BIOSTAT 650 Project

Jaehoon Kim

2024-11-17

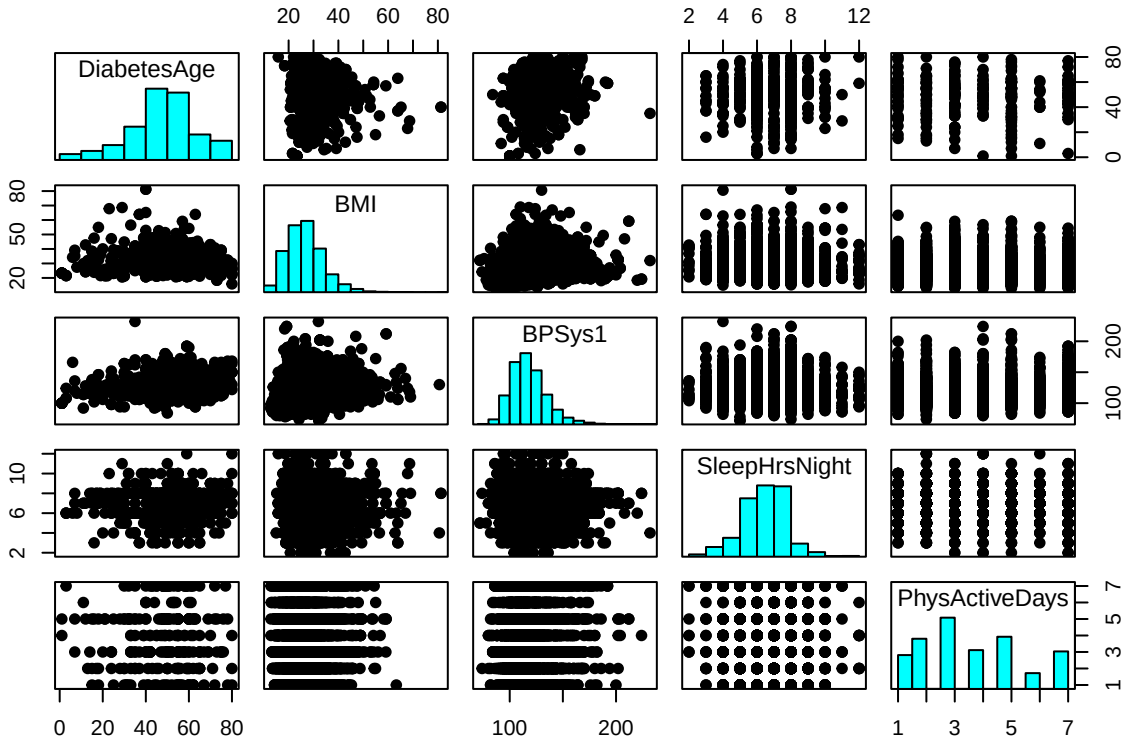
$1 + 1$

```
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
```



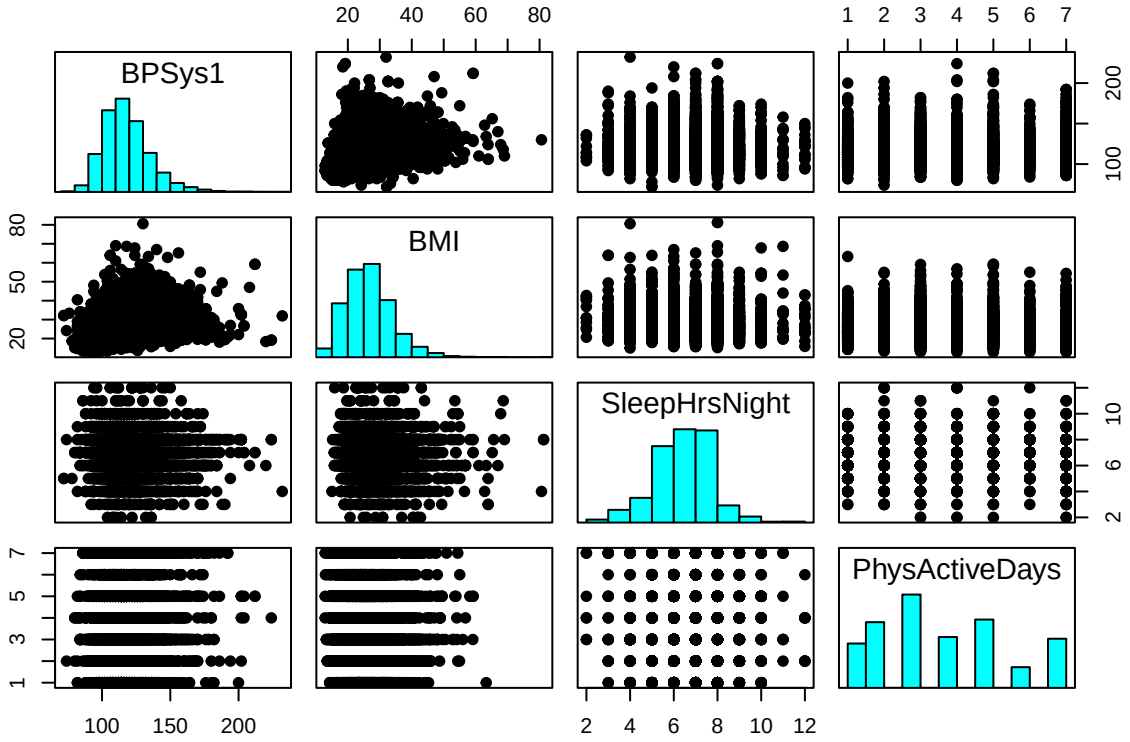

```
df = NHANES
#poverty gender useless?
colnames(df)
```

```
##   [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 ~ HardDrugs, df)
summary(model)
```

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

```
## lm(formula = SexAge ~ HardDrugs, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.804 -1.804 -0.804  1.196 32.196
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.80354    0.05435  327.59   <2e-16 ***
## HardDrugsYes -1.91152    0.12383  -15.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.631 on 5527 degrees of freedom
##   (4471 observations deleted due to missingness)
## Multiple R-squared:  0.04133,    Adjusted R-squared:  0.04116
## F-statistic: 238.3 on 1 and 5527 DF,  p-value: < 2.2e-16
```