       : 328   Texas       : 328
 West North Central Division:533      Florida     : 146   Florida     : 146
 West South Central Division:487      Pennsylvania: 128   Pennsylvania: 128
 East North Central Division:457      Illinois    : 125   Illinois    : 125
 Mountain Division             :457   Minnesota   : 112   Minnesota   : 112
 (Other)                       :895   (Other)     :3005   (Other)     :3005
 asecflag      hflag                            metro
 ASEC:4320   3/8 file:1664   Central city               :1130
             5/8 file:2656   Central city status unknown: 729
                             Not identifiable           :  28
                             Not in metro area          : 805
                             Outside central city       :1628




                                                 metarea
 NIU, household not in a metropolitan area         : 805
 Missing data                                      : 236
 Los Angeles-Long Beach-Santa Ana, CA             : 198
 Washington, DC/MD/VA                             : 123
 New York-Northern New Jersey-Long Island, NY-NJ-PA: 116
 Honolulu, HI                                     : 102
 (Other)                                           :2740
           nfams              ncouples              nmothers
 1 family or N/A:3641   0 couples or NIU:1552   0 mothers or NIU:1059
 2 families     : 547   1               :2695   1               :2935
 3             :  80    2               :  68   2               : 313
 4             :  33    3               :   5    3              :  13
 5             :  18    4               :   0    4              :   0
 6             :   1
 (Other)       :   0
           nfathers      month        pernum          spmeitc
 0 fathers or NIU:1659   March:4320   Min.   : 1.000   Min.   :    0.0
 1               :2584                1st Qu.: 1.000   1st Qu.:    0.0
 2               :  77                Median : 2.000   Median :    0.0
 3               :   0                Mean   : 1.919   Mean   :  780.8
 4               :   0                3rd Qu.: 2.000   3rd Qu.:  487.0
                                      Max.   :10.000   Max.   :10053.0


     wtsupp                  nchild                        relate
 Min.   :  219.5   0 children present:1833   Head/householder    :1957
 1st Qu.: 1534.9   2                 : 941   Spouse              :1019
 Median : 2651.7   1 child present   : 917   Child               : 726
 Mean   : 3248.0   3                 : 455   Unmarried partner   : 176
 3rd Qu.: 4231.1   4                 : 120   Other relatives, n.s.:  96
 Max.   :27746.2   5                 :  36   Housemate/roomate   :  87
                   (Other)           :  18   (Other)             : 259
     age          sex                          race
 Min.   :17.00   Female:2324   White                        :3144
 1st Qu.:27.00   Male  :1996   Black/Negro                  : 669
 Median :37.00                 Asian only                   : 289
 Mean   :37.75                 American Indian/Aleut/Eskimo  :  64
 3rd Qu.:47.00                 Hawaiian/Pacific Islander only:  44
 Max.   :80.00                 White-American Indian         :  33
                               (Other)                      :  77
                   marst                                   hispan
 Divorced              : 398   Not Hispanic                          :3204
 Married, spouse absent :  74   Mexican                              : 773
 Married, spouse present:2164   Central American, (excluding Salvadoran):  68
 Never married/single  :1522   Puerto Rican                          :  65
 Separated             : 105   Other Hispanic                        :  57
 Widowed               :  57   South American                        :  55
                               (Other)                               :  98
                                                 educ
 High school diploma or equivalent                  :1390
 Some college but no degree                         : 963
 Bachelor's degree                                  : 603
 Associate's degree, academic program               : 304
 Associate's degree, occupational/vocational program: 238
```

```
Grade 11                                        : 180
(Other)                                         : 642
                              educ99
High school graduate, or GED     :1390
Some college, no degree          : 963
Bachelors degree                 : 603
Associate degree, academic program : 304
Associate degree, occupational program: 238
11th grade                       : 180
(Other)                          : 642
                           schlcoll
College or university full time           : 268
College or university part time           : 138
Does not attend school, college or university:3261
High school full time                     : 145
High school part time                     :   6
NIU                                       : 502

                    empstat                          labforce
At work                      :4194   NIU                        :   0
Has job, not at work last week: 126  No, not in the labor force:   0
Armed Forces                 :   0   Yes, in the labor force   :4320
NILF, other                  :   0
NILF, retired                :   0
NILF, unable to work         :   0
(Other)                      :   0
     occ           ind        uhrsworkly       uhrsworkt
Min.   :  10   Min.   : 170   Min.   : 1.00   40         :2187
1st Qu.:3800   1st Qu.:4870   1st Qu.:32.00   Hours vary: 285
Median :4700   Median :7580   Median :40.00   20         : 232
Mean   :4972   Mean   :6273   Mean   :36.13   30         : 168
3rd Qu.:6260   3rd Qu.:8470   3rd Qu.:40.00   35         : 163
Max.   :9750   Max.   :9590   Max.   :99.00   25         : 113
                              NA's   :145     (Other)    :1172
   uhrswork1       wksunem1        fullpart        hourwage        paidhour
Min.   : 0.00   Min.   : 0.00   Full-time:3054   Min.   :0.0000   NIU:   0
1st Qu.:25.00   1st Qu.: 0.00   NIU      : 145   1st Qu.:0.0000   No :   0
Median :40.00   Median : 0.00   Part-time:1121   Median :0.0000   Yes:4320
Mean   :32.97   Mean   : 8.02                    Mean   :0.3262
3rd Qu.:40.00   3rd Qu.:12.00                    3rd Qu.:0.0000
Max.   :99.00   Max.   :51.00                    Max.   :7.2500
                NA's   :3346
                  whyunemp                          whyptly
Job leaver           :   0   Could not find full time job: 208
Job loser - on layoff:   0   NIU                        :2818
New entrant          :   0   Other                      : 426
NIU                  :4320   Slack work                 : 310
Other job loser      :   0   Wanted part time           : 558
Re-entrant           :   0
Temporary job ended  :   0
    ftotval          incwage                      offpov          offcutoff
Min.   :      0   Min.   :      0   Above Poverty Line:3907   Min.   :11173
1st Qu.:  31375   1st Qu.:  12000   Below Poverty Line: 413   1st Qu.:15600
Median :  60000   Median :  25000                             Median :18769
Mean   :  73622   Mean   :  31247                             Mean   :21639
3rd Qu.:  98000   3rd Qu.:  41000                             3rd Qu.:24421
Max.   :1147518   Max.   :1107999                             Max.   :51154


                            poverty           eitcred
100-124 percent of the low-income level : 222   Min.   :   0.0
125-149 percent of the low-income level : 221   1st Qu.:   0.0
150 percent and above the low-income level:3464   Median :   0.0
Below poverty                           : 413   Mean   : 431.4
                                                3rd Qu.:   0.0
                                                Max.   :6044.0


                      disabwrk           health
Disability limits or prevents work: 100   Excellent:1305
No disability that affects work   :4220   Fair     : 231
                                          Good     :1152
                                          Poor     :  46
                                          Very good:1586


                    quitsick            diffany         gotwic
No, did not quit job or retire:4296   Has difficulty: 128   NIU:2224
Yes, quit job or retired    :  24   NIU           :   0   No :1988
                                    No difficulty :4192   Yes: 108
```

```
                     ftype        StateAbbreviation         StateName
Nonfamily householder: 306   CA     : 476     CALIFORNIA  : 476
Primary family      :3476    TX     : 328     TEXAS       : 328
Related subfamily    : 201   FL     : 146     FLORIDA     : 146
Secondary individual : 316   PA     : 128     PENNSYLVANIA: 128
Unrelated subfamily  :  21   IL     : 125     ILLINOIS    : 125
                             MN     : 112     MINNESOTA   : 112
                             (Other):3005     (Other)     :3005
StateMinimumWage StateUnemploymentRate      state_name      GDP2014
Min.   :7.25     Min.   :2.800       California : 476   Min.   :  29613
1st Qu.:7.25     1st Qu.:5.100        Texas     : 328   1st Qu.: 132064
Median :7.25     Median :5.900        Florida   : 146   Median : 316204
Mean   :7.80     Mean   :5.929        Pennsylvania: 128 Mean   : 652356
3rd Qu.:8.15     3rd Qu.:6.900        Illinois  : 125   3rd Qu.: 745875
Max.   :9.50     Max.   :8.000        Minnesota : 112   Max.   :2311616
                                      (Other)   :3005
 p_total2014     poverty_indicator    employed      weeks_worked
Min.   : 0.200   Min.   :0.0000    Min.   :0.0000   Min.   : 1.00
1st Qu.: 0.800   1st Qu.:0.0000    1st Qu.:1.0000   1st Qu.:40.00
Median : 1.800   Median :0.0000    Median :1.0000   Median :52.00
Mean   : 3.763   Mean   :0.0956    Mean   :0.9708   Mean   :43.98
3rd Qu.: 4.300   3rd Qu.:0.0000    3rd Qu.:1.0000   3rd Qu.:52.00
Max.   :13.300   Max.   :1.0000    Max.   :1.0000   Max.   :52.00
                                                    NA's   :3346

hours_per_year   hourly_wage          ln_wage          lessHS
Min.   :  16   Min.   :   0.438   Min.   :1.981   Min.   :0.0000
1st Qu.: 780   1st Qu.:  16.618   1st Qu.:1.981   1st Qu.:0.0000
Median :1300   Median :  38.544   Median :1.981   Median :0.0000
Mean   :1336   Mean   :  88.294   Mean   :2.051   Mean   :0.1516
3rd Qu.:1872   3rd Qu.:  84.567   3rd Qu.:2.098   3rd Qu.:0.0000
Max.   :5148   Max.   :2731.525   Max.   :2.251   Max.   :1.0000
NA's   :3346   NA's   :3346
     HS            somecoll          college          white
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
Mean   :0.3218   Mean   :0.3484   Mean   :0.1782   Mean   :0.4819
3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000

     black            asian            hispanic          other
Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
Median :0.0000   Median :0.00000   Median :0.0000   Median :0.00000
Mean   :0.1488   Mean   :0.06551   Mean   :0.2583   Mean   :0.04537
3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.00000
Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
```

In [5]:

```
table(cps2015.include.minWage$poverty_indicator)
```

Out[5]:

```
   0    1
3907  413
```

# Delete this part of the code!!!!!! Turns out, they are the bugs that preventing the prediction function.

In [ ]:

```
# refactorize variables
#varNames = c("black","asian","hispanic","other","lessHS","HS","somecoll")
#cps2015.include.minWage$black=factor(cps2015.include.minWage$black)
#cps2015.include.minWage$asian=factor(cps2015.include.minWage$asian)
#cps2015.include.minWage$hispanic=factor(cps2015.include.minWage$hispanic)
#cps2015.include.minWage$other = factor(cps2015.include.minWage$other)
#cps2015.include.minWage$lessHS=factor(cps2015.include.minWage$lessHS)
#cps2015.include.minWage$HS=factor(cps2015.include.minWage$HS)
#cps2015.include.minWage$somecoll=factor(cps2015.include.minWage$somecoll)
#cps2015.include.minWage$sex=factor(cps2015.include.minWage$sex)
```

In [6]:

```
# rescale GDP
cps2015.include.minWage$GDP2014 = cps2015.include.minWage$GDP2014/10e6
# regression model
lm.model = glm(poverty_indicator~ln_wage+StateUnemploymentRate+sex+GDP2014+employed+
                black+asian+hispanic+other+lessHS+HS+somecoll,family=binomial(link='logit'),
               data=cps2015.include.minWage
              )

# print the table
summary(lm.model)
```

Out[6]:

```
Call:
glm(formula = poverty_indicator ~ ln_wage + StateUnemploymentRate +
    sex + GDP2014 + employed + black + asian + hispanic + other +
    lessHS + HS + somecoll, family = binomial(link = "logit"),
    data = cps2015.include.minWage)

Deviance Residuals:
    Min      1Q   Median       3Q      Max
-0.9067  -0.4917  -0.3982  -0.2984   2.8196

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)           -0.63543    1.55386  -0.409   0.6826
ln_wage               -1.48282    0.79942  -1.855   0.0636 .
StateUnemploymentRate  0.11108    0.05205   2.134   0.0328 *
sexMale               -0.57791    0.10995  -5.256 1.47e-07 ***
GDP2014               -1.87749    0.88695  -2.117   0.0343 *
employed              -0.12875    0.29444  -0.437   0.6619
black                  0.72376    0.14738   4.911 9.06e-07 ***
asian                 -0.30034    0.30006  -1.001   0.3169
hispanic               0.64998    0.13908   4.673 2.96e-06 ***
other                  0.44425    0.24977   1.779   0.0753 .
lessHS                 1.43255    0.22250   6.438 1.21e-10 ***
HS                     1.04791    0.20840   5.028 4.94e-07 ***
somecoll               0.82770    0.20915   3.957 7.57e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2724.3  on 4319  degrees of freedom
Residual deviance: 2584.0  on 4307  degrees of freedom
AIC: 2610

Number of Fisher Scoring iterations: 6
```

In [37]:

```
library(coefplot)
coefplot(lm.model, parm = -1)
```

In [7]:

```
library(car)
```

In [8]:

```
avPlots(lm.model)
```

In [9]:

```
library(visreg)
```

In [10]:

```
visreg(lm.model)
```

```
pframe = read.csv("prediction table.csv")
```

```
pframe[1,]
```

Out[12]:

| | In_wage | StateUnemploymentRate | sex | GDP2014 | employed | black | asian | hispanic | white | other | lessHS | HS | somecoll |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.98 | 5.929491 | Female | 0.06523563 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

```
library(ggplot2)
```

```
prediction=predict.glm(lm.model,newdata=pframe,type="response")
```

|    | prediction |
|----|------------|
| 1  | 0.2674337  |
| 2  | 0.2532302  |
| 3  | 0.1504035  |
| 4  | 0.1990371  |
| 5  | 0.1875362  |
| 6  | 0.1075436  |
| 7  | 0.1662364  |
| 8  | 0.156261   |
| 9  | 0.08816113 |
| 10 | 0.1965428  |
| 11 | 0.1851528  |
| 12 | 0.1060441  |
| 13 | 0.1427439  |
| 14 | 0.1339518  |
| 15 | 0.07471351 |
| 16 | 0.1178551  |
| 17 | 0.1103986  |
| 18 | 0.06084457 |
| 19 | 0.1828715  |
| 20 | 0.172104   |
| 21 | 0.09790074 |
| 22 | 0.132199   |
| 23 | 0.1239625  |
| 24 | 0.06879094 |
| 25 | 0.1089152  |
| 26 | 0.1019589  |
| 27 | 0.05595492 |
| 28 | 0.1481618  |
| 29 | 0.1390901  |
| 30 | 0.07778357 |
| 31 | 0.1058611  |
| 32 | 0.09907819 |
| 33 | 0.05429539 |
| 34 | 0.0867524  |
| 35 | 0.08108281 |
| 36 | 0.04403625 |
| 37 | 0.1100588  |
| 38 | 0.1030379  |
| 39 | 0.05657775 |
| 40 | 0.0776451  |
| 41 | 0.07252344 |
| 42 | 0.03922072 |
| 43 | 0.06326919 |
| 44 | 0.05903511 |
| 45 | 0.03171443 |

In [36]:

```
ggplot(prediction,aes(x=exp(pframe$ln_wage),y=prediction)) +geom_point()+xlab('wage')+ylab('whatever the DV
is')
```

In [ ]: