# BIOSTAT 650 Project

Jaehoon Kim

2024-11-17

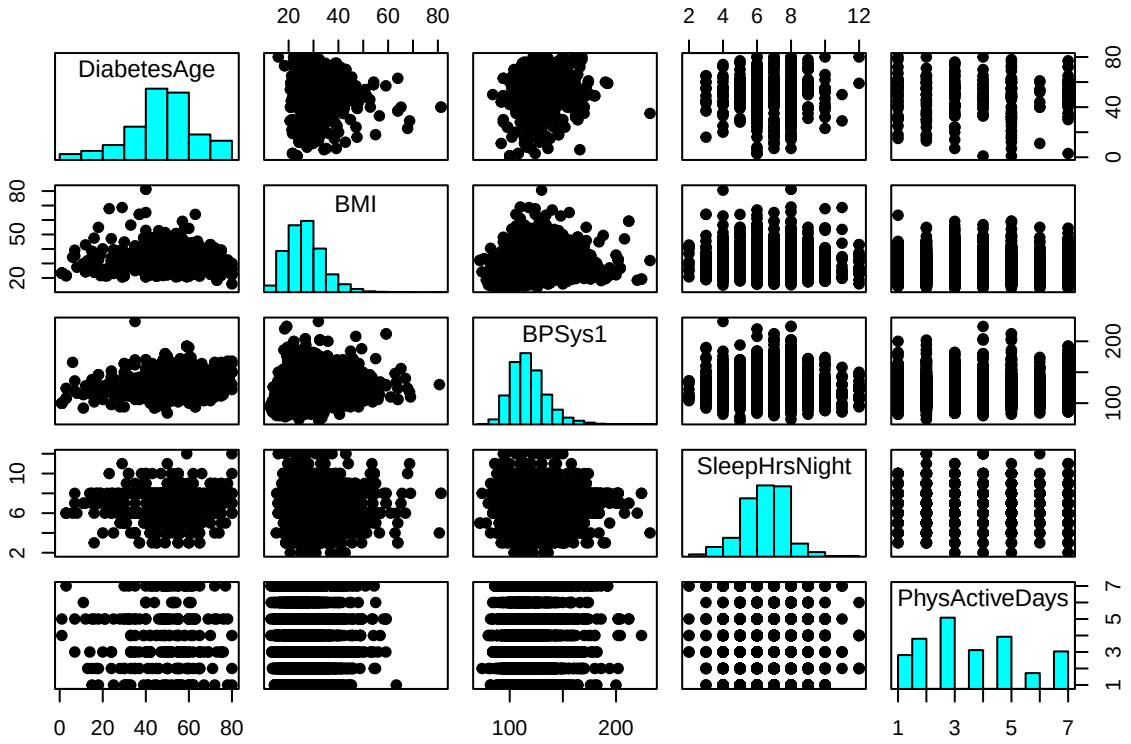
$1 + 1$

```
df = NHANES
#df = NHANES["DiabetesAge" > 20]
colnames(df)
```

```
##  [1] "ID"               "SurveyYr"         "Gender"           "Age"
##  [5] "AgeDecade"        "AgeMonths"        "Race1"            "Race3"
##  [9] "Education"        "MaritalStatus"    "HHIncome"         "HHIncomeMid"
## [13] "Poverty"          "HomeRooms"        "HomeOwn"          "Work"
## [17] "Weight"           "Length"           "HeadCirc"         "Height"
## [21] "BMI"              "BMICatUnder20yrs" "BMI_WHO"          "Pulse"
## [25] "BPSysAve"         "BPDiaAve"         "BPSys1"           "BPDia1"
## [29] "BPSys2"           "BPDia2"           "BPSys3"           "BPDia3"
## [33] "Testosterone"     "DirectChol"       "TotChol"          "UrineVol1"
## [37] "UrineFlow1"       "UrineVol2"        "UrineFlow2"       "Diabetes"
## [41] "DiabetesAge"      "HealthGen"        "DaysPhysHlthBad"  "DaysMentHlthBad"
## [45] "LittleInterest"   "Depressed"        "nPregnancies"     "nBabies"
## [49] "Age1stBaby"       "SleepHrsNight"    "SleepTrouble"     "PhysActive"
## [53] "PhysActiveDays"   "TVHrsDay"         "CompHrsDay"       "TVHrsDayChild"
## [57] "CompHrsDayChild"  "Alcohol12PlusYr"  "AlcoholDay"       "AlcoholYear"
## [61] "SmokeNow"         "Smoke100"         "Smoke100n"        "SmokeAge"
## [65] "Marijuana"        "AgeFirstMarij"    "RegularMarij"     "AgeRegMarij"
## [69] "HardDrugs"        "SexEver"          "SexAge"           "SexNumPartnLife"
## [73] "SexNumPartYear"   "SameSex"          "SexOrientation"   "PregnantNow"
```

```
scatmatrixData = df[,c("DiabetesAge", "BMI", "BPSys1", "SleepHrsNight", "PhysActiveDays")]
panel.hist <- function(x, ...)
{
usr <- par("usr"); on.exit(par(usr))
par(usr = c(usr[1:2], 0, 1.5) )
h <- hist(x, plot = FALSE)
breaks <- h$breaks; nB <- length(breaks)
y <- h$counts; y <- y/max(y)
rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}
pairs(scatmatrixData, pch = 19, diag.panel=panel.hist)
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```



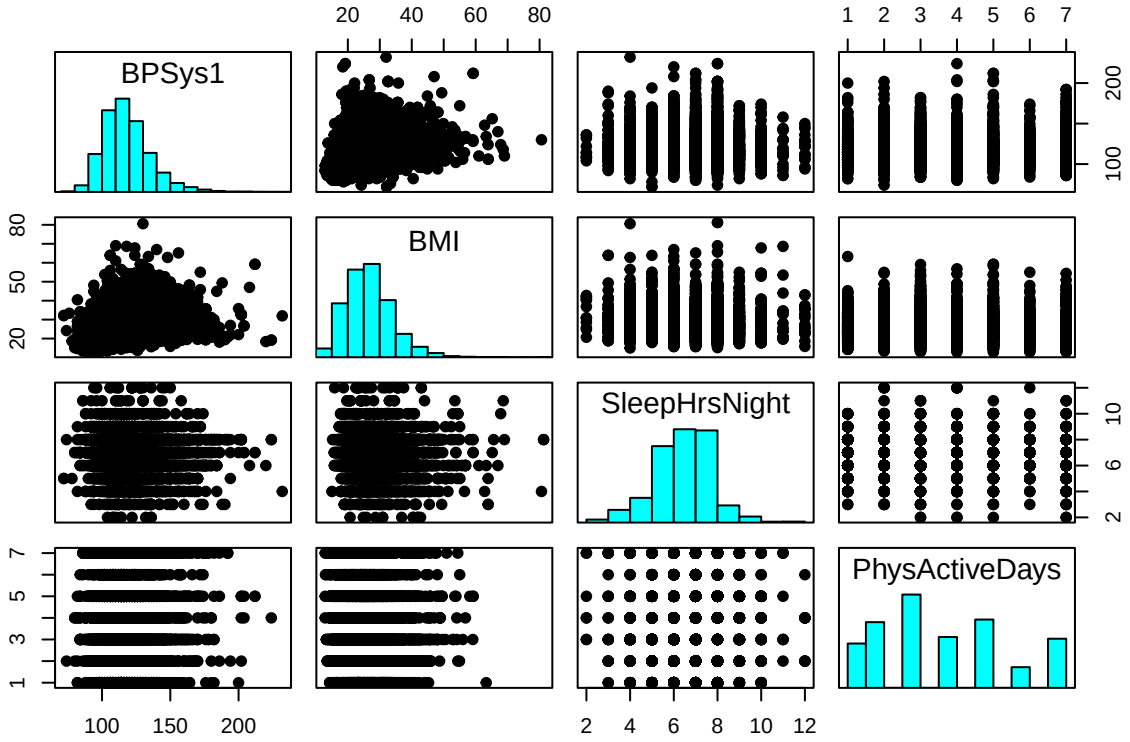

```
df = NHANES
#poverty gender useless?
colnames(df)
```

```
##   [1] "ID"               "SurveyYr"         "Gender"           "Age"
##   [5] "AgeDecade"        "AgeMonths"        "Race1"            "Race3"
##   [9] "Education"        "MaritalStatus"    "HHIncome"         "HHIncomeMid"
##  [13] "Poverty"          "HomeRooms"        "HomeOwn"          "Work"
##  [17] "Weight"           "Length"           "HeadCirc"         "Height"
##  [21] "BMI"              "BMICatUnder20yrs" "BMI_WHO"          "Pulse"
##  [25] "BPSysAve"         "BPDiaAve"         "BPSys1"           "BPDia1"
##  [29] "BPSys2"           "BPDia2"           "BPSys3"           "BPDia3"
##  [33] "Testosterone"     "DirectChol"       "TotChol"          "UrineVol1"
##  [37] "UrineFlow1"       "UrineVol2"        "UrineFlow2"       "Diabetes"
##  [41] "DiabetesAge"      "HealthGen"        "DaysPhysHlthBad"  "DaysMentHlthBad"
##  [45] "LittleInterest"   "Depressed"        "nPregnancies"     "nBabies"
##  [49] "Age1stBaby"       "SleepHrsNight"    "SleepTrouble"     "PhysActive"
##  [53] "PhysActiveDays"   "TVHrsDay"         "CompHrsDay"       "TVHrsDayChild"
##  [57] "CompHrsDayChild"  "Alcohol12PlusYr"  "AlcoholDay"       "AlcoholYear"
##  [61] "SmokeNow"         "Smoke100"         "Smoke100n"        "SmokeAge"
##  [65] "Marijuana"        "AgeFirstMarij"    "RegularMarij"     "AgeRegMarij"
##  [69] "HardDrugs"        "SexEver"          "SexAge"           "SexNumPartnLife"
##  [73] "SexNumPartYear"   "SameSex"          "SexOrientation"   "PregnantNow"
```

```
scatmatrixData = df[,c("BPSys1", "BMI", "SleepHrsNight",
    "PhysActiveDays")]
panel.hist <- function(x, ...)
```

```
{
usr <- par("usr"); on.exit(par(usr))
par(usr = c(usr[1:2], 0, 1.5) )
h <- hist(x, plot = FALSE)
breaks <- h$breaks; nB <- length(breaks)
y <- h$counts; y <- y/max(y)
rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}
pairs(scatmatrixData, pch = 19, diag.panel=panel.hist)
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```




```
model <- lm(DiabetesAge ~ Gender+Poverty+BMI+BPSys1+SleepHrsNight+PhysActiveDays, df)
summary(model)
```

```
##
## Call:
## lm(formula = DiabetesAge ~ Gender + Poverty + BMI + BPSys1 +
##     SleepHrsNight + PhysActiveDays, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.087  -7.907   2.062   8.861  29.318
```

```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    32.96048   10.92836   3.016  0.00287 **
## Gendermale     -2.46465    2.11661  -1.164  0.24553
## Poverty         0.46344    0.62309   0.744  0.45781
## BMI            -0.09236    0.14055  -0.657  0.51180
## BPSys1          0.13469    0.05758   2.339  0.02024 *
## SleepHrsNight   0.25571    0.73547   0.348  0.72841
## PhysActiveDays -0.19888    0.53308  -0.373  0.70945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.09 on 217 degrees of freedom
##   (9776 observations deleted due to missingness)
## Multiple R-squared:  0.04008,    Adjusted R-squared:  0.01354
## F-statistic:  1.51 on 6 and 217 DF,  p-value: 0.176
```

```
model <- lm(BPSys1 ~ Age+Gender+Poverty+BMI+SleepHrsNight+PhysActiveDays+SmokeNow+AlcoholYear+HardDrugs
summary(model)
```

```
##
## Call:
## lm(formula = BPSys1 ~ Age + Gender + Poverty + BMI + SleepHrsNight +
##     PhysActiveDays + SmokeNow + AlcoholYear + HardDrugs, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.397  -8.387  -0.997   7.730  69.906
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    89.959564   3.820975  23.544  < 2e-16 ***
## Age             0.413402   0.035437  11.666  < 2e-16 ***
## Gendermale      5.382522   0.903317   5.959 3.48e-09 ***
## Poverty        -0.843665   0.283924  -2.971  0.00303 **
## BMI             0.345235   0.075337   4.583 5.15e-06 ***
## SleepHrsNight   0.247155   0.331007   0.747  0.45543
## PhysActiveDays -0.021275   0.244823  -0.087  0.93077
## SmokeNowYes     1.325291   0.957252   1.384  0.16651
## AlcoholYear     0.002536   0.004169   0.608  0.54318
## HardDrugsYes    0.141125   0.964282   0.146  0.88367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.18 on 1038 degrees of freedom
##   (8952 observations deleted due to missingness)
## Multiple R-squared:  0.1709, Adjusted R-squared:  0.1637
## F-statistic: 23.78 on 9 and 1038 DF,  p-value: < 2.2e-16
```

```
model <- lm(SexAge ~ Gender+HHIncome+Education+PhysActive+RegularMarij+HardDrugs+RegularMarij*HardDrugs
summary(model)
```

```
##
## Call:
```

```
## lm(formula = SexAge ~ Gender + HHIncome + Education + PhysActive +
##     RegularMarij + HardDrugs + RegularMarij * HardDrugs, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9040 -1.9402 -0.3804  1.2866 27.4776
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   17.51877    0.50254  34.860  < 2e-16 ***
## Gendermale                    -0.03626    0.10663  -0.340   0.7338
## HHIncome 5000-9999            -0.79000    0.54537  -1.449   0.1475
## HHIncome10000-14999           -0.45927    0.46515  -0.987   0.3235
## HHIncome15000-19999           -1.05879    0.46685  -2.268   0.0234 *
## HHIncome20000-24999           -0.44423    0.45914  -0.968   0.3333
## HHIncome25000-34999           -0.36138    0.43797  -0.825   0.4093
## HHIncome35000-44999           -0.15185    0.43796  -0.347   0.7288
## HHIncome45000-54999            0.35892    0.43940   0.817   0.4141
## HHIncome55000-64999           -0.72483    0.44785  -1.618   0.1056
## HHIncome65000-74999            0.38132    0.45346   0.841   0.4004
## HHIncome75000-99999            0.13231    0.42886   0.309   0.7577
## HHIncomemore 99999            -0.22322    0.41947  -0.532   0.5947
## Education9 - 11th Grade        0.14490    0.33425   0.434   0.6647
## EducationHigh School           0.50458    0.31875   1.583   0.1135
## EducationSome College          0.50239    0.31402   1.600   0.1097
## EducationCollege Grad          1.88960    0.32392   5.833 5.84e-09 ***
## PhysActiveYes                 -0.24492    0.11227  -2.182   0.0292 *
## RegularMarijYes               -2.04829    0.15499 -13.215  < 2e-16 ***
## HardDrugsYes                  -1.62404    0.21629  -7.509 7.25e-14 ***
## RegularMarijYes:HardDrugsYes   1.49673    0.29129   5.138 2.90e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.399 on 4205 degrees of freedom
##   (5774 observations deleted due to missingness)
## Multiple R-squared:  0.1358, Adjusted R-squared:  0.1317
## F-statistic: 33.05 on 20 and 4205 DF,  p-value: < 2.2e-16
```

```
model <- lm(SexAge ~ Gender+HHIncome+Education+SameSex+PhysActive+RegularMarij+HardDrugs+RegularMarij*Ha
summary(model)
```

```
##
## Call:
## lm(formula = SexAge ~ Gender + HHIncome + Education + SameSex +
##     PhysActive + RegularMarij + HardDrugs + RegularMarij * HardDrugs,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9073 -1.9665 -0.4121  1.2964 27.4144
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   17.54801    0.50328  34.867  < 2e-16 ***
## Gendermale                    -0.07223    0.10749  -0.672   0.5016
```

```
## HHIncome 5000-9999           -0.79270    0.54506  -1.454   0.1459
## HHIncome10000-14999          -0.44989    0.46490  -0.968   0.3332
## HHIncome15000-19999          -1.06281    0.46658  -2.278   0.0228 *
## HHIncome20000-24999          -0.44484    0.45888  -0.969   0.3324
## HHIncome25000-34999          -0.38598    0.43784  -0.882   0.3781
## HHIncome35000-44999          -0.18232    0.43789  -0.416   0.6772
## HHIncome45000-54999           0.35222    0.43915   0.802   0.4226
## HHIncome55000-64999          -0.73119    0.44760  -1.634   0.1024
## HHIncome65000-74999           0.32731    0.45372   0.721   0.4707
## HHIncome75000-99999           0.08799    0.42898   0.205   0.8375
## HHIncomemore 99999           -0.25391    0.41941  -0.605   0.5449
## Education9 - 11th Grade       0.16340    0.33500   0.488   0.6257
## EducationHigh School          0.52625    0.31954   1.647   0.0997 .
## EducationSome College         0.53590    0.31488   1.702   0.0888 .
## EducationCollege Grad         1.93066    0.32478   5.945 3.00e-09 ***
## SameSexYes                   -0.49517    0.19924  -2.485   0.0130 *
## PhysActiveYes                -0.24524    0.11221  -2.186   0.0289 *
## RegularMarijYes              -2.01369    0.15549 -12.950  < 2e-16 ***
## HardDrugsYes                 -1.54232    0.21857  -7.056 1.99e-12 ***
## RegularMarijYes:HardDrugsYes  1.46429    0.29139   5.025 5.24e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.397 on 4203 degrees of freedom
##   (5775 observations deleted due to missingness)
## Multiple R-squared:  0.1372, Adjusted R-squared:  0.1328
## F-statistic: 31.81 on 21 and 4203 DF,  p-value: < 2.2e-16
```

```
model <- lm(SexNumPartnLife ~ Gender+HHIncome+Education+PhysActive+RegularMarij+HardDrugs+RegularMarij*
summary(model)
```

```
##
## Call:
## lm(formula = SexNumPartnLife ~ Gender + HHIncome + Education +
##     PhysActive + RegularMarij + HardDrugs + RegularMarij * HardDrugs,
##     data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43.88 -11.51  -4.29   2.76 985.61
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -3.10099    7.13864  -0.434   0.6640
## Gendermale                    8.77546    1.51990   5.774 8.30e-09 ***
## HHIncome 5000-9999           14.54638    7.76891   1.872   0.0612 .
## HHIncome10000-14999           3.78538    6.62111   0.572   0.5675
## HHIncome15000-19999           0.04752    6.67954   0.007   0.9943
## HHIncome20000-24999           8.46345    6.59501   1.283   0.1995
## HHIncome25000-34999          11.18533    6.26544   1.785   0.0743 .
## HHIncome35000-44999           1.12603    6.27352   0.179   0.8576
## HHIncome45000-54999           1.67325    6.29487   0.266   0.7904
## HHIncome55000-64999           2.52128    6.40564   0.394   0.6939
## HHIncome65000-74999           3.25426    6.51323   0.500   0.6174
## HHIncome75000-99999           4.36560    6.14932   0.710   0.4778
```

```
## HHIncomemore 99999            4.36177    6.01363   0.725   0.4683
## Education9 - 11th Grade       5.45707    4.69156   1.163   0.2448
## EducationHigh School          4.54384    4.45914   1.019   0.3083
## EducationSome College         1.14179    4.38485   0.260   0.7946
## EducationCollege Grad        -2.03712    4.52072  -0.451   0.6523
## PhysActiveYes                 3.02096    1.60090   1.887   0.0592 .
## RegularMarijYes              13.61541    2.23551   6.091 1.22e-09 ***
## HardDrugsYes                 12.66710    3.11864   4.062 4.96e-05 ***
## RegularMarijYes:HardDrugsYes -4.10977    4.21049  -0.976   0.3291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 49.13 on 4323 degrees of freedom
##   (5656 observations deleted due to missingness)
## Multiple R-squared:  0.05162,    Adjusted R-squared:  0.04723
## F-statistic: 11.77 on 20 and 4323 DF,  p-value: < 2.2e-16
```

```
model <- lm(SexNumPartnLife ~ Gender+HHIncome+Education+PhysActive+SameSex+RegularMarij+HardDrugs+Regula
summary(model)
```

```
##
## Call:
## lm(formula = SexNumPartnLife ~ Gender + HHIncome + Education +
##     PhysActive + SameSex + RegularMarij + HardDrugs + RegularMarij *
##     HardDrugs, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43.99 -11.32  -4.30   2.80 985.80
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -2.83227    7.15102  -0.396   0.6921
## Gendermale                    8.62320    1.53271   5.626 1.96e-08 ***
## HHIncome 5000-9999           14.55906    7.77014   1.874   0.0610 .
## HHIncome10000-14999           3.86482    6.62286   0.584   0.5595
## HHIncome15000-19999           0.06679    6.68064   0.010   0.9920
## HHIncome20000-24999           8.50076    6.59625   1.289   0.1976
## HHIncome25000-34999          11.17764    6.26741   1.783   0.0746 .
## HHIncome35000-44999           1.02913    6.27553   0.164   0.8697
## HHIncome45000-54999           1.68879    6.29584   0.268   0.7885
## HHIncome55000-64999           2.53680    6.40663   0.396   0.6922
## HHIncome65000-74999           3.05708    6.51876   0.469   0.6391
## HHIncome75000-99999           4.21680    6.15303   0.685   0.4932
## HHIncomemore 99999            4.27884    6.01544   0.711   0.4769
## Education9 - 11th Grade       5.35105    4.70437   1.137   0.2554
## EducationHigh School          4.45800    4.47243   0.997   0.3189
## EducationSome College         1.10825    4.39882   0.252   0.8011
## EducationCollege Grad        -2.03806    4.53482  -0.449   0.6531
## PhysActiveYes                 3.00891    1.60123   1.879   0.0603 .
## SameSexYes                   -2.32060    2.88395  -0.805   0.4211
## RegularMarijYes              13.77346    2.24501   6.135 9.27e-10 ***
## HardDrugsYes                 13.04387    3.15518   4.134 3.63e-05 ***
## RegularMarijYes:HardDrugsYes -4.26299    4.21578  -1.011   0.3120
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 49.14 on 4321 degrees of freedom
##   (5657 observations deleted due to missingness)
## Multiple R-squared:  0.05177,    Adjusted R-squared:  0.04716
## F-statistic: 11.23 on 21 and 4321 DF,  p-value: < 2.2e-16
```