Valery Delgado and Safinaz Ali Final Project

2022-12-18

suppressMessages(library(stargazer))  
suppressMessages(library(coefplot))  
suppressMessages(library(AER))  
suppressMessages(library(dplyr))  
suppressMessages(library(ggplot2))  
suppressMessages(library(scales))  
suppressMessages(library(plyr))  
suppressMessages(library(table1))  
suppressMessages(library(coefplot))  
suppressMessages(library(readr))  
  
#cleaned data  
library(readr)  
suppressMessages(NHIS21\_CLEANED\_R <- read\_csv("NHIS21 CLEANED.R.csv"))  
suppressMessages(attach(NHIS21\_CLEANED\_R))  
use\_varb <- (age >=18) & (age <84) & (fulltime == 1)   
#work 35+ hours, 18-84 years  
data\_use <- subset(NHIS21\_CLEANED\_R,use\_varb)   
#The analytic sample included 13,256 adults ages 18–85 with an employment status of working 35 hours looking at injury   
attach(data\_use)

## The following objects are masked from NHIS21\_CLEANED\_R:  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek

## The following object is masked from package:survival:  
##   
## veteran

table1(~genderf + age+ white+ black+ asian+ aian+ medicaid + private+ uninsured+ single + married + divorced +educ\_bach +educ\_nohs + educ\_hs + educ\_as + educ\_smcoll + educ\_adv + injury +health+ Construction+ Retail+ Agriculture+Healthcare+Manufacturing+ Services+Transportation+Wholesale+ Mining, data = data\_use)

## Get nicer `table1` .docx output by simply installing the `flextable` package

##   Overall  
## 1 (N=13256)  
## 2 genderf   
## 3   Mean (SD) 0.465 (0.499)  
## 4   Median [Min, Max] 0 [0, 1.00]  
## 5 age   
## 6   Mean (SD) 44.0 (13.2)  
## 7   Median [Min, Max] 43.0 [18.0, 83.0]  
## 8 white   
## 9   Mean (SD) 0.730 (0.444)  
## 10   Median [Min, Max] 1.00 [0, 1.00]  
## 11 black   
## 12   Mean (SD) 0.110 (0.313)  
## 13   Median [Min, Max] 0 [0, 1.00]  
## 14 asian   
## 15   Mean (SD) 0.0729 (0.260)  
## 16   Median [Min, Max] 0 [0, 1.00]  
## 17 aian   
## 18   Mean (SD) 0.00686 (0.0826)  
## 19   Median [Min, Max] 0 [0, 1.00]  
## 20 medicaid   
## 21   Mean (SD) 0.0546 (0.227)  
## 22   Median [Min, Max] 0 [0, 1.00]  
## 23   Missing 734 (5.5%)  
## 24 private   
## 25   Mean (SD) 0.827 (0.379)  
## 26   Median [Min, Max] 1.00 [0, 1.00]  
## 27   Missing 734 (5.5%)  
## 28 uninsured   
## 29   Mean (SD) 0.0938 (0.292)  
## 30   Median [Min, Max] 0 [0, 1.00]  
## 31   Missing 734 (5.5%)  
## 32 single   
## 33   Mean (SD) 0.308 (0.462)  
## 34   Median [Min, Max] 0 [0, 1.00]  
## 35 married   
## 36   Mean (SD) 0.495 (0.500)  
## 37   Median [Min, Max] 0 [0, 1.00]  
## 38 divorced   
## 39   Mean (SD) 0.152 (0.359)  
## 40   Median [Min, Max] 0 [0, 1.00]  
## 41 educ\_bach   
## 42   Mean (SD) 0.293 (0.455)  
## 43   Median [Min, Max] 0 [0, 1.00]  
## 44 educ\_nohs   
## 45   Mean (SD) 0.0488 (0.215)  
## 46   Median [Min, Max] 0 [0, 1.00]  
## 47 educ\_hs   
## 48   Mean (SD) 0.210 (0.407)  
## 49   Median [Min, Max] 0 [0, 1.00]  
## 50 educ\_as   
## 51   Mean (SD) 0.124 (0.330)  
## 52   Median [Min, Max] 0 [0, 1.00]  
## 53 educ\_smcoll   
## 54   Mean (SD) 0.131 (0.337)  
## 55   Median [Min, Max] 0 [0, 1.00]  
## 56 educ\_adv   
## 57   Mean (SD) 0.189 (0.391)  
## 58   Median [Min, Max] 0 [0, 1.00]  
## 59 injury   
## 60   Mean (SD) 0.0918 (0.289)  
## 61   Median [Min, Max] 0 [0, 1.00]  
## 62 health   
## 63   Excellent 3641 (27.5%)  
## 64   Fair 798 (6.0%)  
## 65   Good 3486 (26.3%)  
## 66   Poor 88 (0.7%)  
## 67   Very good 5243 (39.6%)  
## 68 Construction   
## 69   Mean (SD) 0.0677 (0.251)  
## 70   Median [Min, Max] 0 [0, 1.00]  
## 71   Missing 18 (0.1%)  
## 72 Retail   
## 73   Mean (SD) 0.0803 (0.272)  
## 74   Median [Min, Max] 0 [0, 1.00]  
## 75   Missing 18 (0.1%)  
## 76 Agriculture   
## 77   Mean (SD) 0.0136 (0.116)  
## 78   Median [Min, Max] 0 [0, 1.00]  
## 79   Missing 18 (0.1%)  
## 80 Healthcare   
## 81   Mean (SD) 0.141 (0.348)  
## 82   Median [Min, Max] 0 [0, 1.00]  
## 83   Missing 18 (0.1%)  
## 84 Manufacturing   
## 85   Mean (SD) 0.107 (0.310)  
## 86   Median [Min, Max] 0 [0, 1.00]  
## 87   Missing 18 (0.1%)  
## 88 Services   
## 89   Mean (SD) 0.0370 (0.189)  
## 90   Median [Min, Max] 0 [0, 1.00]  
## 91   Missing 18 (0.1%)  
## 92 Transportation   
## 93   Mean (SD) 0.0462 (0.210)  
## 94   Median [Min, Max] 0 [0, 1.00]  
## 95   Missing 18 (0.1%)  
## 96 Wholesale   
## 97   Mean (SD) 0.0187 (0.135)  
## 98   Median [Min, Max] 0 [0, 1.00]  
## 99   Missing 18 (0.1%)  
## 100 Mining   
## 101   Mean (SD) 0.00423 (0.0649)  
## 102   Median [Min, Max] 0 [0, 1.00]  
## 103   Missing 18 (0.1%)

ddply(NHIS21\_CLEANED\_R, .(health), summarize, tmean = mean(injury), tsd = sd(injury), n\_obs = length(injury))

## health tmean tsd n\_obs  
## 1 Excellent 0.07210455 0.2586804 6657  
## 2 Fair 0.12376976 0.3293678 3353  
## 3 Good 0.09532934 0.2936869 8350  
## 4 Not Ascertained 0.00000000 0.0000000 7  
## 5 Poor 0.17330677 0.3787009 1004  
## 6 Refused 0.16666667 0.4082483 6  
## 7 Very good 0.08975755 0.2858482 10105

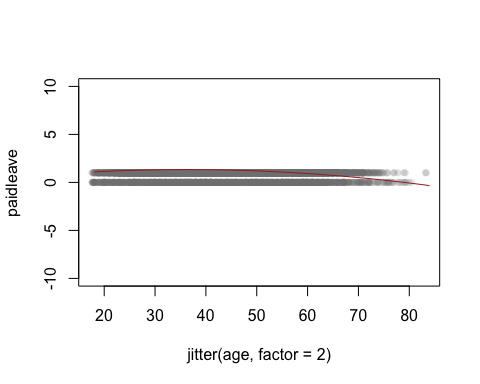
xtabs(~health+injury)

## injury  
## health 0 1  
## Excellent 3386 255  
## Fair 704 94  
## Good 3141 345  
## Poor 73 15  
## Very good 4735 508

model\_temp1 <- glm(paidleave ~ genderf + age+ I(age^2)+ white+black+asian+aian+medicaid + private+ uninsured+ single + married+divorced +educ\_bach+educ\_as+educ\_adv+ injury +health+Healthcare +Construction+Mining+Agriculture+Manufacturing+Services+Transportation+Wholesale+Retail, family = binomial, data = data\_use)  
summary(model\_temp1)

##   
## Call:  
## glm(formula = paidleave ~ genderf + age + I(age^2) + white +   
## black + asian + aian + medicaid + private + uninsured + single +   
## married + divorced + educ\_bach + educ\_as + educ\_adv + injury +   
## health + Healthcare + Construction + Mining + Agriculture +   
## Manufacturing + Services + Transportation + Wholesale + Retail,   
## family = binomial, data = data\_use)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4273 0.4100 0.4994 0.6284 1.8659   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.1477924 0.3618293 -0.408 0.682937   
## genderf 0.1531738 0.0519839 2.947 0.003213 \*\*   
## age 0.0532336 0.0149767 3.554 0.000379 \*\*\*  
## I(age^2) -0.0007336 0.0001709 -4.293 1.76e-05 \*\*\*  
## white -0.2927295 0.0865014 -3.384 0.000714 \*\*\*  
## black -0.1342672 0.1082144 -1.241 0.214698   
## asian -0.2582457 0.1261416 -2.047 0.040632 \*   
## aian 0.1168386 0.2817197 0.415 0.678337   
## medicaid -0.5110977 0.1495874 -3.417 0.000634 \*\*\*  
## private 0.8838685 0.1282881 6.890 5.59e-12 \*\*\*  
## uninsured -0.7461053 0.1405886 -5.307 1.11e-07 \*\*\*  
## single 0.2555152 0.1230867 2.076 0.037904 \*   
## married 0.2088695 0.1170848 1.784 0.074437 .   
## divorced 0.2750329 0.1264775 2.175 0.029663 \*   
## educ\_bach 0.4533304 0.0622054 7.288 3.15e-13 \*\*\*  
## educ\_as 0.0979224 0.0724052 1.352 0.176240   
## educ\_adv 0.6402051 0.0808539 7.918 2.41e-15 \*\*\*  
## injury 0.1076196 0.0833423 1.291 0.196601   
## healthFair -0.2808917 0.1010935 -2.779 0.005460 \*\*   
## healthGood -0.0534035 0.0657726 -0.812 0.416825   
## healthPoor -0.4087076 0.2651547 -1.541 0.123221   
## healthVery good 0.0133865 0.0603726 0.222 0.824523   
## Healthcare 0.0400140 0.0782215 0.512 0.608968   
## Construction -0.9077036 0.0865689 -10.485 < 2e-16 \*\*\*  
## Mining -0.1367253 0.3527168 -0.388 0.698286   
## Agriculture -1.2764513 0.1829583 -6.977 3.02e-12 \*\*\*  
## Manufacturing -0.0507092 0.0818070 -0.620 0.535348   
## Services -0.8404451 0.1100200 -7.639 2.19e-14 \*\*\*  
## Transportation -0.5520187 0.1031160 -5.353 8.63e-08 \*\*\*  
## Wholesale 0.2751034 0.1951427 1.410 0.158612   
## Retail -0.0769427 0.0890014 -0.865 0.387307   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 12959 on 12504 degrees of freedom  
## Residual deviance: 11357 on 12474 degrees of freedom  
## (751 observations deleted due to missingness)  
## AIC: 11419  
##   
## Number of Fisher Scoring iterations: 4

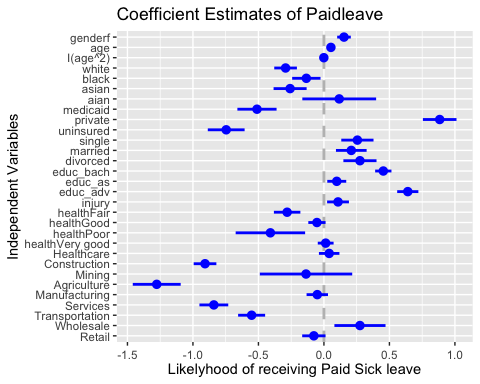
suppressMessages(require(AER))  
NNobs <- length(paidleave)  
set.seed(12345)   
graph\_obs <-(runif(NNobs) < 0.5)   
dat\_graph <-subset(data\_use,graph\_obs)   
plot(paidleave ~ jitter(age, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(-10,10), data = dat\_graph)  
to\_be\_predicted1 <- data.frame(age = 18:84,genderf =1,white=1,black=0,aian=0,asian=0, medicaid=0, private=1, uninsured=0, single=0, married=1,divorced=0,educ\_as=0, educ\_bach=1,educ\_adv=0, injury=1, health="Excellent",Healthcare=1,Construction=1,Mining=0,Agriculture=0,Manufacturing=1,Services=0,Transportation=0,Wholesale=0,Retail=1)  
to\_be\_predicted1$yhat <- predict(model\_temp1, newdata = to\_be\_predicted1)  
lines(yhat ~ age, data = to\_be\_predicted1, col = "brown")



summary(to\_be\_predicted1$yhat)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.3336 0.6217 1.1317 0.9035 1.2837 1.3367

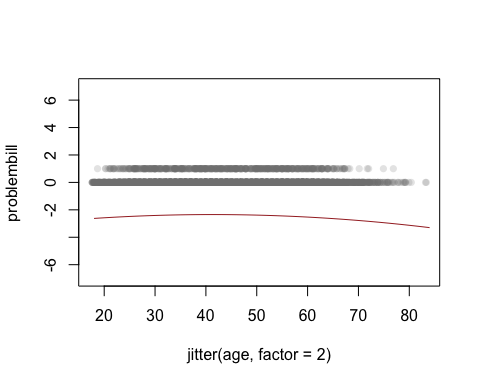
coefplot(model\_temp1, innerCI = 1, outerCI = 0, intercept = FALSE, title = "Coefficient Estimates of Paidleave ",  
 ylab = "Independent Variables", xlab = "Likelyhood of receiving Paid Sick leave", decreasing = TRUE)



model\_temp2 <- glm(problembill ~ genderf + age+I(age^2)+ white+black+asian+aian+medicaid + private+ uninsured+ single + married+divorced +educ\_bach+educ\_as+educ\_adv+ injury +health+Healthcare +Construction+Mining+Agriculture+Manufacturing+Services+Transportation+Wholesale+Retail, family = binomial, data = data\_use)  
summary(model\_temp2)

##   
## Call:  
## glm(formula = problembill ~ genderf + age + I(age^2) + white +   
## black + asian + aian + medicaid + private + uninsured + single +   
## married + divorced + educ\_bach + educ\_as + educ\_adv + injury +   
## health + Healthcare + Construction + Mining + Agriculture +   
## Manufacturing + Services + Transportation + Wholesale + Retail,   
## family = binomial, data = data\_use)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4108 -0.4628 -0.3509 -0.2652 2.9428   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.8839383 0.5227050 -7.430 1.08e-13 \*\*\*  
## genderf 0.4182275 0.0710173 5.889 3.88e-09 \*\*\*  
## age 0.0429722 0.0212303 2.024 0.042960 \*   
## I(age^2) -0.0005210 0.0002427 -2.147 0.031824 \*   
## white 0.0153071 0.1123252 0.136 0.891604   
## black 0.4919605 0.1323863 3.716 0.000202 \*\*\*  
## asian -0.3480274 0.1910842 -1.821 0.068557 .   
## aian 0.2105659 0.3243681 0.649 0.516237   
## medicaid 0.1493172 0.2454946 0.608 0.543035   
## private 0.0880947 0.2160778 0.408 0.683495   
## uninsured 0.8029643 0.2284572 3.515 0.000440 \*\*\*  
## single -0.1060525 0.1623555 -0.653 0.513620   
## married -0.0831306 0.1558875 -0.533 0.593845   
## divorced 0.0849720 0.1646845 0.516 0.605876   
## educ\_bach -0.6070462 0.0879178 -6.905 5.03e-12 \*\*\*  
## educ\_as -0.0975195 0.0946328 -1.031 0.302773   
## educ\_adv -0.8749575 0.1187607 -7.367 1.74e-13 \*\*\*  
## injury 0.5086780 0.0965013 5.271 1.36e-07 \*\*\*  
## healthFair 1.5137327 0.1215639 12.452 < 2e-16 \*\*\*  
## healthGood 0.8776940 0.0976895 8.985 < 2e-16 \*\*\*  
## healthPoor 2.4801707 0.2463920 10.066 < 2e-16 \*\*\*  
## healthVery good 0.3418497 0.0979448 3.490 0.000483 \*\*\*  
## Healthcare 0.0688635 0.0971829 0.709 0.478574   
## Construction 0.1023801 0.1304630 0.785 0.432603   
## Mining -0.3384852 0.6075101 -0.557 0.577413   
## Agriculture -0.1429278 0.2959367 -0.483 0.629119   
## Manufacturing 0.1002600 0.1089073 0.921 0.357260   
## Services 0.1328438 0.1625803 0.817 0.413873   
## Transportation 0.0670521 0.1504259 0.446 0.655779   
## Wholesale 0.0919345 0.2402532 0.383 0.701974   
## Retail 0.0410914 0.1210342 0.340 0.734231   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 7674.9 on 12504 degrees of freedom  
## Residual deviance: 6980.2 on 12474 degrees of freedom  
## (751 observations deleted due to missingness)  
## AIC: 7042.2  
##   
## Number of Fisher Scoring iterations: 6

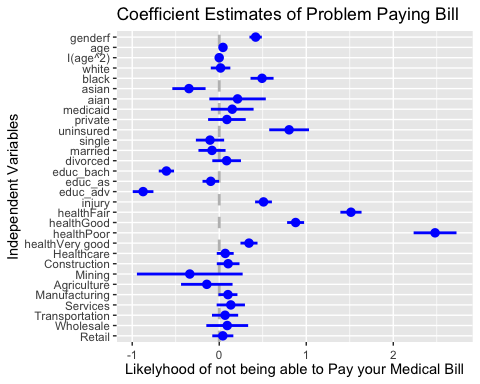
suppressMessages(require(AER))  
NNobs <- length(problembill)  
set.seed(12345)   
graph\_obs <- (runif(NNobs) < 0.5)   
dat\_graph <-subset(data\_use,graph\_obs)   
plot(problembill ~ jitter(age, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(-7,7), data = dat\_graph)  
to\_be\_predicted2 <- data.frame(age = 18:84,genderf =1,white=1,black=0,aian=0,asian=0, medicaid=0, private=1, uninsured=0, single=0, married=1,divorced=0,educ\_as=0, educ\_bach=1,educ\_adv=0, injury=1, health="Excellent",Healthcare=1,Construction=1,Mining=0,Agriculture=0,Manufacturing=1,Services=0,Transportation=0,Wholesale=0,Retail=1)  
to\_be\_predicted2$yhat <- predict(model\_temp2, newdata = to\_be\_predicted2)  
lines(yhat ~ age, data = to\_be\_predicted2, col = "brown")



summary(to\_be\_predicted2$yhat)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -3.298 -2.705 -2.492 -2.590 -2.383 -2.345

coefplot(model\_temp2, innerCI = 1, outerCI = 0, intercept = FALSE, title = "Coefficient Estimates of Problem Paying Bill ",  
 ylab = "Independent Variables", xlab = "Likelyhood of not being able to Pay your Medical Bill", decreasing = TRUE)



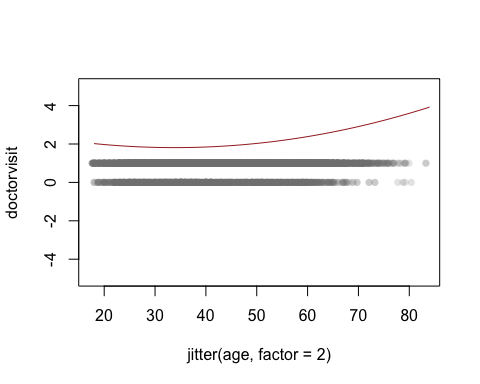
suppressMessages(require(stargazer))  
suppressWarnings(stargazer(model\_temp1,model\_temp2, type = "text"))

##   
## ==============================================  
## Dependent variable:   
## ----------------------------  
## paidleave problembill   
## (1) (2)   
## ----------------------------------------------  
## genderf 0.153\*\*\* 0.418\*\*\*   
## (0.052) (0.071)   
##   
## age 0.053\*\*\* 0.043\*\*   
## (0.015) (0.021)   
##   
## I(age2) -0.001\*\*\* -0.001\*\*   
## (0.0002) (0.0002)   
##   
## white -0.293\*\*\* 0.015   
## (0.087) (0.112)   
##   
## black -0.134 0.492\*\*\*   
## (0.108) (0.132)   
##   
## asian -0.258\*\* -0.348\*   
## (0.126) (0.191)   
##   
## aian 0.117 0.211   
## (0.282) (0.324)   
##   
## medicaid -0.511\*\*\* 0.149   
## (0.150) (0.245)   
##   
## private 0.884\*\*\* 0.088   
## (0.128) (0.216)   
##   
## uninsured -0.746\*\*\* 0.803\*\*\*   
## (0.141) (0.228)   
##   
## single 0.256\*\* -0.106   
## (0.123) (0.162)   
##   
## married 0.209\* -0.083   
## (0.117) (0.156)   
##   
## divorced 0.275\*\* 0.085   
## (0.126) (0.165)   
##   
## educ\_bach 0.453\*\*\* -0.607\*\*\*   
## (0.062) (0.088)   
##   
## educ\_as 0.098 -0.098   
## (0.072) (0.095)   
##   
## educ\_adv 0.640\*\*\* -0.875\*\*\*   
## (0.081) (0.119)   
##   
## injury 0.108 0.509\*\*\*   
## (0.083) (0.097)   
##   
## healthFair -0.281\*\*\* 1.514\*\*\*   
## (0.101) (0.122)   
##   
## healthGood -0.053 0.878\*\*\*   
## (0.066) (0.098)   
##   
## healthPoor -0.409 2.480\*\*\*   
## (0.265) (0.246)   
##   
## healthVery good 0.013 0.342\*\*\*   
## (0.060) (0.098)   
##   
## Healthcare 0.040 0.069   
## (0.078) (0.097)   
##   
## Construction -0.908\*\*\* 0.102   
## (0.087) (0.130)   
##   
## Mining -0.137 -0.338   
## (0.353) (0.608)   
##   
## Agriculture -1.276\*\*\* -0.143   
## (0.183) (0.296)   
##   
## Manufacturing -0.051 0.100   
## (0.082) (0.109)   
##   
## Services -0.840\*\*\* 0.133   
## (0.110) (0.163)   
##   
## Transportation -0.552\*\*\* 0.067   
## (0.103) (0.150)   
##   
## Wholesale 0.275 0.092   
## (0.195) (0.240)   
##   
## Retail -0.077 0.041   
## (0.089) (0.121)   
##   
## Constant -0.148 -3.884\*\*\*   
## (0.362) (0.523)   
##   
## ----------------------------------------------  
## Observations 12,505 12,505   
## Log Likelihood -5,678.265 -3,490.087   
## Akaike Inf. Crit. 11,418.530 7,042.174   
## ==============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

model\_temp3 <- glm(doctorvisit ~ genderf + age+I(age^2)+white+black+asian+aian+medicaid + private+ uninsured+ single + married+divorced +educ\_bach+educ\_as+educ\_adv+ injury +health+Healthcare +Construction+Mining+Agriculture+Manufacturing+Services+Transportation+Wholesale+Retail, family = binomial, data = data\_use)  
summary(model\_temp3)

##   
## Call:  
## glm(formula = doctorvisit ~ genderf + age + I(age^2) + white +   
## black + asian + aian + medicaid + private + uninsured + single +   
## married + divorced + educ\_bach + educ\_as + educ\_adv + injury +   
## health + Healthcare + Construction + Mining + Agriculture +   
## Manufacturing + Services + Transportation + Wholesale + Retail,   
## family = binomial, data = data\_use)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5290 0.3662 0.5422 0.7107 1.6592   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.6201706 0.3758311 4.311 1.63e-05 \*\*\*  
## genderf 0.7800511 0.0510102 15.292 < 2e-16 \*\*\*  
## age -0.0567515 0.0150856 -3.762 0.000169 \*\*\*  
## I(age^2) 0.0008386 0.0001756 4.775 1.80e-06 \*\*\*  
## white 0.0258739 0.0816562 0.317 0.751347   
## black 0.4668354 0.1096869 4.256 2.08e-05 \*\*\*  
## asian -0.5156672 0.1094836 -4.710 2.48e-06 \*\*\*  
## aian 0.7715989 0.3183687 2.424 0.015367 \*   
## medicaid -0.3622446 0.1954077 -1.854 0.063769 .   
## private -0.3639163 0.1685564 -2.159 0.030849 \*   
## uninsured -1.6704475 0.1782671 -9.370 < 2e-16 \*\*\*  
## single 0.1205399 0.1280657 0.941 0.346584   
## married 0.4682965 0.1240040 3.776 0.000159 \*\*\*  
## divorced 0.1864029 0.1331073 1.400 0.161395   
## educ\_bach 0.1250293 0.0593960 2.105 0.035290 \*   
## educ\_as 0.1347821 0.0764624 1.763 0.077947 .   
## educ\_adv 0.2922773 0.0736266 3.970 7.20e-05 \*\*\*  
## injury 0.1877267 0.0816454 2.299 0.021488 \*   
## healthFair 0.7136220 0.1177642 6.060 1.36e-09 \*\*\*  
## healthGood 0.4021908 0.0639409 6.290 3.17e-10 \*\*\*  
## healthPoor 1.7482070 0.4423136 3.952 7.74e-05 \*\*\*  
## healthVery good 0.1310297 0.0547712 2.392 0.016743 \*   
## Healthcare 0.1921298 0.0789452 2.434 0.014945 \*   
## Construction -0.1737248 0.0884560 -1.964 0.049534 \*   
## Mining 0.1047917 0.3335836 0.314 0.753415   
## Agriculture 0.2077379 0.2036436 1.020 0.307678   
## Manufacturing -0.0159601 0.0763754 -0.209 0.834472   
## Services 0.0448418 0.1236279 0.363 0.716817   
## Transportation 0.0587479 0.1100259 0.534 0.593379   
## Wholesale 0.2906004 0.1795225 1.619 0.105503   
## Retail -0.0739911 0.0850622 -0.870 0.384384   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 13106 on 12504 degrees of freedom  
## Residual deviance: 11932 on 12474 degrees of freedom  
## (751 observations deleted due to missingness)  
## AIC: 11994  
##   
## Number of Fisher Scoring iterations: 5

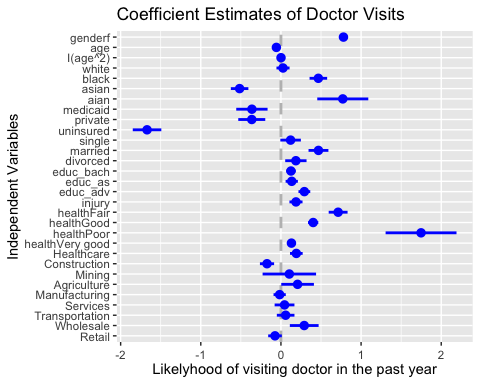
suppressMessages(require(AER))  
NNobs <- length(doctorvisit)  
set.seed(12345)   
graph\_obs <- (runif(NNobs) < 0.5)   
dat\_graph <-subset(data\_use,graph\_obs)   
plot(doctorvisit ~ jitter(age, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(-5,5), data = dat\_graph)  
to\_be\_predicted3 <- data.frame(age = 18:84,genderf =1,white=1,black=0,aian=0,asian=0, medicaid=0, private=1, uninsured=0, single=0, married=1,divorced=0,educ\_as=0, educ\_bach=1,educ\_adv=0, injury=1, health="Excellent",Healthcare=1,Construction=1,Mining=0,Agriculture=0,Manufacturing=1,Services=0,Transportation=0,Wholesale=0,Retail=1)  
to\_be\_predicted3$yhat <- predict(model\_temp3, newdata = to\_be\_predicted3)  
lines(yhat ~ age, data = to\_be\_predicted3, col = "brown")



summary(to\_be\_predicted3$yhat)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.812 1.872 2.058 2.372 2.762 3.922

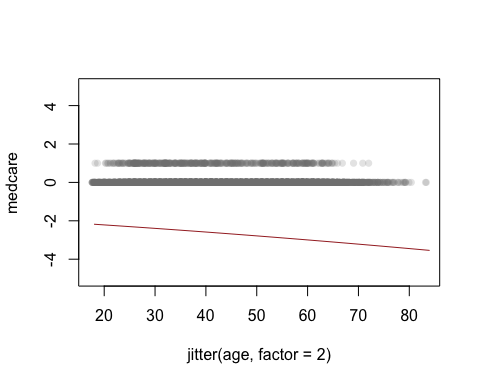
coefplot(model\_temp3, innerCI = 1, outerCI = 0, intercept = FALSE, title = "Coefficient Estimates of Doctor Visits ",  
 ylab = "Independent Variables", xlab = "Likelyhood of visiting doctor in the past year", decreasing = TRUE)



model\_temp4 <- glm(medcare ~ genderf + age+ I(age^2)+ white+black+asian+aian+medicaid + private+ uninsured+ single + married+divorced +educ\_bach+educ\_as+educ\_adv+ injury +health+Healthcare +Construction+Mining+Agriculture+Manufacturing+Services+Transportation+Wholesale+Retail, family = binomial, data = data\_use)  
summary(model\_temp4)

##   
## Call:  
## glm(formula = medcare ~ genderf + age + I(age^2) + white + black +   
## asian + aian + medicaid + private + uninsured + single +   
## married + divorced + educ\_bach + educ\_as + educ\_adv + injury +   
## health + Healthcare + Construction + Mining + Agriculture +   
## Manufacturing + Services + Transportation + Wholesale + Retail,   
## family = binomial, data = data\_use)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4105 -0.4176 -0.3199 -0.2454 2.9486   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.125e+00 5.827e-01 -5.363 8.18e-08 \*\*\*  
## genderf 3.677e-01 7.626e-02 4.822 1.42e-06 \*\*\*  
## age -1.561e-02 2.251e-02 -0.694 0.487981   
## I(age^2) -4.936e-05 2.612e-04 -0.189 0.850117   
## white 1.828e-01 1.209e-01 1.512 0.130547   
## black -2.639e-01 1.611e-01 -1.638 0.101402   
## asian -6.716e-01 2.245e-01 -2.992 0.002772 \*\*   
## aian -3.296e-01 3.924e-01 -0.840 0.400900   
## medicaid 3.666e-02 3.484e-01 0.105 0.916217   
## private 6.046e-01 3.017e-01 2.004 0.045062 \*   
## uninsured 1.887e+00 3.096e-01 6.095 1.10e-09 \*\*\*  
## single -3.389e-01 1.749e-01 -1.938 0.052632 .   
## married -5.392e-01 1.693e-01 -3.184 0.001451 \*\*   
## divorced 1.582e-01 1.751e-01 0.903 0.366330   
## educ\_bach 1.133e-01 9.045e-02 1.253 0.210294   
## educ\_as 1.636e-01 1.072e-01 1.526 0.126939   
## educ\_adv -1.247e-01 1.196e-01 -1.043 0.296964   
## injury 6.037e-01 9.878e-02 6.112 9.86e-10 \*\*\*  
## healthFair 1.502e+00 1.340e-01 11.209 < 2e-16 \*\*\*  
## healthGood 9.662e-01 1.037e-01 9.315 < 2e-16 \*\*\*  
## healthPoor 2.068e+00 2.757e-01 7.501 6.33e-14 \*\*\*  
## healthVery good 3.628e-01 1.022e-01 3.550 0.000385 \*\*\*  
## Healthcare 1.314e-01 1.025e-01 1.281 0.200169   
## Construction -9.960e-02 1.449e-01 -0.687 0.491820   
## Mining -1.351e-01 6.121e-01 -0.221 0.825275   
## Agriculture -6.130e-02 3.087e-01 -0.199 0.842576   
## Manufacturing -8.693e-02 1.251e-01 -0.695 0.487015   
## Services 6.440e-02 1.750e-01 0.368 0.712951   
## Transportation -1.118e-01 1.815e-01 -0.616 0.538047   
## Wholesale -3.545e-02 2.762e-01 -0.128 0.897870   
## Retail -3.443e-02 1.331e-01 -0.259 0.795902   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 6812.6 on 12504 degrees of freedom  
## Residual deviance: 6172.0 on 12474 degrees of freedom  
## (751 observations deleted due to missingness)  
## AIC: 6234  
##   
## Number of Fisher Scoring iterations: 6

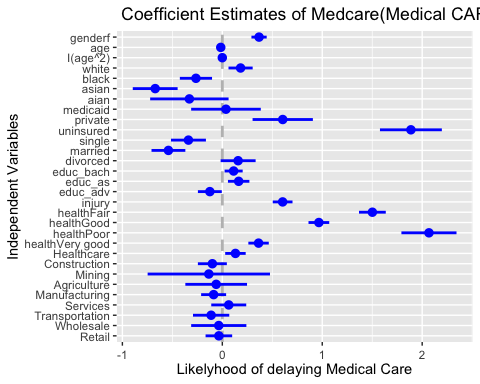
suppressMessages(require(AER))  
NNobs <- length(medcare)  
set.seed(12345)   
graph\_obs <- (runif(NNobs) < 0.5)   
dat\_graph <-subset(data\_use,graph\_obs)   
plot(medcare ~ jitter(age, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(-5,5), data = dat\_graph)  
to\_be\_predicted4 <- data.frame(age = 18:84,genderf =1,white=1,black=0,aian=0,asian=0, medicaid=0, private=1, uninsured=0, single=0, married=1,divorced=0,educ\_as=0, educ\_bach=1,educ\_adv=0, injury=1, health="Excellent",Healthcare=1,Construction=1,Mining=0,Agriculture=0,Manufacturing=1,Services=0,Transportation=0,Wholesale=0,Retail=1)  
to\_be\_predicted4$yhat <- predict(model\_temp4, newdata = to\_be\_predicted4)  
lines(yhat ~ age, data = to\_be\_predicted4, col = "brown")



summary(to\_be\_predicted4$yhat)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -3.541 -3.160 -2.806 -2.824 -2.479 -2.179

coefplot(model\_temp4, innerCI = 1, outerCI = 0, intercept = FALSE, title = " Coefficient Estimates of Medcare(Medical CARE) ",  
 ylab = "Independent Variables", xlab = "Likelyhood of delaying Medical Care", decreasing = TRUE)



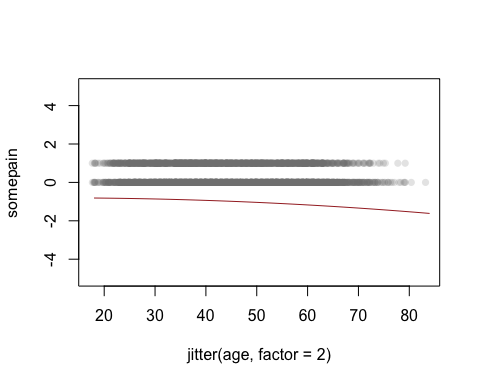
suppressMessages(require(stargazer))  
suppressWarnings(stargazer(model\_temp3,model\_temp4, type = "text"))

##   
## ==============================================  
## Dependent variable:   
## ----------------------------  
## doctorvisit medcare   
## (1) (2)   
## ----------------------------------------------  
## genderf 0.780\*\*\* 0.368\*\*\*   
## (0.051) (0.076)   
##   
## age -0.057\*\*\* -0.016   
## (0.015) (0.023)   
##   
## I(age2) 0.001\*\*\* -0.00005   
## (0.0002) (0.0003)   
##   
## white 0.026 0.183   
## (0.082) (0.121)   
##   
## black 0.467\*\*\* -0.264   
## (0.110) (0.161)   
##   
## asian -0.516\*\*\* -0.672\*\*\*   
## (0.109) (0.224)   
##   
## aian 0.772\*\* -0.330   
## (0.318) (0.392)   
##   
## medicaid -0.362\* 0.037   
## (0.195) (0.348)   
##   
## private -0.364\*\* 0.605\*\*   
## (0.169) (0.302)   
##   
## uninsured -1.670\*\*\* 1.887\*\*\*   
## (0.178) (0.310)   
##   
## single 0.121 -0.339\*   
## (0.128) (0.175)   
##   
## married 0.468\*\*\* -0.539\*\*\*   
## (0.124) (0.169)   
##   
## divorced 0.186 0.158   
## (0.133) (0.175)   
##   
## educ\_bach 0.125\*\* 0.113   
## (0.059) (0.090)   
##   
## educ\_as 0.135\* 0.164   
## (0.076) (0.107)   
##   
## educ\_adv 0.292\*\*\* -0.125   
## (0.074) (0.120)   
##   
## injury 0.188\*\* 0.604\*\*\*   
## (0.082) (0.099)   
##   
## healthFair 0.714\*\*\* 1.502\*\*\*   
## (0.118) (0.134)   
##   
## healthGood 0.402\*\*\* 0.966\*\*\*   
## (0.064) (0.104)   
##   
## healthPoor 1.748\*\*\* 2.068\*\*\*   
## (0.442) (0.276)   
##   
## healthVery good 0.131\*\* 0.363\*\*\*   
## (0.055) (0.102)   
##   
## Healthcare 0.192\*\* 0.131   
## (0.079) (0.103)   
##   
## Construction -0.174\*\* -0.100   
## (0.088) (0.145)   
##   
## Mining 0.105 -0.135   
## (0.334) (0.612)   
##   
## Agriculture 0.208 -0.061   
## (0.204) (0.309)   
##   
## Manufacturing -0.016 -0.087   
## (0.076) (0.125)   
##   
## Services 0.045 0.064   
## (0.124) (0.175)   
##   
## Transportation 0.059 -0.112   
## (0.110) (0.182)   
##   
## Wholesale 0.291 -0.035   
## (0.180) (0.276)   
##   
## Retail -0.074 -0.034   
## (0.085) (0.133)   
##   
## Constant 1.620\*\*\* -3.125\*\*\*   
## (0.376) (0.583)   
##   
## ----------------------------------------------  
## Observations 12,505 12,505   
## Log Likelihood -5,965.842 -3,085.976   
## Akaike Inf. Crit. 11,993.680 6,233.951   
## ==============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

model\_temp5 <- glm(somepain ~ genderf + age+I(age^2)+ white+black+asian+aian+medicaid + private+ uninsured+ single + married+divorced +educ\_bach+educ\_as+educ\_adv+ injury +health+Healthcare +Construction+Mining+Agriculture+Manufacturing+Services+Transportation+Wholesale+Retail, family = binomial, data = data\_use)  
summary(model\_temp5)

##   
## Call:  
## glm(formula = somepain ~ genderf + age + I(age^2) + white + black +   
## asian + aian + medicaid + private + uninsured + single +   
## married + divorced + educ\_bach + educ\_as + educ\_adv + injury +   
## health + Healthcare + Construction + Mining + Agriculture +   
## Manufacturing + Services + Transportation + Wholesale + Retail,   
## family = binomial, data = data\_use)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4941 -0.9015 -0.7573 1.3137 1.9785   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.1501901 0.4298324 -2.676 0.00745 \*\*   
## genderf 0.3641490 0.0555347 6.557 5.48e-11 \*\*\*  
## age 0.0030761 0.0174609 0.176 0.86016   
## I(age^2) -0.0001493 0.0001982 -0.753 0.45132   
## white 0.0555494 0.1023721 0.543 0.58739   
## black -0.2428693 0.1279016 -1.899 0.05758 .   
## asian 0.2222395 0.1437228 1.546 0.12203   
## aian 0.4028503 0.2915632 1.382 0.16707   
## medicaid -0.2735211 0.1878797 -1.456 0.14544   
## private -0.2961577 0.1531460 -1.934 0.05313 .   
## uninsured -0.4477947 0.1767730 -2.533 0.01130 \*   
## single 0.2522241 0.1502344 1.679 0.09318 .   
## married 0.2254141 0.1446884 1.558 0.11925   
## divorced 0.2920340 0.1524060 1.916 0.05534 .   
## educ\_bach 0.1176647 0.0668248 1.761 0.07827 .   
## educ\_as -0.0147365 0.0825522 -0.179 0.85832   
## educ\_adv 0.0408471 0.0799061 0.511 0.60922   
## injury 0.5233584 0.0727375 7.195 6.24e-13 \*\*\*  
## healthFair 0.9108910 0.1087234 8.378 < 2e-16 \*\*\*  
## healthGood 0.6643499 0.0778647 8.532 < 2e-16 \*\*\*  
## healthPoor 0.4579862 0.2673679 1.713 0.08672 .   
## healthVery good 0.3082113 0.0741573 4.156 3.24e-05 \*\*\*  
## Healthcare -0.0696130 0.0773455 -0.900 0.36811   
## Construction -0.1027953 0.1117032 -0.920 0.35744   
## Mining 0.2599902 0.3722012 0.699 0.48485   
## Agriculture 0.3701076 0.2318968 1.596 0.11049   
## Manufacturing -0.2702967 0.0895294 -3.019 0.00254 \*\*   
## Services -0.1147273 0.1435952 -0.799 0.42431   
## Transportation -0.2528217 0.1338799 -1.888 0.05897 .   
## Wholesale -0.1651189 0.2017203 -0.819 0.41304   
## Retail -0.2179506 0.1011171 -2.155 0.03113 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9123.2 on 7211 degrees of freedom  
## Residual deviance: 8839.3 on 7181 degrees of freedom  
## (6044 observations deleted due to missingness)  
## AIC: 8901.3  
##   
## Number of Fisher Scoring iterations: 4

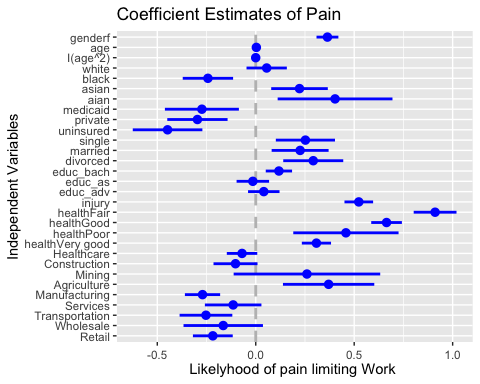
suppressMessages(require(AER))  
NNobs <- length(somepain)  
set.seed(12345)   
graph\_obs <- (runif(NNobs) < 0.5)   
dat\_graph <-subset(data\_use,graph\_obs)   
plot(somepain ~ jitter(age, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(-5,5), data = dat\_graph)  
to\_be\_predicted5 <- data.frame(age = 18:84,genderf =1,white=1,black=0,aian=0,asian=0, medicaid=0, private=1, uninsured=0, single=0, married=1,divorced=0,educ\_as=0, educ\_bach=1,educ\_adv=0, injury=1, health="Excellent",Healthcare=1,Construction=1,Mining=0,Agriculture=0,Manufacturing=1,Services=0,Transportation=0,Wholesale=0,Retail=1)  
to\_be\_predicted5$yhat <- predict(model\_temp5, newdata = to\_be\_predicted5)  
lines(yhat ~ age, data = to\_be\_predicted5, col = "brown")



summary(to\_be\_predicted5$yhat)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -1.6157 -1.2933 -1.0522 -1.1080 -0.8924 -0.8139

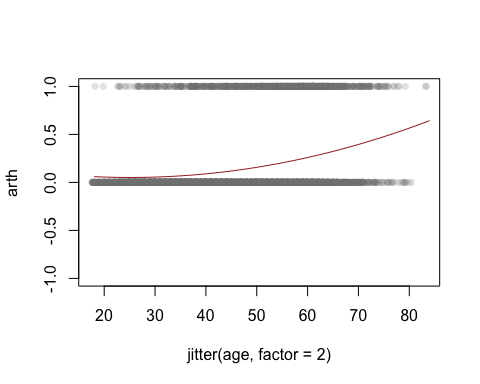
coefplot(model\_temp5, innerCI = 1, outerCI = 0, intercept = FALSE, title = "Coefficient Estimates of Pain ",  
 ylab = "Independent Variables", xlab = "Likelyhood of pain limiting Work ", decreasing = TRUE)



#OLS REGRESSION CLEANED DATA  
model\_temp6 <- lm(arth ~ genderf + age+ I(age^2)+ white+black+asian+aian+medicaid + private+ uninsured+ single + married+divorced +educ\_bach+educ\_as+educ\_adv+ injury +health+Healthcare +Construction+Mining+Agriculture+Manufacturing+Services+Transportation+Wholesale+Retail, data = data\_use)  
summary(model\_temp6)

##   
## Call:  
## lm(formula = arth ~ genderf + age + I(age^2) + white + black +   
## asian + aian + medicaid + private + uninsured + single +   
## married + divorced + educ\_bach + educ\_as + educ\_adv + injury +   
## health + Healthcare + Construction + Mining + Agriculture +   
## Manufacturing + Services + Transportation + Wholesale + Retail,   
## data = data\_use)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.60145 -0.15874 -0.06841 -0.00091 1.08905   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.448e-01 4.460e-02 3.247 0.001170 \*\*   
## genderf 2.770e-02 5.896e-03 4.699 2.65e-06 \*\*\*  
## age -8.544e-03 1.807e-03 -4.728 2.29e-06 \*\*\*  
## I(age^2) 1.704e-04 2.066e-05 8.248 < 2e-16 \*\*\*  
## white 4.054e-02 1.026e-02 3.950 7.86e-05 \*\*\*  
## black 2.118e-02 1.275e-02 1.661 0.096729 .   
## asian -1.155e-02 1.412e-02 -0.818 0.413284   
## aian 8.033e-02 3.431e-02 2.341 0.019222 \*   
## medicaid -6.746e-02 2.100e-02 -3.213 0.001318 \*\*   
## private -7.892e-02 1.755e-02 -4.496 6.97e-06 \*\*\*  
## uninsured -1.106e-01 1.962e-02 -5.638 1.76e-08 \*\*\*  
## single -2.118e-02 1.553e-02 -1.364 0.172498   
## married -1.680e-02 1.490e-02 -1.127 0.259648   
## divorced -1.230e-02 1.591e-02 -0.773 0.439672   
## educ\_bach -1.922e-02 7.149e-03 -2.689 0.007181 \*\*   
## educ\_as 1.958e-02 8.991e-03 2.177 0.029462 \*   
## educ\_adv -1.131e-02 8.482e-03 -1.334 0.182353   
## injury 3.335e-02 9.441e-03 3.533 0.000412 \*\*\*  
## healthFair 2.214e-01 1.261e-02 17.564 < 2e-16 \*\*\*  
## healthGood 1.025e-01 7.627e-03 13.438 < 2e-16 \*\*\*  
## healthPoor 2.839e-01 3.442e-02 8.248 < 2e-16 \*\*\*  
## healthVery good 4.126e-02 6.793e-03 6.074 1.28e-09 \*\*\*  
## Healthcare 6.424e-03 8.455e-03 0.760 0.447392   
## Construction 1.102e-02 1.172e-02 0.941 0.346879   
## Mining -5.454e-02 4.182e-02 -1.304 0.192197   
## Agriculture 7.474e-04 2.556e-02 0.029 0.976668   
## Manufacturing 7.676e-03 9.361e-03 0.820 0.412228   
## Services -1.019e-02 1.481e-02 -0.688 0.491659   
## Transportation -2.156e-02 1.350e-02 -1.597 0.110340   
## Wholesale -7.528e-03 2.075e-02 -0.363 0.716704   
## Retail 1.887e-03 1.061e-02 0.178 0.858870   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3049 on 12474 degrees of freedom  
## (751 observations deleted due to missingness)  
## Multiple R-squared: 0.118, Adjusted R-squared: 0.1159   
## F-statistic: 55.62 on 30 and 12474 DF, p-value: < 2.2e-16

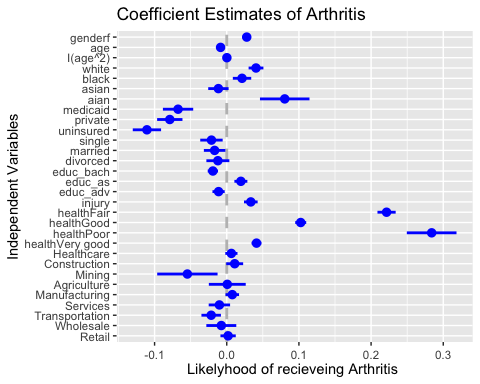
suppressMessages(require(AER))  
NNobs <- length(arth)  
set.seed(12345)   
graph\_obs <- (runif(NNobs) < 0.5)   
dat\_graph <-subset(data\_use,graph\_obs)   
plot(arth~ jitter(age, factor = 2), pch = 16, col = rgb(0.5, 0.5, 0.5, alpha = 0.2), ylim = c(-1,1), data = dat\_graph)  
to\_be\_predicted6 <- data.frame(age = 18:84,genderf =1,white=1,black=0,aian=0,asian=0, medicaid=0, private=1, uninsured=0, single=0, married=1,divorced=0,educ\_as=0, educ\_bach=1,educ\_adv=0, injury=1, health="Excellent",Healthcare=1,Construction=1,Mining=0,Agriculture=0,Manufacturing=1,Services=0,Transportation=0,Wholesale=0,Retail=1)  
to\_be\_predicted6$yhat <- predict(model\_temp6, newdata = to\_be\_predicted6)  
lines(yhat ~ age, data = to\_be\_predicted6, col = "brown")



summary(to\_be\_predicted6$yhat)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.05136 0.06654 0.16588 0.22960 0.35807 0.64293

coefplot(model\_temp6, innerCI = 1, outerCI = 0, intercept = FALSE, title = "Coefficient Estimates of Arthritis ",  
 ylab = "Independent Variables", xlab = "Likelyhood of recieveing Arthritis ", decreasing = TRUE)



suppressMessages(require(stargazer))  
suppressWarnings(stargazer(model\_temp5,model\_temp6, type = "text"))

##   
## =========================================================  
## Dependent variable:   
## -------------------------------------  
## somepain arth   
## logistic OLS   
## (1) (2)   
## ---------------------------------------------------------  
## genderf 0.364\*\*\* 0.028\*\*\*   
## (0.056) (0.006)   
##   
## age 0.003 -0.009\*\*\*   
## (0.017) (0.002)   
##   
## I(age2) -0.0001 0.0002\*\*\*   
## (0.0002) (0.00002)   
##   
## white 0.056 0.041\*\*\*   
## (0.102) (0.010)   
##   
## black -0.243\* 0.021\*   
## (0.128) (0.013)   
##   
## asian 0.222 -0.012   
## (0.144) (0.014)   
##   
## aian 0.403 0.080\*\*   
## (0.292) (0.034)   
##   
## medicaid -0.274 -0.067\*\*\*   
## (0.188) (0.021)   
##   
## private -0.296\* -0.079\*\*\*   
## (0.153) (0.018)   
##   
## uninsured -0.448\*\* -0.111\*\*\*   
## (0.177) (0.020)   
##   
## single 0.252\* -0.021   
## (0.150) (0.016)   
##   
## married 0.225 -0.017   
## (0.145) (0.015)   
##   
## divorced 0.292\* -0.012   
## (0.152) (0.016)   
##   
## educ\_bach 0.118\* -0.019\*\*\*   
## (0.067) (0.007)   
##   
## educ\_as -0.015 0.020\*\*   
## (0.083) (0.009)   
##   
## educ\_adv 0.041 -0.011   
## (0.080) (0.008)   
##   
## injury 0.523\*\*\* 0.033\*\*\*   
## (0.073) (0.009)   
##   
## healthFair 0.911\*\*\* 0.221\*\*\*   
## (0.109) (0.013)   
##   
## healthGood 0.664\*\*\* 0.102\*\*\*   
## (0.078) (0.008)   
##   
## healthPoor 0.458\* 0.284\*\*\*   
## (0.267) (0.034)   
##   
## healthVery good 0.308\*\*\* 0.041\*\*\*   
## (0.074) (0.007)   
##   
## Healthcare -0.070 0.006   
## (0.077) (0.008)   
##   
## Construction -0.103 0.011   
## (0.112) (0.012)   
##   
## Mining 0.260 -0.055   
## (0.372) (0.042)   
##   
## Agriculture 0.370 0.001   
## (0.232) (0.026)   
##   
## Manufacturing -0.270\*\*\* 0.008   
## (0.090) (0.009)   
##   
## Services -0.115 -0.010   
## (0.144) (0.015)   
##   
## Transportation -0.253\* -0.022   
## (0.134) (0.014)   
##   
## Wholesale -0.165 -0.008   
## (0.202) (0.021)   
##   
## Retail -0.218\*\* 0.002   
## (0.101) (0.011)   
##   
## Constant -1.150\*\*\* 0.145\*\*\*   
## (0.430) (0.045)   
##   
## ---------------------------------------------------------  
## Observations 7,212 12,505   
## R2 0.118   
## Adjusted R2 0.116   
## Log Likelihood -4,419.626   
## Akaike Inf. Crit. 8,901.252   
## Residual Std. Error 0.305 (df = 12474)   
## F Statistic 55.616\*\*\* (df = 30; 12474)  
## =========================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

suppressMessages(attach(NHIS21\_CLEANED\_R))  
model\_temp7 <- lm(arth ~ genderf + age+ I(age^2)+ white+ black+ asian+ aian+ medicaid + private+ single + married + divorced +educ\_bach + injury +health+ Construction+ Retail+ Agriculture+Healthcare+Manufacturing+ Services+Transportation+Wholesale+ Mining, data = data\_use)  
summary(model\_temp7)

##   
## Call:  
## lm(formula = arth ~ genderf + age + I(age^2) + white + black +   
## asian + aian + medicaid + private + single + married + divorced +   
## educ\_bach + injury + health + Construction + Retail + Agriculture +   
## Healthcare + Manufacturing + Services + Transportation +   
## Wholesale + Mining, data = data\_use)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.53125 -0.15693 -0.06867 -0.00393 1.09410   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.580e-02 4.199e-02 1.567 0.117140   
## genderf 2.762e-02 5.899e-03 4.683 2.86e-06 \*\*\*  
## age -9.012e-03 1.794e-03 -5.023 5.15e-07 \*\*\*  
## I(age^2) 1.759e-04 2.051e-05 8.577 < 2e-16 \*\*\*  
## white 4.264e-02 1.024e-02 4.165 3.14e-05 \*\*\*  
## black 2.435e-02 1.275e-02 1.910 0.056136 .   
## asian -1.271e-02 1.400e-02 -0.908 0.363985   
## aian 7.699e-02 3.435e-02 2.241 0.025043 \*   
## medicaid 2.006e-02 1.425e-02 1.407 0.159414   
## private 6.047e-03 8.811e-03 0.686 0.492532   
## single -2.319e-02 1.554e-02 -1.493 0.135535   
## married -1.651e-02 1.491e-02 -1.108 0.267964   
## divorced -1.172e-02 1.593e-02 -0.736 0.462049   
## educ\_bach -1.801e-02 6.152e-03 -2.927 0.003429 \*\*   
## injury 3.248e-02 9.449e-03 3.437 0.000589 \*\*\*  
## healthFair 2.235e-01 1.253e-02 17.842 < 2e-16 \*\*\*  
## healthGood 1.041e-01 7.567e-03 13.751 < 2e-16 \*\*\*  
## healthPoor 2.850e-01 3.445e-02 8.275 < 2e-16 \*\*\*  
## healthVery good 4.239e-02 6.789e-03 6.244 4.41e-10 \*\*\*  
## Construction 9.173e-03 1.157e-02 0.793 0.427901   
## Retail 3.057e-03 1.052e-02 0.291 0.771311   
## Agriculture -3.420e-03 2.551e-02 -0.134 0.893366   
## Healthcare 7.961e-03 8.448e-03 0.942 0.346011   
## Manufacturing 9.012e-03 9.284e-03 0.971 0.331748   
## Services -8.997e-03 1.480e-02 -0.608 0.543116   
## Transportation -1.884e-02 1.341e-02 -1.404 0.160255   
## Wholesale -6.269e-03 2.071e-02 -0.303 0.762167   
## Mining -5.342e-02 4.185e-02 -1.276 0.201832   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3054 on 12477 degrees of freedom  
## (751 observations deleted due to missingness)  
## Multiple R-squared: 0.1151, Adjusted R-squared: 0.1132   
## F-statistic: 60.11 on 27 and 12477 DF, p-value: < 2.2e-16

norm\_varb <- function(X\_in) {  
 (X\_in - min(X\_in, na.rm = TRUE))/( max(X\_in, na.rm = TRUE) - min(X\_in, na.rm = TRUE) )  
}  
#a factor  
status<- as.factor(health)  
#added a numerical value  
pain <- norm\_varb(backpain+handpain+toothpain+hipspain+arth+abdominalpain)  
#added another numerical value  
race <-norm\_varb(black+white+asian+aian)   
  
norm\_pain <- norm\_varb(pain)  
norm\_race <- norm\_varb(race)  
  
data\_use\_prelim <- cbind(age, I(age^2),genderf,single,married, divorced, private, medicaid, uninsured,educ\_bach,injury, race, pain)  
data\_use\_prelim <- data.frame(data\_use\_prelim)  
  
good\_obs\_data\_use <- complete.cases(data\_use\_prelim,status)  
data\_use <- subset(data\_use\_prelim,good\_obs\_data\_use)  
y\_use <- subset(status,good\_obs\_data\_use)  
status0 <- complete.cases(data\_use\_prelim, status)  
summary(data\_use\_prelim)

## age V2 genderf single   
## Min. :18.00 Min. : 324 Min. :0.0000 Min. :0.0000   
## 1st Qu.:37.00 1st Qu.:1369 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :53.00 Median :2809 Median :1.0000 Median :0.0000   
## Mean :52.63 Mean :3111 Mean :0.5462 Mean :0.2468   
## 3rd Qu.:68.00 3rd Qu.:4624 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :99.00 Max. :9801 Max. :1.0000 Max. :1.0000   
##   
## married divorced private medicaid   
## Min. :0.0000 Min. :0.0000 Min. :0.0 Min. :0.000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0 1st Qu.:0.000   
## Median :0.0000 Median :0.0000 Median :1.0 Median :0.000   
## Mean :0.4558 Mean :0.1495 Mean :0.7 Mean :0.137   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0 3rd Qu.:0.000   
## Max. :1.0000 Max. :1.0000 Max. :1.0 Max. :1.000   
## NA's :8908 NA's :8908   
## uninsured educ\_bach injury race   
## Min. :0.000 Min. :0.0000 Min. :0.00000 Min. :0.0000   
## 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:1.0000   
## Median :0.000 Median :0.0000 Median :0.00000 Median :1.0000   
## Mean :0.111 Mean :0.2363 Mean :0.09406 Mean :0.9259   
## 3rd Qu.:0.000 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:1.0000   
## Max. :1.000 Max. :1.0000 Max. :1.00000 Max. :1.0000   
## NA's :8908   
## pain   
## Min. :0.000   
## 1st Qu.:0.167   
## Median :0.167   
## Mean :0.258   
## 3rd Qu.:0.333   
## Max. :1.000   
## NA's :11171

set.seed(12345)  
NN\_obs <- sum(good\_obs\_data\_use == 1)  
select1 <- (runif(NN\_obs) < 0.8)  
train\_data <- subset(data\_use,select1)  
test\_data <- subset(data\_use,(!select1))  
cl\_data <- y\_use[select1]  
true\_data <- y\_use[!select1]  
  
summary(cl\_data)

## Excellent Fair Good Not Ascertained Poor   
## 1740 1202 2916 0 372   
## Refused Very good   
## 4 3412

prop.table(summary(cl\_data))

## Excellent Fair Good Not Ascertained Poor   
## 0.1803856521 0.1246112378 0.3023014721 0.0000000000 0.0385652084   
## Refused Very good   
## 0.0004146797 0.3537217499

summary(train\_data)

## age V2 genderf single   
## Min. :18.00 Min. : 324 Min. :0.0000 Min. :0.0000   
## 1st Qu.:34.00 1st Qu.:1156 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :46.00 Median :2116 Median :1.0000 Median :0.0000   
## Mean :44.93 Mean :2188 Mean :0.5426 Mean :0.3047   
## 3rd Qu.:57.00 3rd Qu.:3249 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :97.00 Max. :9409 Max. :1.0000 Max. :1.0000   
## married divorced private medicaid   
## 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 :1.0000 Median :0.0000   
## Mean :0.4735 Mean :0.1606 Mean :0.6913 Mean :0.1492   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## uninsured educ\_bach injury race   
## Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.0000   
## Median :0.00000 Median :0.0000 Median :0.0000 Median :1.0000   
## Mean :0.09631 Mean :0.2447 Mean :0.1354 Mean :0.9217   
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000   
## Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## pain   
## Min. :0.0000   
## 1st Qu.:0.1667   
## Median :0.1667   
## Mean :0.2381   
## 3rd Qu.:0.3333   
## Max. :1.0000

require(class)

## Loading required package: class

for (indx in seq(1,15, by= 2)) {  
 pred\_status<- knn(train\_data, test\_data, cl\_data, k = indx, l = 0, prob = FALSE, use.all = TRUE)  
 num\_correct\_labels <- sum(pred\_status == true\_data)  
 correct\_rate <- num\_correct\_labels/length(true\_data)  
 print(c(indx,correct\_rate))  
}

## [1] 1.0000000 0.3025706  
## [1] 3.0000000 0.3240624  
## [1] 5.000000 0.346397  
## [1] 7.0000000 0.3447113  
## [1] 9.0000000 0.3409187  
## [1] 11.0000000 0.3476612  
## [1] 13.0000000 0.3544037  
## [1] 15.0000000 0.3535609

cl\_data\_n <- as.numeric(cl\_data)  
model\_ols1 <- lm(cl\_data\_n ~ train\_data$genderf + train\_data$single + train\_data$married+ train\_data$divorced + train\_data$age + train\_data$uninsured +train\_data$medicaid +train\_data$private+ train\_data$educ\_bach + train\_data$injury )  
summary(model\_ols1)

##   
## Call:  
## lm(formula = cl\_data\_n ~ train\_data$genderf + train\_data$single +   
## train\_data$married + train\_data$divorced + train\_data$age +   
## train\_data$uninsured + train\_data$medicaid + train\_data$private +   
## train\_data$educ\_bach + train\_data$injury)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5608 -1.7977 -0.9234 2.8021 3.6227   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.658658 0.179023 20.437 < 2e-16 \*\*\*  
## train\_data$genderf 0.095283 0.048823 1.952 0.0510 .   
## train\_data$single 0.015535 0.111997 0.139 0.8897   
## train\_data$married 0.145699 0.105156 1.386 0.1659   
## train\_data$divorced -0.046752 0.114406 -0.409 0.6828   
## train\_data$age -0.003503 0.002089 -1.677 0.0936 .   
## train\_data$uninsured 0.211838 0.124134 1.707 0.0879 .   
## train\_data$medicaid -0.010474 0.115766 -0.090 0.9279   
## train\_data$private 0.424395 0.100782 4.211 2.57e-05 \*\*\*  
## train\_data$educ\_bach 0.293634 0.057142 5.139 2.82e-07 \*\*\*  
## train\_data$injury 0.034194 0.070435 0.485 0.6274   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.363 on 9635 degrees of freedom  
## Multiple R-squared: 0.01271, Adjusted R-squared: 0.01168   
## F-statistic: 12.4 on 10 and 9635 DF, p-value: < 2.2e-16

y\_hat <- fitted.values(model\_ols1)  
  
mean(y\_hat[cl\_data\_n == 1])

## [1] 4.072153

mean(y\_hat[cl\_data\_n == 2])

## [1] 3.857105

cl\_data\_n1 <- as.numeric(cl\_data\_n == 1)  
  
model\_ols\_v1 <- lm(cl\_data\_n1 ~ train\_data$genderf +train\_data$single + train\_data$married+ train\_data$divorced + train\_data$age + train\_data$uninsured +train\_data$medicaid +train\_data$private+ train\_data$educ\_bach + train\_data$injury)  
y\_hat\_v1 <- fitted.values(model\_ols\_v1)  
mean(y\_hat\_v1[cl\_data\_n1 == 1])

## [1] 0.2054121

mean(y\_hat\_v1[cl\_data\_n1 == 0])

## [1] 0.1748777

library(readr)  
suppressMessages(NHIS21\_CLEANED\_R <- read\_csv("NHIS21 CLEANED.R.csv"))  
suppressMessages(attach(NHIS21\_CLEANED\_R))  
use\_varb <- (age >=18) & (age <84) & (fulltime == 1)   
data\_use <- subset(NHIS21\_CLEANED\_R,use\_varb)   
attach(data\_use)

## The following objects are masked from NHIS21\_CLEANED\_R (pos = 3):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek

## The following objects are masked from NHIS21\_CLEANED\_R (pos = 5):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek

## The following objects are masked from data\_use (pos = 6):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek

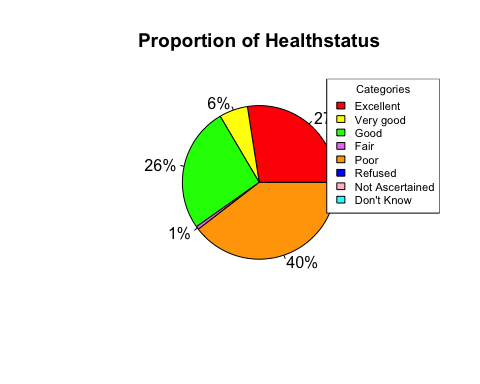
## The following objects are masked from NHIS21\_CLEANED\_R (pos = 7):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek

## The following object is masked from package:survival:  
##   
## veteran

table(data\_use$health)

##   
## Excellent Fair Good Poor Very good   
## 3641 798 3486 88 5243

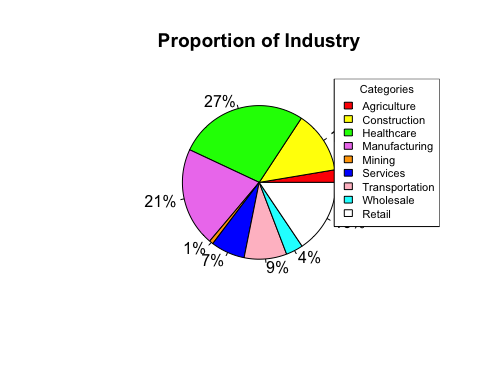
pie(table(data\_use$health),  
 labels = paste(round(prop.table(table(data\_use$health))\*100), "%", sep = ""),   
 col = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan"), main = "Proportion of Healthstatus")  
legend("topright", legend = c("Excellent","Very good", "Good", "Fair", "Poor", "Refused", "Not Ascertained","Don't Know"),   
 fill = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan"), title = "Categories", cex = .7)



data\_use$industry <- factor((data\_use$Agriculture + 2\*data\_use$Construction + 3\*data\_use$Healthcare + 4\*data\_use$Manufacturing + 5\*data\_use$Mining + 6\*data\_use$Services + 7\*data\_use$Transportation + 8\*data\_use$Wholesale + 9\*data\_use$Retail  
),   
levels=c(1,2,3,4,5,6,7,8,9), labels = c("Agriculture","Construction","Healthcare","Manufacturing","Mining","Services","Transportation","Wholesale","Retail"))  
table(data\_use$industry)

##   
## Agriculture Construction Healthcare Manufacturing Mining   
## 180 896 1861 1422 56   
## Services Transportation Wholesale Retail   
## 490 612 247 1063

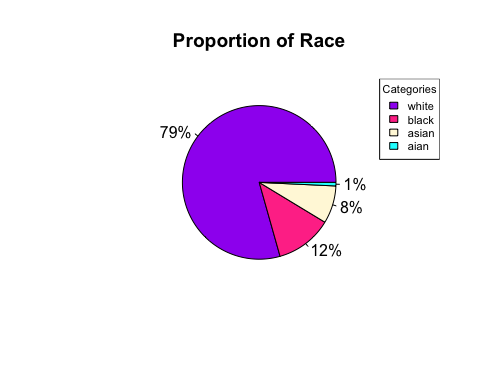
pie(table(data\_use$industry),  
 labels = paste(round(prop.table(table(data\_use$industry))\*100), "%", sep = ""),   
 col = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan", "white"), main = "Proportion of Industry")  
legend("topright", legend = c("Agriculture","Construction","Healthcare","Manufacturing","Mining","Services","Transportation","Wholesale","Retail"),   
 fill = c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan", "white"), title = "Categories", cex = .7)



data\_use$Race <- factor((data\_use$white + 2\*data\_use$black + 3\*data\_use$asian + 4\*data\_use$aian),   
levels=c(1,2,3,4), labels = c("white","black","asian","aian"))  
table(data\_use$Race)

##   
## white black asian aian   
## 9683 1457 967 91

pie(table(data\_use$Race),  
 labels = paste(round(prop.table(table(data\_use$Race))\*100), "%", sep = ""),   
 col = c("purple", "violetred1", "cornsilk", "cyan", "white"), main = "Proportion of Race")  
legend("topright", legend = c("white", "black", "asian","aian"),   
 fill = c("purple", "violetred1", "cornsilk", "cyan", "white"), title = "Categories", cex = .7)



library(readr)  
NHIS21\_CLEANED\_R <- read\_csv("NHIS21 CLEANED.R.csv")

## New names:  
## Rows: 29482 Columns: 54  
## ── Column specification  
## ──────────────────────────────────────────────────────── Delimiter: "," chr  
## (2): region, health dbl (52): ...1, uninsured, Retail, Agriculture,  
## Construction, Healthcare, Ma...  
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
## Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## • `` -> `...1`

attach(NHIS21\_CLEANED\_R)

## The following objects are masked from data\_use (pos = 3):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek  
##   
## The following objects are masked from NHIS21\_CLEANED\_R (pos = 4):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek  
##   
## The following objects are masked from NHIS21\_CLEANED\_R (pos = 6):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek  
##   
## The following objects are masked from data\_use (pos = 7):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek  
##   
## The following objects are masked from NHIS21\_CLEANED\_R (pos = 8):  
##   
## ...1, abdominalpain, age, Agriculture, aian, arth, asian, backpain,  
## black, Construction, divorced, doctorvisit, educ\_adv, educ\_as,  
## educ\_bach, educ\_hs, educ\_nohs, educ\_smcoll, ER, foodworker,  
## fulltime, genderf, handpain, health, Healthcare, hipspain, injury,  
## Manufacturing, married, medcare, medicaid, Mining, missed,  
## paidleave, payworry, private, problembill, reducework, region,  
## Retail, rinjury, seperated, Services, single, somepain, stopwork,  
## toothpain, Transportation, uninsured, veteran, white, Wholesale,  
## widowed, worklastweek  
##   
## The following object is masked from package:survival:  
##   
## veteran

use\_varb <- (age >=18) & (age <84) & (fulltime == 1)   
data\_use <- subset(NHIS21\_CLEANED\_R,use\_varb)   
  
data\_use$industry <- factor((data\_use$Agriculture + 2\*data\_use$Construction + 3\*data\_use$Healthcare + 4\*data\_use$Manufacturing + 5\*data\_use$Mining + 6\*data\_use$Services + 7\*data\_use$Transportation + 8\*data\_use$Wholesale + 9\*data\_use$Retail  
),   
levels=c(1,2,3,4,5,6,7,8,9), labels = c("Agriculture","Construction","Healthcare","Manufacturing","Mining","Services","Transportation","Wholesale","Retail"))  
  
table1 <- table(data\_use$industry,data\_use$paidleave)  
table1

##   
## 0 1  
## Agriculture 104 76  
## Construction 412 484  
## Healthcare 315 1546  
## Manufacturing 274 1148  
## Mining 11 45  
## Services 190 300  
## Transportation 199 413  
## Wholesale 36 211  
## Retail 251 812

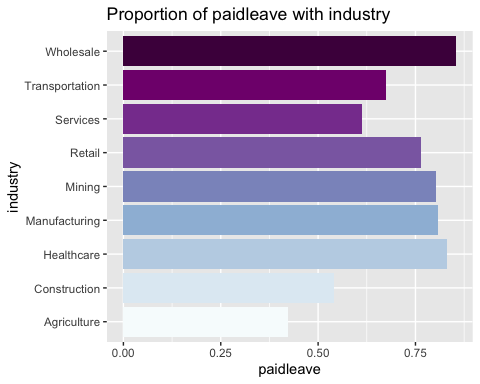
x=c(table1[1,2]/(table1[1,1]+table1[1,2]),  
 table1[2,2]/(table1[2,1]+table1[2,2]),  
 table1[3,2]/(table1[3,1]+table1[3,2]),  
 table1[4,2]/(table1[4,1]+table1[4,2]),  
 table1[5,2]/(table1[5,1]+table1[5,2]),  
 table1[6,2]/(table1[6,1]+table1[6,2]),  
 table1[7,2]/(table1[7,1]+table1[7,2]),  
 table1[8,2]/(table1[8,1]+table1[8,2]),  
 table1[9,2]/(table1[9,1]+table1[9,2]))  
x

## [1] 0.4222222 0.5401786 0.8307362 0.8073136 0.8035714 0.6122449 0.6748366  
## [8] 0.8542510 0.7638758

paidleaveV\_prop\_table<-data.frame(row.names=row.names(table1),Prop\_paidleave=x)  
paidleaveV\_prop\_table

## Prop\_paidleave  
## Agriculture 0.4222222  
## Construction 0.5401786  
## Healthcare 0.8307362  
## Manufacturing 0.8073136  
## Mining 0.8035714  
## Services 0.6122449  
## Transportation 0.6748366  
## Wholesale 0.8542510  
## Retail 0.7638758

require(ggplot2)  
ggplot(data=paidleaveV\_prop\_table, aes(y=row.names(paidleaveV\_prop\_table), x=Prop\_paidleave, fill=row.names(paidleaveV\_prop\_table))) +   
 geom\_bar(stat="identity") + scale\_fill\_brewer(palette = "BuPu") + ggtitle("Proportion of paidleave with industry") + theme(legend.position = "none") + labs( x="paidleave", y="industry")



data\_use$insurance<- factor((data\_use$medicaid + 2\*data\_use$private+ 3\*data\_use$uninsured),   
 levels=c(1,2,3), labels = c("medicaid","private","uninsured"))  
table3 <- table(data\_use$insurance,data\_use$medcare)  
table3

##   
## 0 1  
## medicaid 649 35  
## private 9672 678  
## uninsured 930 245

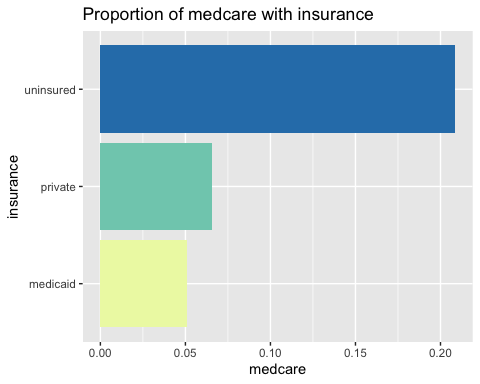
x=c(table3[1,2]/(table3[1,1]+table3[1,2]),  
 table3[2,2]/(table3[2,1]+table3[2,2]),  
 table3[3,2]/(table3[3,1]+table3[3,2]))  
x

## [1] 0.05116959 0.06550725 0.20851064

medcareV\_prop\_table<-data.frame(row.names=row.names(table3), Prop\_medcare=x)  
medcareV\_prop\_table

## Prop\_medcare  
## medicaid 0.05116959  
## private 0.06550725  
## uninsured 0.20851064

require(ggplot2)  
ggplot(data=medcareV\_prop\_table, aes(y=row.names(medcareV\_prop\_table), x=Prop\_medcare, fill=row.names(medcareV\_prop\_table))) +   
 geom\_bar(stat="identity") + scale\_fill\_brewer(palette = "YlGnBu") + ggtitle("Proportion of medcare with insurance") + theme(legend.position = "none") + labs( x="medcare", y="insurance")



data\_use$Race <- factor((data\_use$white + 2\*data\_use$black + 3\*data\_use$asian + 4\*data\_use$aian),   
levels=c(1,2,3,4), labels = c("white","black","asian","aian"))  
  
table4 <- table(data\_use$Race,data\_use$doctorvisit)  
table4

##   
## 0 1  
## white 2008 7675  
## black 220 1237  
## asian 279 688  
## aian 14 77

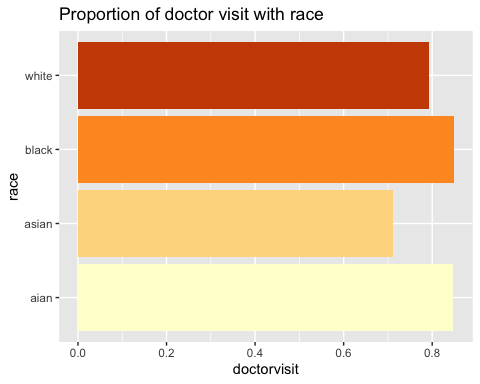
x=c(table4[1,2]/(table4[1,1]+table4[1,2]),  
 table4[2,2]/(table4[2,1]+table4[2,2]),  
 table4[3,2]/(table4[3,1]+table4[3,2]),  
 table4[4,2]/(table4[4,1]+table4[4,2]))  
x

## [1] 0.7926263 0.8490048 0.7114788 0.8461538

doctorvisitV\_prop\_table<-data.frame(row.names=row.names(table4), Prop\_doctorvisit=x)  
doctorvisitV\_prop\_table

## Prop\_doctorvisit  
## white 0.7926263  
## black 0.8490048  
## asian 0.7114788  
## aian 0.8461538

require(ggplot2)  
ggplot(data=doctorvisitV\_prop\_table, aes(y=row.names(doctorvisitV\_prop\_table), x=Prop\_doctorvisit, fill=row.names(doctorvisitV\_prop\_table))) +   
 geom\_bar(stat="identity") + scale\_fill\_brewer(palette = "YlOrBr") + ggtitle("Proportion of doctor visit with race") + theme(legend.position = "none") + labs( x="doctorvisit", y="race")



data\_use$maritial <- factor((data\_use$single + 2\*data\_use$widowed + 3\*data\_use$married + 4\*data\_use$divorced),   
 levels=c(1,2,3,4), labels = c("single","widowed","divorced","married"))  
table5 <- table(data\_use$maritial,data\_use$problembill)  
table5

##   
## 0 1  
## single 3712 375  
## widowed 273 33  
## divorced 6040 527  
## married 1791 228

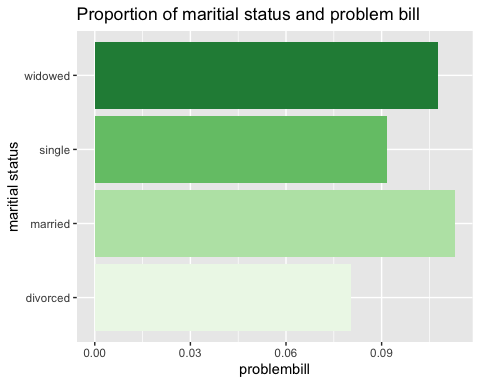
x=c(table5[1,2]/(table5[1,1]+table5[1,2]),  
 table5[2,2]/(table5[2,1]+table5[2,2]),  
 table5[3,2]/(table5[3,1]+table5[3,2]),  
 table5[4,2]/(table5[4,1]+table5[4,2]))  
x

## [1] 0.09175434 0.10784314 0.08024973 0.11292719

problembillV\_prop\_table<-data.frame(row.names=row.names(table5), Prop\_problembill=x)  
problembillV\_prop\_table

## Prop\_problembill  
## single 0.09175434  
## widowed 0.10784314  
## divorced 0.08024973  
## married 0.11292719

require(ggplot2)  
ggplot(data=problembillV\_prop\_table, aes(y=row.names(problembillV\_prop\_table), x=Prop\_problembill, fill=row.names(problembillV\_prop\_table))) +   
 geom\_bar(stat="identity") + scale\_fill\_brewer(palette = "Greens") + ggtitle("Proportion of maritial status and problem bill") + theme(legend.position = "none") + labs( x="problembill", y="maritial status")



data\_use$bodypain <- factor((data\_use$backpain + 2\*data\_use$hipspain + 3\*data\_use$handpain+ 4\*data\_use$abdominalpain+5\*data\_use$toothpain),   
 levels=c(1,2,3,4,5),labels = c("backpain","hipspain","abdominalpain","handpain","toothpain"))  
table6 <- table(data\_use$bodypain,data\_use$somepain)  
table6

##   
## 0 1  
## backpain 635 281  
## hipspain 506 173  
## abdominalpain 782 305  
## handpain 386 179  
## toothpain 443 231

x=c(table6[1,2]/(table6[1,1]+table6[1,2]),  
 table6[2,2]/(table6[2,1]+table6[2,2]),  
 table6[3,2]/(table6[3,1]+table6[3,2]),  
 table6[4,2]/(table6[4,1]+table6[4,2]),  
 table6[5,2]/(table6[5,1]+table6[5,2]))  
x

## [1] 0.3067686 0.2547865 0.2805888 0.3168142 0.3427300

somepainV\_prop\_table<-data.frame(row.names=row.names(table6), Prop\_somepain=x)  
somepainV\_prop\_table

## Prop\_somepain  
## backpain 0.3067686  
## hipspain 0.2547865  
## abdominalpain 0.2805888  
## handpain 0.3168142  
## toothpain 0.3427300

require(ggplot2)  
ggplot(data=somepainV\_prop\_table, aes(y=row.names(somepainV\_prop\_table), x=Prop\_somepain, fill=row.names(somepainV\_prop\_table))) +   
 geom\_bar(stat="identity") + scale\_fill\_brewer(palette = "Reds") + ggtitle("Types of pain with some pain that limits work and life") + theme(legend.position = "none") + labs( x="somepain", y=" type of pain")

