BIOSTAT 650 Project

Jaehoon Kim

2024-11-17

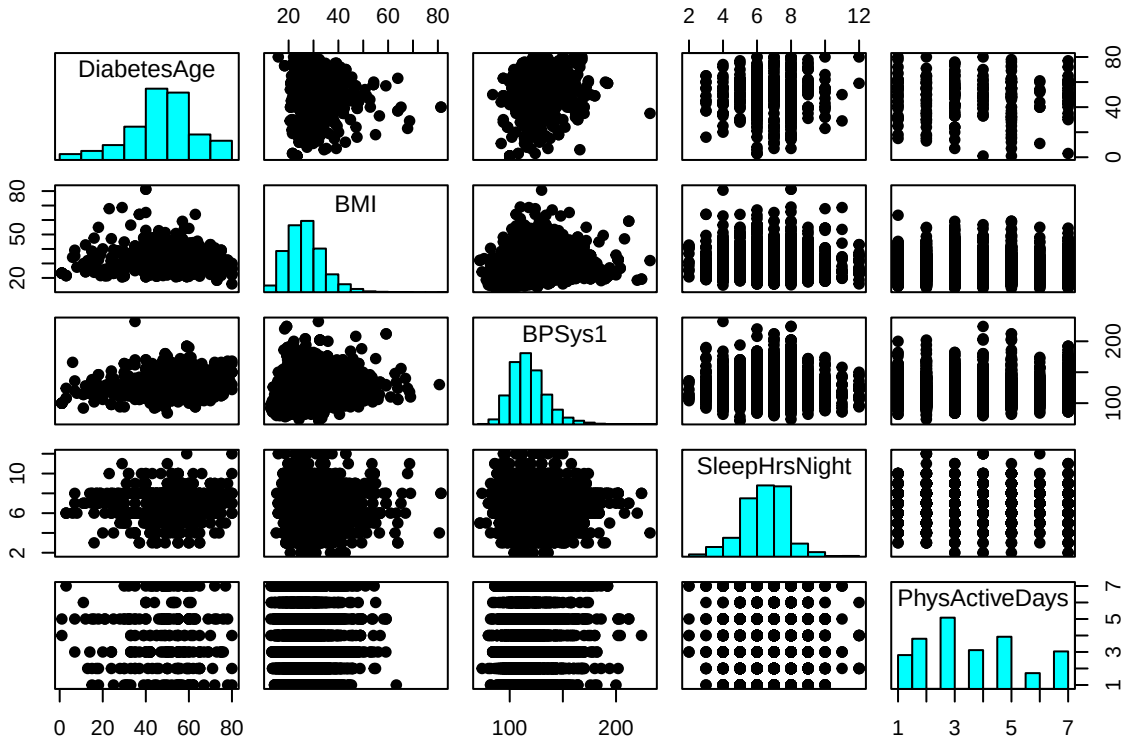
$1 + 1$

```
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
```



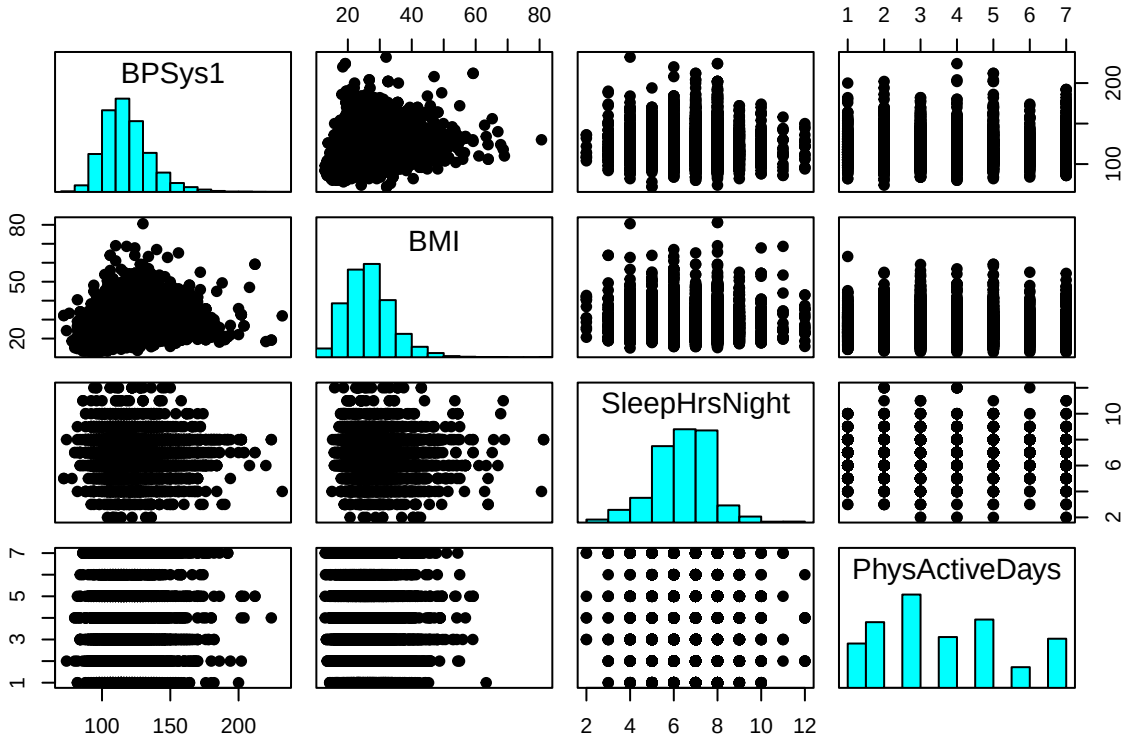

```
df = NHANES
#poverty gender useless?
colnames(df)
```

```
##   [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
```




```
#cor(df)

#outlier analysis
#sensitivity analysis
#QQ-Plot -> quadratic terms(transformation necessary?)
#Categorization -> non linearity?
```

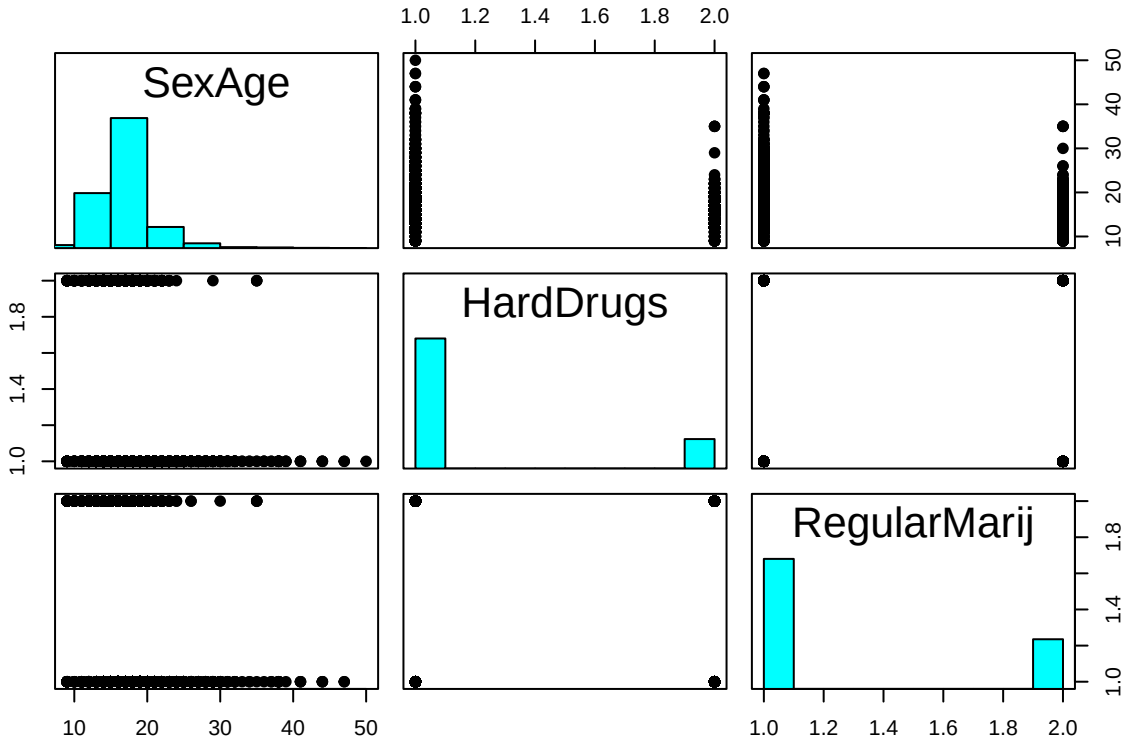
```
df = NHANES
#df = NHANES["DiabetesAge" > 20]
colnames(df)
```

```
##  [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("SexAge", "HardDrugs", "RegularMarij")]
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
```




```
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 ~ Depressed+LittleInterest+HealthGen+Gender+HHIncome+Education+PhysActive+RegularMar
summary(model)
```

```
##
## Call:
## lm(formula = SexAge ~ Depressed + LittleInterest + HealthGen +
##     Gender + HHIncome + Education + PhysActive + RegularMarij +
##     HardDrugs + RegularMarij * HardDrugs + Depressed * HardDrugs +
##     SmokeAge, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2968 -1.4972 -0.1227  1.1686 20.5223
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  16.342991   0.624806  26.157  < 2e-16 ***
## DepressedSeveral             -0.177236   0.241818  -0.733 0.463700
## DepressedMost                -1.291956   0.374178  -3.453 0.000568 ***
## LittleInterestSeveral        -0.231825   0.191238  -1.212 0.225587
```

```
## LittleInterestMost             0.322324   0.277909   1.160 0.246281
## HealthGenVgood                 0.200654   0.267130   0.751 0.452665
## HealthGenGood                 -0.340287   0.264213  -1.288 0.197942
## HealthGenFair                 -0.002334   0.300057  -0.008 0.993793
## HealthGenPoor                 -0.184880   0.467620  -0.395 0.692623
## Gendermale                     0.304082   0.129913   2.341 0.019362 *
## HHIncome 5000-9999            -1.348405   0.557167  -2.420 0.015618 *
## HHIncome10000-14999           -1.088389   0.480505  -2.265 0.023629 *
## HHIncome15000-19999           -1.294652   0.483536  -2.677 0.007488 **
## HHIncome20000-24999           -1.369399   0.477907  -2.865 0.004215 **
## HHIncome25000-34999           -0.949078   0.460535  -2.061 0.039469 *
## HHIncome35000-44999           -1.471535   0.469899  -3.132 0.001767 **
## HHIncome45000-54999           -0.426089   0.466347  -0.914 0.361014
## HHIncome55000-64999           -1.784112   0.478566  -3.728 0.000199 ***
## HHIncome65000-74999           -0.933033   0.488515  -1.910 0.056305 .
## HHIncome75000-99999           -1.144292   0.456791  -2.505 0.012333 *
## HHIncomemore 99999            -1.242224   0.442429  -2.808 0.005045 **
## Education9 - 11th Grade       -0.218123   0.341017  -0.640 0.522501
## EducationHigh School          -0.179374   0.332905  -0.539 0.590085
## EducationSome College          0.189442   0.332127   0.570 0.568486
## EducationCollege Grad          1.445331   0.352639   4.099 4.35e-05 ***
## PhysActiveYes                 -0.599686   0.133608  -4.488 7.65e-06 ***
## RegularMarijYes               -1.256137   0.167049  -7.520 8.74e-14 ***
## HardDrugsYes                  -0.891059   0.248838  -3.581 0.000352 ***
## SmokeAge                       0.100107   0.013415   7.462 1.34e-13 ***
## RegularMarijYes:HardDrugsYes   0.834558   0.290879   2.869 0.004166 **
## DepressedSeveral:HardDrugsYes -0.184463   0.332563  -0.555 0.579190
## DepressedMost:HardDrugsYes     0.565576   0.465395   1.215 0.224432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.624 on 1744 degrees of freedom
##   (8224 observations deleted due to missingness)
## Multiple R-squared:  0.1699, Adjusted R-squared:  0.1551
## F-statistic: 11.51 on 31 and 1744 DF,  p-value: < 2.2e-16
```

```
model <- lm(SexAge ~ RegularMarij+HardDrugs+RegularMarij*HardDrugs, df)
summary(model)
```

```
##
## Call:
## lm(formula = SexAge ~ RegularMarij + HardDrugs + RegularMarij *
##     HardDrugs, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0399 -2.0399 -0.3123  1.1842 28.9601
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  18.03995    0.06268 287.823  < 2e-16 ***
## RegularMarijYes              -2.22420    0.14750 -15.080  < 2e-16 ***
## HardDrugsYes                 -1.72766    0.20925  -8.256  < 2e-16 ***
## RegularMarijYes:HardDrugsYes  1.44824    0.28116   5.151  2.7e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.464 on 4712 degrees of freedom
##   (5284 observations deleted due to missingness)
## Multiple R-squared:  0.08977,    Adjusted R-squared:  0.08919
## F-statistic: 154.9 on 3 and 4712 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)
```

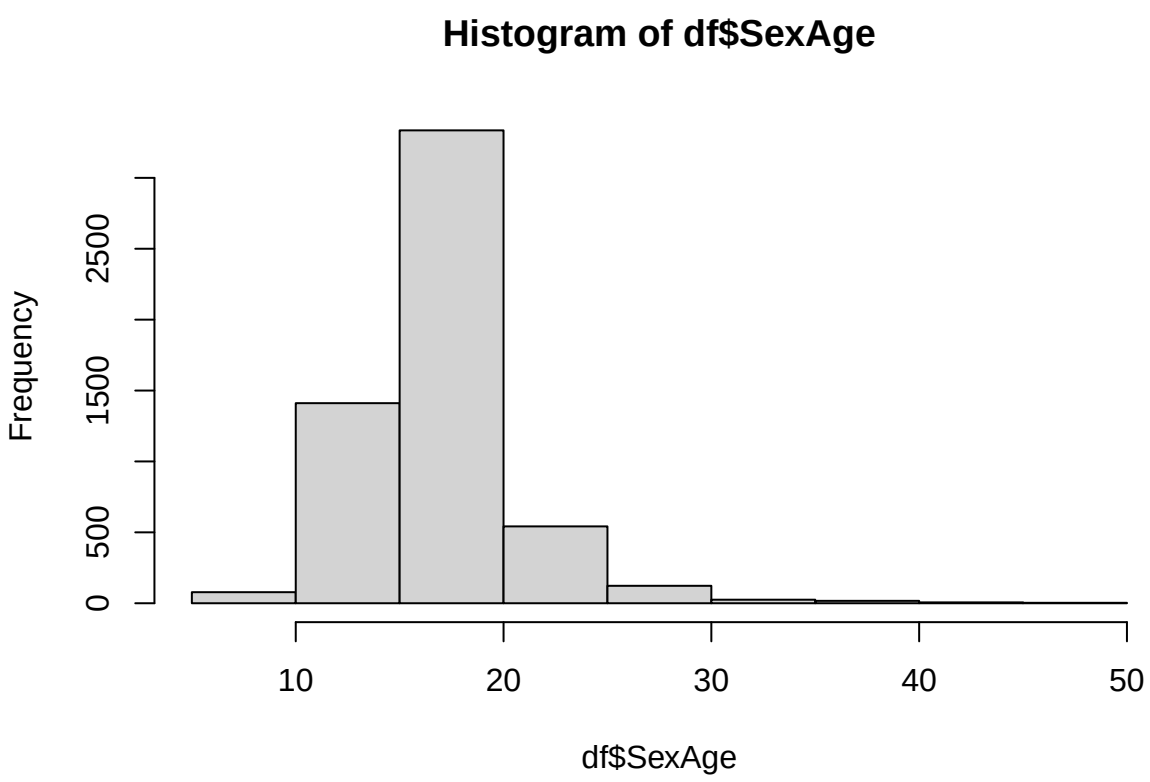
```
##
## Call:
## lm(formula = SexNumPartnLife ~ Gender + HHIncome + Education +
##     PhysActive + RegularMarij + HardDrugs + RegularMarij * HardDrugs,
##     data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43.88 -11.51  -4.29   2.76 985.61
##
## Coefficients:
##                               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
```

```
hist(df$SexAge)
```




```
sort(unique(df$SexAge))
```

```
##  [1]  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
## [26] 34 35 36 37 38 39 41 44 47 50
```

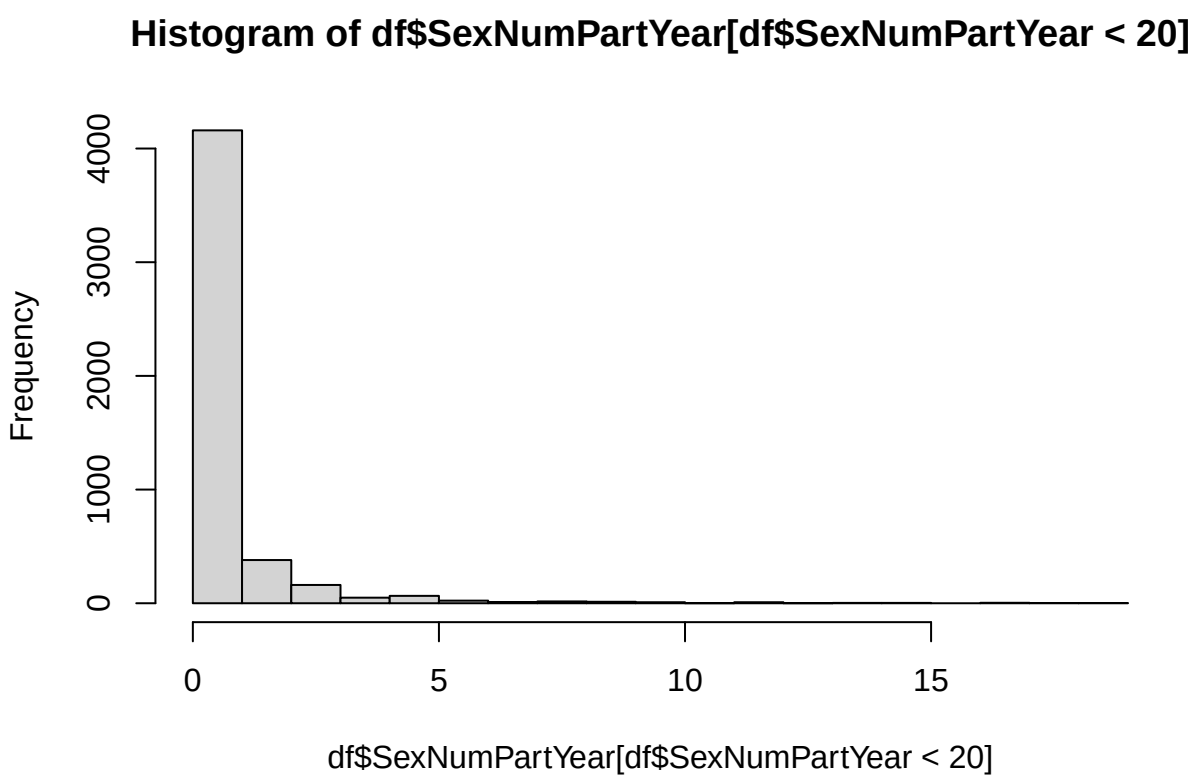
```
typeof(df$SexAge)
```

```
## [1] "integer"
```

```
subset(df, SexAge == 9 & !is.na(SexAge))$SexNumPartnLife
```

```
##  [1]  30  30  90  90  55  55 120   5   5   5   5  19   3   3   3   5   5   9  88
## [20]  98  27  27  25  30 150 150 150  NA   2  11  85 500 200 200   5   1  23   2
## [39]   8  19  20  20  20   3 100  50  40  40   6 360 150  20  80   3   3   3   5
## [58]  50   7
```

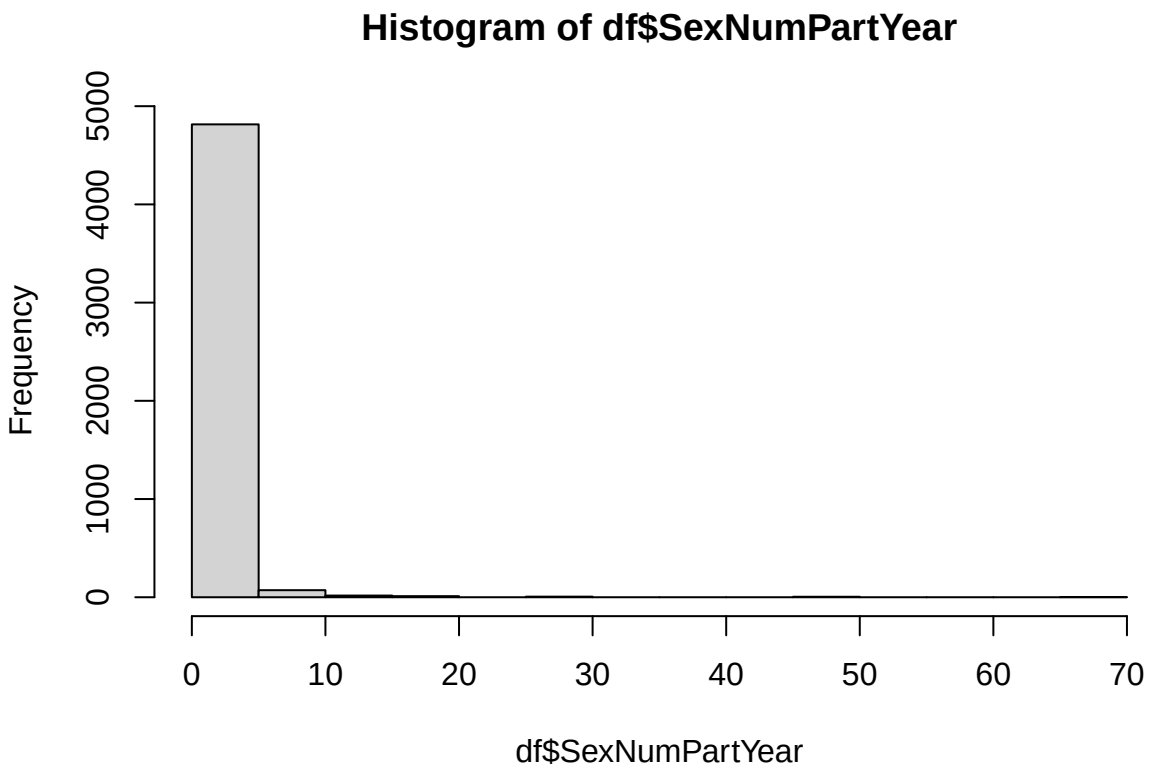
```
hist(df$SexNumPartYear[df$SexNumPartYear < 20])
```



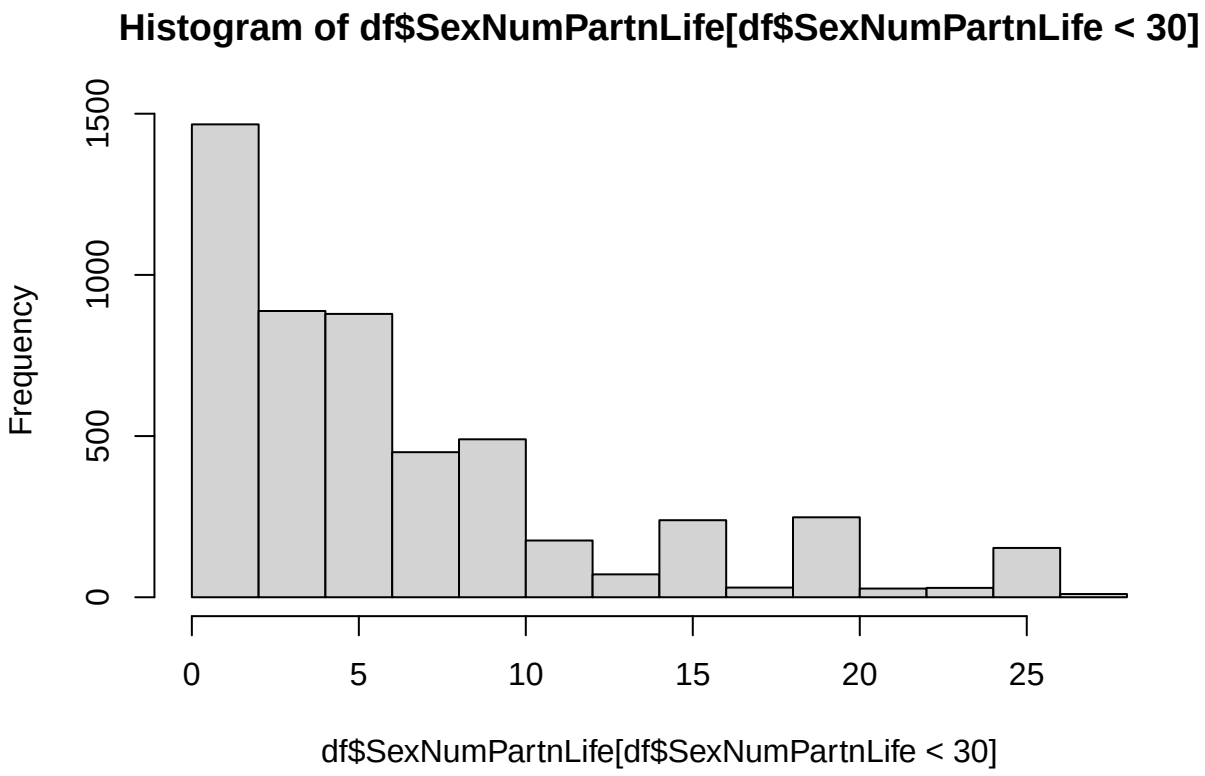

```
sort(unique(df$SexNumPartYear))
```

```
##  [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 17 18 19 20 30 50 69
```

```
hist(df$SexNumPartYear)
```



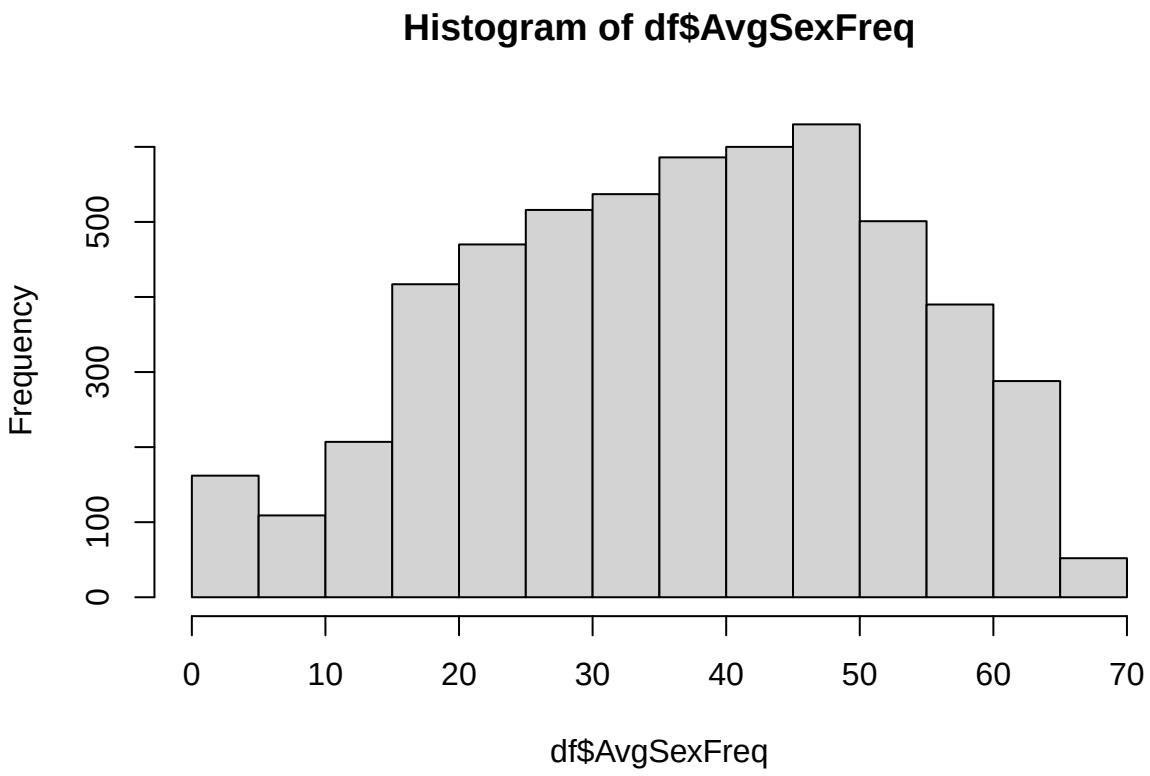

```
hist(df$SexNumPartnLife[df$SexNumPartnLife < 30])
```




```
unique(df$SexAge)
```

```
##  [1] 16 NA 12 13 17 22 27 20 18 14 23 15 21 24 28 30 19 32 29 26 37 33 35  9 38
## [26] 11 25 10 34 31 50 39 36 44 41 47
```

```
df = mutate(df, AvgSexFreq = Age-SexAge/SexNumPartnLife)
hist(df$AvgSexFreq)
```




```
df$AvgSexFreq[is.infinite(df$AvgSexFreq)] = NA
unique(df$AvgSexFreq)
```

```
##    [1] 32.00000       NA 47.80000 44.35000 64.86667 54.85714 53.88000  6.00000
##    [9] 40.00000 46.00000 53.40000 53.30000 15.00000 35.15000 42.75000 24.33333
##   [17] 35.00000 48.50000  7.00000 35.40000  4.00000 30.86667 22.50000 18.00000
##   [25] 23.00000 55.77778 24.00000 51.50000 25.06667 29.56667 33.00000 18.50000
##   [33] 57.00000 34.25000 31.33333 10.00000 20.00000 36.00000 36.75000 43.00000
##   [41] 39.00000  9.00000 44.00000 34.00000 29.28571 28.12500 50.50000 42.80000
##   [49] 45.33333 22.33333 54.20000 52.93333 37.85714 31.00000 27.00000  0.00000
##   [57] 16.00000 35.64444 37.00000 21.00000 41.75000 44.70000 32.71429 55.00000
##   [65]  1.00000 56.66667 37.80000 46.57778 30.50000 15.40000 63.16667  2.00000
##   [73] 57.20000 28.00000 55.52500 46.66667 46.83333 39.83333 60.33333 24.70000
##   [81] 63.60000 28.50000 60.94792 25.80000 41.72308 54.00000 23.50000 38.66667
##   [89] 39.57143 30.33333 42.50000 54.10000 59.75000 36.30000 45.75000 14.00000
##   [97] 53.36000 50.00000 19.30000 22.00000 37.33333 36.25000 24.44000 21.60000
##  [105] 26.85714 44.33333 53.77333 38.30000 30.11111 55.05000 11.00000 66.50000
##  [113] 16.22222 26.30000 38.40000 29.00000 45.00000 54.50000 44.52941 25.62500
##  [121] 16.66667 47.70000 27.66667 60.00000 24.50000 61.64000 59.10000 38.50000
##  [129] 28.20000 29.50000 41.00000 41.42857 21.75000 62.56667  3.00000 55.16667
##  [137] 16.80000 14.66667 28.40000 37.82667 51.87000 51.71429 10.50000 39.60000
##  [145] 35.93333 53.33333 16.83333 58.05000 56.22222 22.41667 62.66667 26.83333
##  [153] 38.35000 28.64286 34.31818 31.46667 47.00000 57.33333 38.57143 61.16667
##  [161] 61.40000 56.40000 45.80000 23.22222 58.33333 48.00000 42.28571 47.20000
##  [169] 17.40909 40.16667 38.51429 39.33333 26.00000 62.50000 36.50000 50.59524
##  [177] 28.83333 62.41667 57.70000 51.00000 33.40000 49.00000 33.06667 13.33333
```

```
##  [185] 23.83333 30.12500 34.50000 39.50000 23.12500 30.40000 40.25000 33.66667
##  [193] 21.11111 31.75000 52.00000 48.40000 30.00000 17.85714 49.50000 41.33333
##  [201] 34.46667 19.25000 64.40000 23.75000 50.40000 25.40000 52.21053 49.20000
##  [209] 17.33333 62.83333 56.88889 64.80000 18.80000 42.36364 48.60000 52.25000
##  [217] 61.00000 46.23810 21.88000 42.00000 36.33333 49.60000 29.66667 52.66667
##  [225] 20.33333 46.25000 34.33333 53.75000 22.70000 37.30000 59.25000 18.16667
##  [233] 41.40000 17.50000  5.00000 54.33333 19.28571 19.00000 29.42857 33.60000
##  [241] 26.10000 37.12500 67.77500 27.50000 64.50000 46.46667 25.00000 40.33333
##  [249] 27.57895 52.40000 38.63415 41.29167 56.20000 18.87500 49.86000 64.73333
##  [257] 27.40741 46.68000 26.66667 20.20000 23.20000 30.66667 59.00000 41.87500
##  [265] 48.75000 40.83333 12.66667 46.87500 53.71667 28.86667 38.11111 58.25000
##  [273] 38.75000 59.40000 42.66667 66.10000 40.20000 52.48571 27.80000 18.93333
##  [281] 51.57143 35.50000 17.00000 49.86667 53.00000 40.50000 21.37500 29.84000
##  [289] 52.33333 36.10000 59.28571 62.60976 42.58333 26.33333 61.80000 25.90000
##  [297] 25.85714 33.50000 51.10000 12.00000 59.76364 32.23810 47.15000 20.73333
##  [305] 22.30000 58.32000 27.40000 48.20000 49.28000 38.33333 28.66000 34.66667
##  [313] 22.60000 43.83333 48.66667 60.61538 39.84444 46.33333 43.48571 35.71429
##  [321] 59.87500 25.20000 36.90476 63.33333 43.80000 56.33333 18.25000 32.25000
##  [329] 19.67442 44.25000 33.30000 30.62500 65.66667 62.12500 28.10000 21.40000
##  [337] 26.20000 45.73333 57.66667 22.66667 66.25000 21.85714 39.66667 54.12500
##  [345] 49.36000 48.86667 43.10000 21.66667 26.13333 32.33333 24.57143 60.60000
##  [353] 50.20000 39.25000 44.93333 22.06667 62.94020 49.43333 39.12500 42.25000
##  [361] 28.36364 64.67273 39.20000 27.83636 50.90076 41.80000 13.00000 61.70000
##  [369] 42.06667 55.50000 23.66667 16.20000 66.00000 50.85714 43.40000 53.66667
##  [377] 31.42857 34.93333 15.50000 39.58333 35.60000 43.46667 36.42857 63.10000
##  [385] 24.60000 40.75000 50.60000 15.60000 24.40000 34.87778 18.35000 45.11111
##  [393] 59.05000 40.15000 61.46667 35.30435 33.70000 21.12500 40.48000 62.15000
##  [401] 60.87500 42.52000 25.50000 17.40000 27.88000 15.20000 66.70000 38.92500
##  [409] 53.93333 36.60000 18.58333 29.40000 51.20000 55.66667 21.33333 26.50000
##  [417] 40.66667 53.31818 41.66667 63.70000 31.10000 65.60000 28.73469 33.18182
##  [425] 15.80000 35.30000 63.66667 21.53333 45.86667 32.40000 30.75000 42.46667
##  [433] 26.60000 51.83333 32.35000 63.46667 43.20000 35.29167 38.76364 59.20000
##  [441] 25.46667 18.33333 55.55556 47.25000 63.00000 52.50000 52.80000 55.72000
##  [449] 45.25000 29.27778 48.71429 46.20000 60.10000 26.58333 37.16667 23.33333
##  [457] 28.87500 46.40000 48.73333 53.16667 35.66667 36.92308 17.80000 44.50000
##  [465] 51.75000 18.60000 20.87500 46.28000 55.83000 68.79000 37.20000 28.33333
##  [473] 31.06667 50.71429 36.74286 45.41667 22.80000 27.33333 39.36842  8.00000
##  [481] 32.20000 32.50000 27.56667 60.80000 61.13333 33.33333 63.66000 18.52632
##  [489] 57.86667 58.78571 65.34783 33.80000 34.80000 46.50000 38.25000 52.20000
##  [497] 27.35000 56.68000 30.20000 18.62500 58.00000 47.33333 50.75000 63.62500
##  [505] 46.36000 59.43333 56.73333 32.72000 62.55000 53.50000 36.80000 17.60000
##  [513] 36.20000 58.80000 25.70000 18.68750 30.60000 32.29167 25.77778 29.92308
##  [521] 31.71429 42.40000 31.80000 60.81000 48.58333 38.41667 14.33333 42.33333
##  [529] 40.60000 38.14286 58.62500 22.65000 29.33333 41.60000 37.93333 50.37500
##  [537] 55.83333 33.75000 45.76923 38.00000 58.44000 32.60000 31.50000 54.60000
##  [545] 20.50000 40.40000 66.44444 42.60000  9.50000 31.20000 62.89773 32.90000
##  [553] 62.75000 37.50000 37.76923 15.66667 60.78000 28.75000 39.46667 47.62222
##  [561] 16.50000 49.75000 20.80000 22.71429 58.91333 28.90816 17.55556 26.81818
##  [569] 58.66667 32.21053 36.57143 38.71429 27.10000 62.68000 26.25000 41.71429
##  [577] 67.86667 43.50000 56.50000 46.80000 26.12500 53.86667 37.37500 47.66667
##  [585] 29.80000 24.12500 39.36000 44.06667 68.28000 49.33333 43.33333 19.42857
##  [593] 29.69091 47.79012 24.81818 23.07143 54.41667 34.40000 26.44444 45.72000
##  [601] 33.57143 20.57143 43.44444 47.50000 37.24000 55.61765 26.36000 17.75000
##  [609] 59.35000 20.45000 57.50000 20.13333 56.75000 64.75000 38.15000 32.80000
```

```
##  [617] 28.57143 28.60000 53.80000 31.05556 62.62500 53.87500 19.40000 53.98700
##  [625] 56.62000 33.62000 56.72727 59.58333 21.87500 51.33333 42.20000 55.10000
##  [633] 36.70370 46.82955 58.50000 40.64000 50.86667 27.70000 26.94000 55.20000
##  [641] 29.30000 39.93333 41.83333 29.60000 54.40000 62.05000 33.16667 28.16667
##  [649] 42.93200 58.87500 48.25000 32.16667 30.16667 60.75000 28.71429 29.65000
##  [657] 47.57500 56.44118 24.23810 45.50000 52.89333 42.68000 47.72000 20.86667
##  [665] 53.48000 47.30000 35.36000 53.62500 50.11111 41.36364 49.48000 52.10000
##  [673] 31.53333 19.20000 27.71429 29.75000 39.68889 63.25000 18.55556 28.70000
##  [681] 39.85000 40.18182 13.50000 20.05882 31.64444 35.54286 41.20000 41.50000
##  [689] 39.30000 45.42857 37.25000 56.80000 26.80000 46.88889 22.42857 20.75000
##  [697] 28.25000 40.30000 35.55556 55.40000 45.15000 45.20000 29.10000 43.89412
##  [705] 34.57895 43.16667 54.36000 21.50000 39.32000 43.93333 46.10000 47.05000
##  [713] 55.45714 33.86000 58.20000 49.62857 51.42857 45.52000 32.16000 13.66667
##  [721] 62.00000 47.40000 51.40000 51.88696 43.75000 17.91667 23.91667 11.50000
##  [729] 20.85714 48.74000 17.12500 39.16667 30.44444 21.30000 17.66667 43.12500
##  [737] 54.25000 60.50000 20.30000 62.98200 39.82667 25.66667 50.33333 31.87500
##  [745] 56.30000 50.46667 52.30000 29.85714 35.20000 57.15000 39.64103 59.41667
##  [753] 25.15000 45.16667 54.80000 49.71429 60.66667 16.33333 53.76000 39.85714
##  [761] 30.05000 29.14286 52.88889 18.85714 27.21053 46.58333 25.13333 49.10000
##  [769] 17.44444 29.95500 27.87500 20.45455 20.06667 31.30000 50.30000 43.85714
##  [777] 59.81000 34.60000 61.20000 47.32000 37.66667 38.87500 56.98700 54.75000
##  [785] 63.50000 62.42857 46.86667 45.56667 21.48000 30.70000 41.97500 19.80000
##  [793] 20.16667 54.66667 33.58333 20.40000 25.33333 22.12500 48.06667 64.33333
##  [801] 25.57143 59.83333 38.20000 64.25000 40.87500 59.66667 40.35000 40.05882
##  [809] 58.86667 47.87500 51.66667 14.60000 35.12500 43.68333 41.58333 23.65714
##  [817] 37.75000 18.44444 40.84000 49.30000 48.57143 59.33333 35.25000 25.36000
##  [825] 63.75000 56.13333 40.85714 59.36667 63.82353 18.63636 56.43333 25.71429
##  [833] 43.64000 34.30000 43.66667 51.74000 49.85714 34.82051 52.68000 49.80000
##  [841] 38.16667 12.50000 21.62500 34.51429 21.47826 63.46154 59.80000 19.60000
##  [849] 32.36000 55.80000 15.33333 50.06667 19.92308 46.60000 28.85000 29.74000
##  [857] 62.90000 62.81000 20.51852 25.07143 60.86000 28.93333 51.25000 64.60870
##  [865] 34.57143 24.88182 31.14286 46.26667 57.64000 68.67500 35.11111 45.60000
##  [873] 34.83333 58.83333 52.87500 40.46667 37.40000 61.68000 26.43333 48.90000
##  [881] 36.76923 29.67500 35.87500 19.71429 66.80870 45.87500 36.45000 35.85714
##  [889] 48.98571 19.50000 27.60000 30.80000 27.66000 41.55000 56.32000 38.60000
##  [897] 28.85714 47.24000 23.80000 37.95667 51.36000 61.12500 34.85714 27.22222
##  [905] 45.40000 33.36364 59.60000 48.33333 39.80000 28.80000 47.88000 24.86667
##  [913] 37.83333 37.17647 23.86667 24.54545 54.87000 21.80000 29.25000 14.50000
##  [921] 17.87500 22.40000 64.60000 36.66667 44.40000 16.60000 44.10000 38.58333
##  [929] 44.46667 19.33333 30.66667 37.93000 41.28571 20.37500 56.14286 53.52000
##  [937] 15.25000 57.37500 62.60000 41.11111 47.66000 32.93333 53.83333 30.10000
##  [945] 49.66667 55.57143 60.46667 54.30000 42.12000 46.75000 56.00000 64.00000
##  [953] 31.66667 41.93333 62.33333 24.76923 18.70000 26.40000 40.68000 64.43333
##  [961] 25.75000 58.29167 61.87500 39.40000 21.20000 20.92857 18.57143 27.13333
##  [969] 59.50000 28.32000 66.41667 60.71667 34.94333 37.60000 47.81250 33.22222
##  [977] 43.30000 48.83333 17.21429 51.15000 56.25000 45.70000 30.74000 57.40000
##  [985] 65.00000 62.40000 32.58333 39.86000 63.20000 48.30000 18.20000 33.12500
##  [993] 49.38462 50.87500 40.70000 46.61538 29.34615 25.11111 65.40000 36.90000
## [1001] 59.85714 58.82000 31.96333 38.18182 41.25000 51.55556 20.66667 65.87500
## [1009] 46.12500 65.80000 55.25000 57.06667 18.28571 34.20000 60.20000 21.42857
## [1017] 66.37500 57.52500 49.65714 37.87500 27.45833 24.36364 27.93750 20.55556
## [1025] 51.16667 20.35000 50.25000 49.90000 17.20000 21.94333 42.55263 35.81333
## [1033] 44.66667 30.82105 45.28000 61.88889 68.40000 67.55000 51.80000 17.52632
## [1041] 28.90400 27.09091 47.42857 31.78750 37.95000 38.46667 53.78667 42.83333
```

```
## [1049] 43.57143 63.55000 33.42857 26.28571 40.22222 35.14286 34.75000 43.72000
## [1057] 39.05882 45.78571 16.25000 49.40000 42.30000 49.35000 62.16667 51.43333
## [1065] 20.36000 61.44444 67.39130 30.33333 57.80000 49.41667 46.43333 30.14286
## [1073] 32.73333 29.20000 62.25000 29.76923 47.58333 43.83750 34.55000 49.62791
## [1081] 19.83333 34.53125 28.83000 52.57143 28.15789 17.36000 30.37500 18.75000
## [1089] 60.25000 49.06667 48.36000 42.84615 23.16667 63.12500 67.10000 22.86667
## [1097] 34.42857 19.57143 24.25000 41.85000 21.84444 17.25000 19.33333 38.90000
## [1105] 53.70000 62.37143 21.72727 27.15000 32.62857 58.30000 60.84000 48.21053
## [1113] 42.90000 39.91000 35.10000 65.33333 23.73333 22.57143 55.38889 55.93750
## [1121] 31.40000 44.87500 17.45455 30.32000 23.41667 23.75926 64.70000 17.93333
## [1129] 55.33333 62.28000 61.90000 56.60000 23.06667 61.60000 22.96765 39.35000
## [1137] 53.57500 28.34783 62.51429 44.42857 44.63636 45.46154 36.40000 51.84000
## [1145] 56.93333 31.12500 59.94444 34.61290 34.87500 29.83333 44.60000 54.82000
## [1153] 32.46667 43.48000 46.14286 29.33333 60.28000 58.40000 18.06667 18.71429
## [1161] 44.86667 61.83333 24.10000 43.74286 21.25000 32.30000 52.73333 62.55556
## [1169] 27.57143 50.66667 55.35000 56.28571 58.75000 45.66667 38.87000 49.70000
## [1177] 35.80000 20.44444 24.66667 57.16667 31.15000 37.42857 67.67500 14.75000
## [1185] 30.25000 28.28571 23.30000 26.46429 56.16667 45.77500 22.11111 24.46154
## [1193] 41.51429 32.44000 38.28000 37.57143 31.85000 16.93333 51.18182 27.93333
## [1201] 44.62500 48.16667 57.76000 57.73333 55.30000 39.93750 21.36000 43.25000
## [1209] 33.36000 44.88000 52.82759 44.88889 24.87500 50.53333 23.42857 19.85714
## [1217] 30.28000 40.14286 62.82500 53.72000 22.14286 44.72727 45.30000 31.92308
## [1225] 26.84615 64.42857 41.62500 33.67391 57.93333 49.98500 39.40741 23.25000
## [1233] 62.79487 22.16667 63.86667 49.54286 44.36364 24.80000 48.61538 55.68000
## [1241] 46.78333 42.42857 58.22727 49.44444 35.44000 19.16667 23.35000 59.08000
## [1249] 53.98699 24.71429 34.12500 58.22222 16.57143 45.87000 20.53333 62.20000
## [1257] 57.36000 51.68000 63.87500 63.90000 52.83000 41.15000 40.11111 39.75000
## [1265] 16.70000 31.06250 60.83333 42.88000 63.15000 32.05000 44.21053 41.10000
## [1273] 56.45714 22.83333 22.25000 61.42857 61.68750 40.99350 32.66667 47.76000
## [1281] 34.63636 60.40000 38.70000 64.83838 27.85714 16.40000 19.66667 61.97500
## [1289] 20.12500 37.16000 41.30000 23.60000 52.85000 53.27273 56.11111 54.94000
## [1297] 35.77143 53.96750 51.15789 61.66000 60.76000 56.46429 48.80000 34.27273
## [1305] 54.83333 47.93333 25.35000 17.22222 64.95333 39.85057 34.16667 52.56000
## [1313] 36.93333 35.96000 58.88889 55.42857 54.53333 48.35000 56.92857 16.75000
## [1321] 54.13333 25.25000 33.20000 21.79167 30.78333 53.78788 21.44444 56.85714
## [1329] 53.81250 31.48571 23.40000 67.25000 17.34783 16.71429 24.44444 33.84000
## [1337] 17.71429 29.70000 29.46667 59.44000 52.16667 43.15000 50.79630 48.85714
## [1345] 43.46154 34.71429 33.93333 67.46667 48.62500 39.48000 56.71429 32.30435
## [1353] 53.55000 51.87500 40.80000 54.22222 36.18182 19.87500 56.62500 55.93333
## [1361] 53.73333 22.33333 27.88750 53.42857 58.10000 65.90000 61.54545 24.22222
## [1369] 28.66667 42.85714 36.12500 64.44444 24.75000 25.60000 53.91667 48.88000
## [1377] 27.25000 49.87500 62.84444 29.93333 49.98250 35.62500 44.53333 18.85455
## [1385] 31.86667 47.60000 63.40000 39.82000 21.58333 66.30000 19.75000 57.77647
## [1393] 42.53846 35.98000 48.91667 26.15000 39.23810 41.54545 26.93333 55.37037
## [1401] 39.22222 26.70000 36.58333 62.57143 44.58333 65.71429 18.11111 37.06667
## [1409] 61.50000 53.20000 52.84000 27.72500 15.75000
```

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

```
##
## Call:
## lm(formula = AvgSexFreq ~ Gender + HHIncome + Education + PhysActive +
##     SameSex + RegularMarij + HardDrugs + RegularMarij * HardDrugs,
##     data = df)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.477  -9.336   0.669   9.500  28.063
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    30.4123     1.9171  15.864  < 2e-16 ***
## Gendermale                      0.7954     0.4072   1.954  0.05081 .
## HHIncome 5000-9999              2.9511     2.0736   1.423  0.15476
## HHIncome10000-14999             2.6100     1.7674   1.477  0.13981
## HHIncome15000-19999             1.2376     1.7718   0.698  0.48491
## HHIncome20000-24999             0.9757     1.7466   0.559  0.57645
## HHIncome25000-34999             1.6320     1.6639   0.981  0.32675
## HHIncome35000-44999             3.5636     1.6627   2.143  0.03215 *
## HHIncome45000-54999             3.7606     1.6688   2.253  0.02428 *
## HHIncome55000-64999             6.6339     1.7007   3.901 9.74e-05 ***
## HHIncome65000-74999             5.5228     1.7226   3.206  0.00136 **
## HHIncome75000-99999             4.8601     1.6299   2.982  0.00288 **
## HHIncomemore 99999              6.8785     1.5955   4.311 1.66e-05 ***
## Education9 - 11th Grade         0.3910     1.2795   0.306  0.75994
## EducationHigh School           -1.3333     1.2203  -1.093  0.27464
## EducationSome College          -2.0428     1.2033  -1.698  0.08965 .
## EducationCollege Grad          -0.1608     1.2402  -0.130  0.89685
## PhysActiveYes                  -3.1296     0.4239  -7.383 1.86e-13 ***
## SameSexYes                     -1.9134     0.7800  -2.453  0.01420 *
## RegularMarijYes                 2.9732     0.5889   5.049 4.64e-07 ***
## HardDrugsYes                    8.4849     0.8255  10.279  < 2e-16 ***
## RegularMarijYes:HardDrugsYes   -3.0504     1.1015  -2.769  0.00564 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.76 on 4155 degrees of freedom
##   (5823 observations deleted due to missingness)
## Multiple R-squared:  0.09745,    Adjusted R-squared:  0.09289
## F-statistic: 21.36 on 21 and 4155 DF,  p-value: < 2.2e-16
```