TP3 - Logistic regression

Bealy MECH

10/18/2021

I. Health application: diagnostis of a heart attack

```
tab <- read.table(file="./SAheart.txt", header = TRUE, sep = ",", row.names = 1)
names(tab)
```

```
## [1] "sbp"       "tobacco"  "ldl"       "adiposity" "famhist"   "typea"
## [7] "obesity"   "alcohol"  "age"       "chd"
```

```
dim(tab) #dimension of data set
```

```
## [1] 462   10
```

```
head(tab)
```

```
##    sbp tobacco  ldl adiposity famhist typea obesity alcohol age chd
## 1 160   12.00 5.73     23.11 Present    49   25.30   97.20  52   1
## 2 144    0.01 4.41     28.61  Absent    55   28.87    2.06  63   1
## 3 118    0.08 3.48     32.28 Present    52   29.14    3.81  46   0
## 4 170    7.50 6.41     38.03 Present    51   31.99   24.26  58   1
## 5 134   13.60 3.50     27.78 Present    60   25.99   57.34  49   1
## 6 132    6.20 6.47     36.21 Present    62   30.77   14.14  45   0
```

```
tab$famhist <- as.numeric(tab$famhist == "Present")
head(tab)
```

```
##    sbp tobacco  ldl adiposity famhist typea obesity alcohol age chd
## 1 160   12.00 5.73     23.11       1    49   25.30   97.20  52   1
## 2 144    0.01 4.41     28.61       0    55   28.87    2.06  63   1
## 3 118    0.08 3.48     32.28       1    52   29.14    3.81  46   0
## 4 170    7.50 6.41     38.03       1    51   31.99   24.26  58   1
## 5 134   13.60 3.50     27.78       1    60   25.99   57.34  49   1
## 6 132    6.20 6.47     36.21       1    62   30.77   14.14  45   0
```
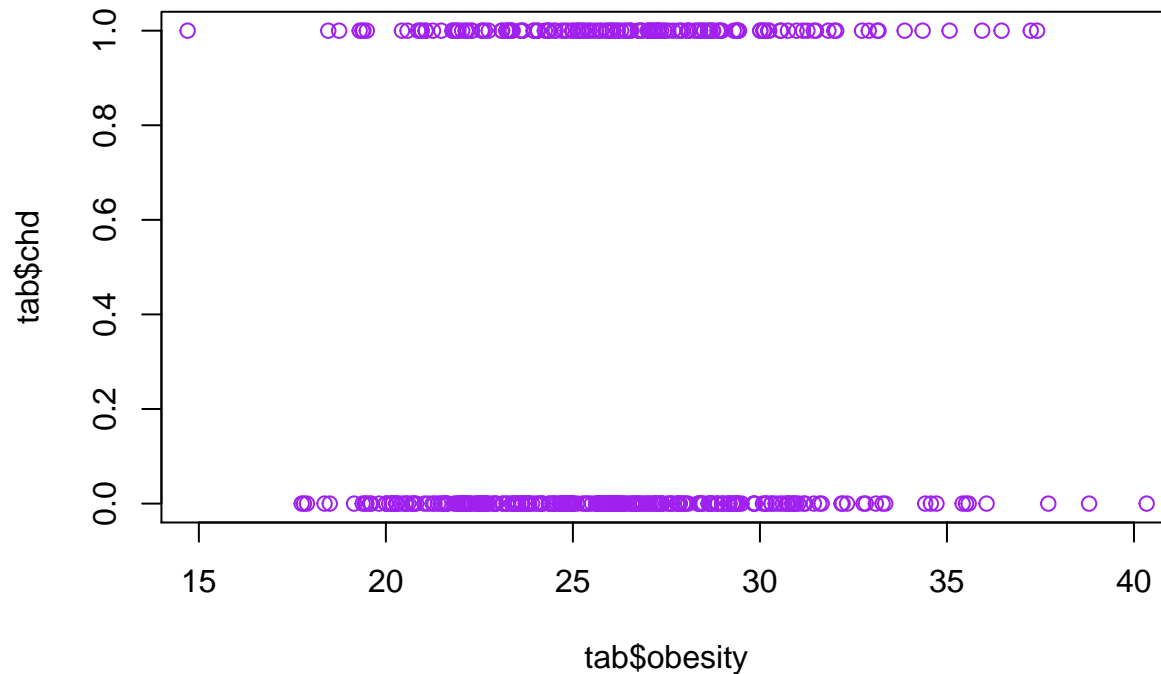
```
plot(tab, pch=20, col=rgb(0, 1, 0.5))
```
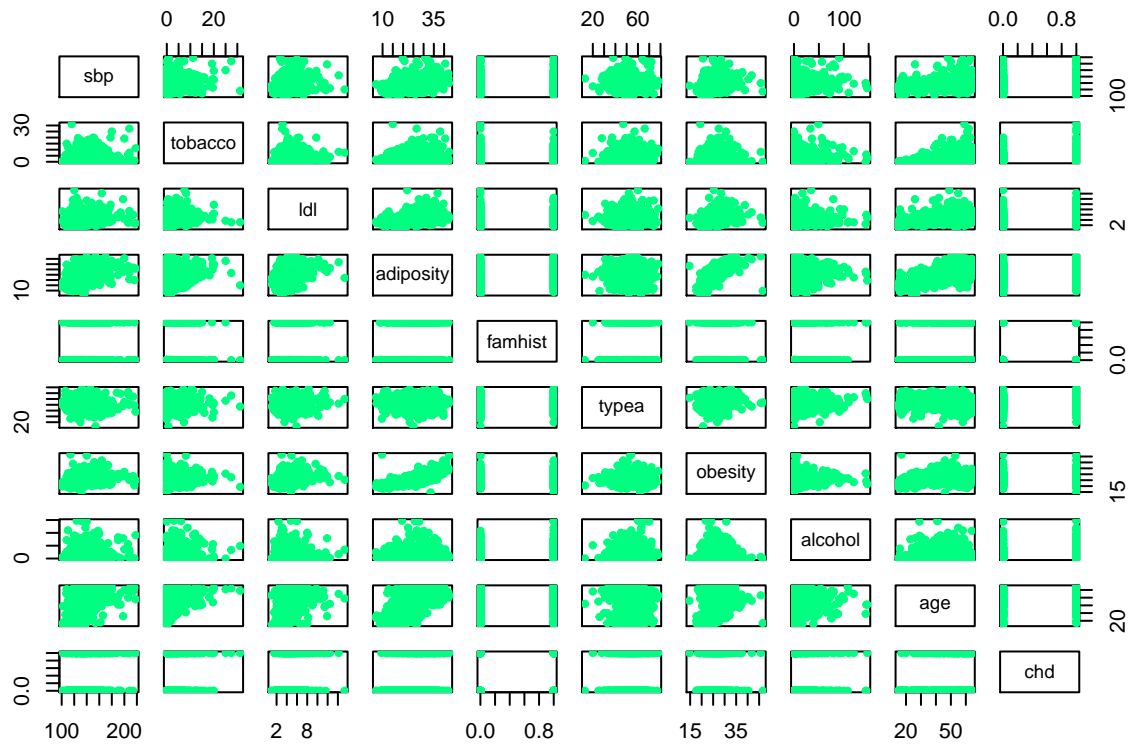
```
pairs(tab[,-5],pch=22,bg=c("red","blue")[unclass(factor(tab[,"chd"]))])
```
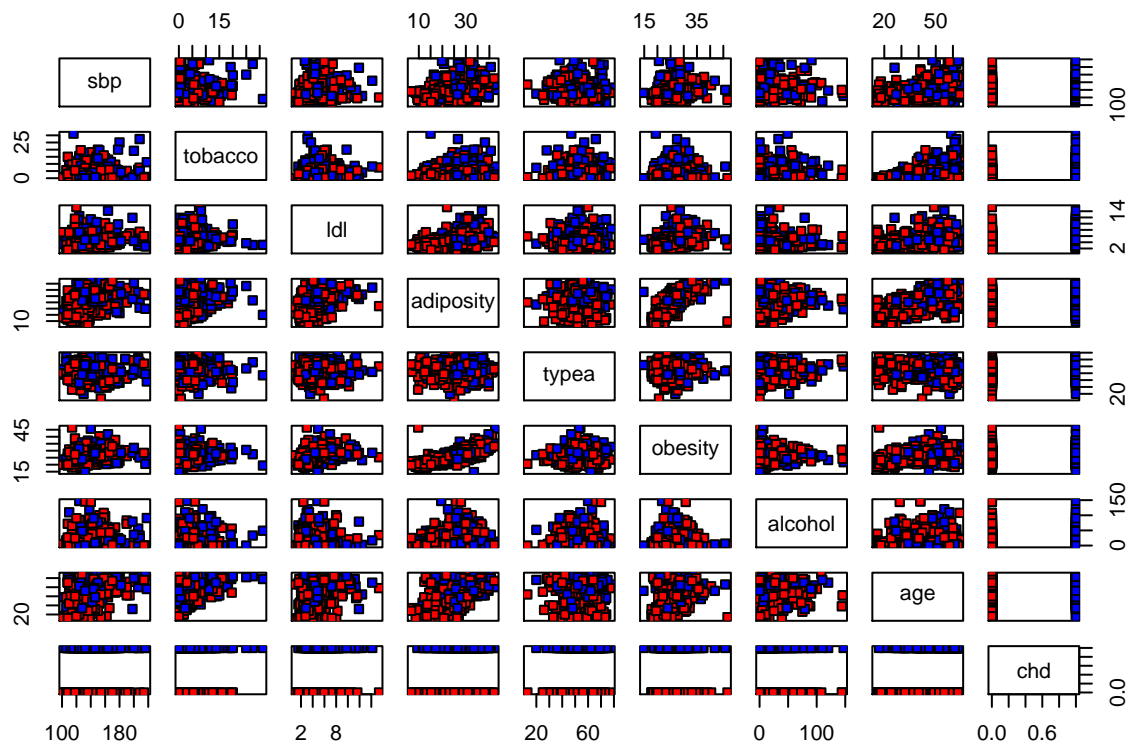


```
plot(tab$chd~tab$obesity,xlim=c(15,40), col="purple")
```

```
cor(tab)
```

```
##                   sbp     tobacco          ldl   adiposity     famhist
## sbp        1.00000000  0.21224652  0.15829633  0.35650008 0.08564531
## tobacco    0.21224652  1.00000000  0.15890546  0.28664037 0.08860143
## ldl        0.15829633  0.15890546  1.00000000  0.44043175 0.16135306
## adiposity  0.35650008  0.28664037  0.44043175  1.00000000 0.18172101
## famhist    0.08564531  0.08860143  0.16135306  0.18172101 1.00000000
## typea     -0.05745431 -0.01460788  0.04404758 -0.04314364 0.04480858
## obesity    0.23806661  0.12452941  0.33050586  0.71655625 0.11559508
## alcohol    0.14009559  0.20081339 -0.03340340  0.10033013 0.08051969
## age        0.38877060  0.45033016  0.31179923  0.62595442 0.23966742
## chd        0.19235411  0.29971754  0.26305268  0.25412139 0.27237273
##                 typea      obesity     alcohol         age         chd
## sbp       -0.05745431 0.23806661  0.14009559  0.3887706 0.19235411
## tobacco   -0.01460788 0.12452941  0.20081339  0.4503302 0.29971754
## ldl        0.04404758 0.33050586 -0.03340340  0.3117992 0.26305268
## adiposity -0.04314364 0.71655625  0.10033013  0.6259544 0.25412139
## famhist    0.04480858 0.11559508  0.08051969  0.2396674 0.27237273
## typea      1.00000000 0.07400610  0.03949794 -0.1026063 0.10315583
## obesity    0.07400610 1.00000000  0.05161957  0.2917771 0.10009508
## alcohol    0.03949794 0.05161957  1.00000000  0.1011246 0.06253068
## age       -0.10260632 0.29177713  0.10112465  1.0000000 0.37297334
## chd        0.10315583 0.10009508  0.06253068  0.3729733 1.00000000
```

```
library(corrplot)
```

```
## corrplot 0.90 loaded
```

```
corrplot(cor(tab), is.corr = TRUE, method = "ellipse", number.cex = .6, addCoef.col = "black")
```