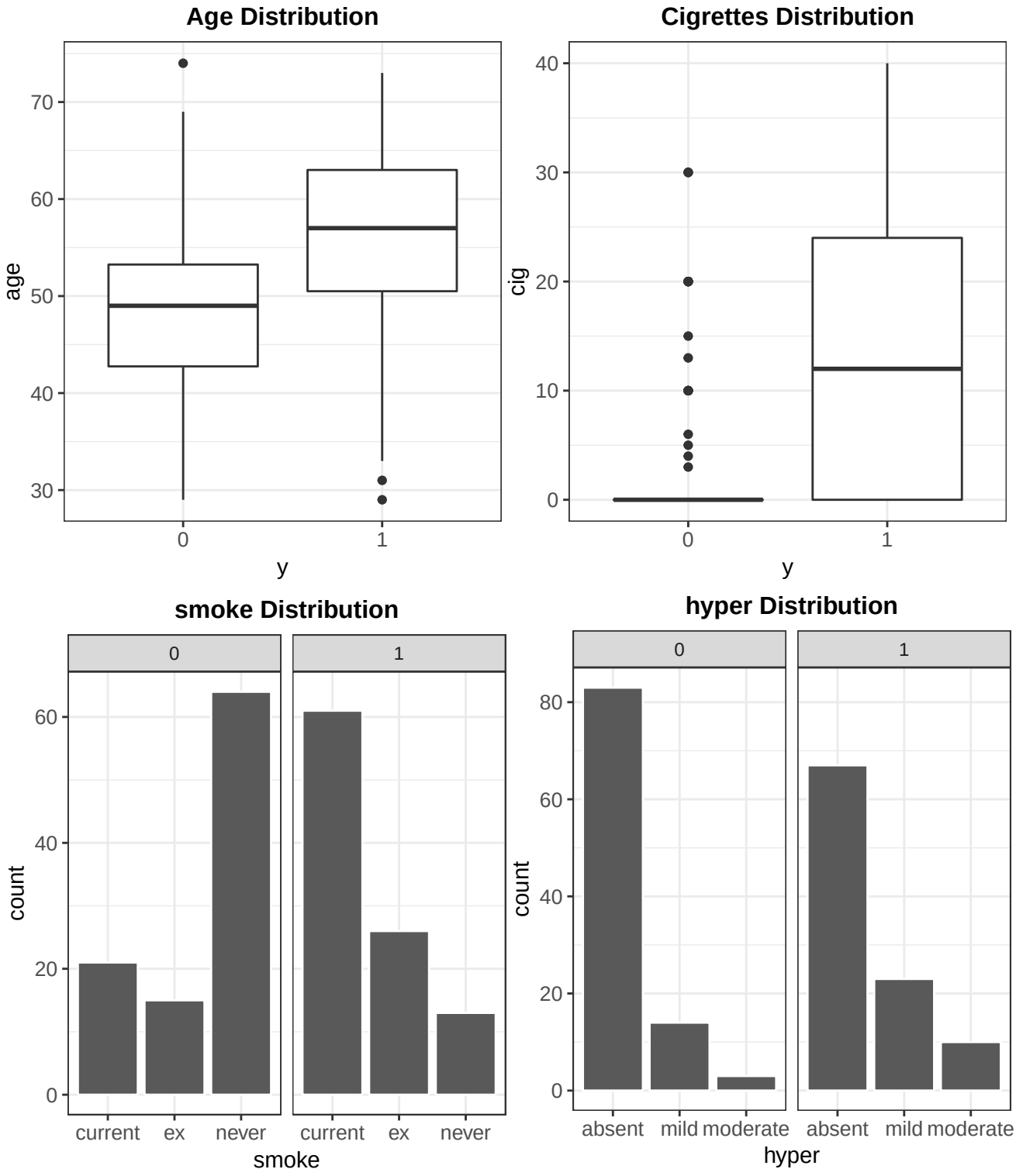
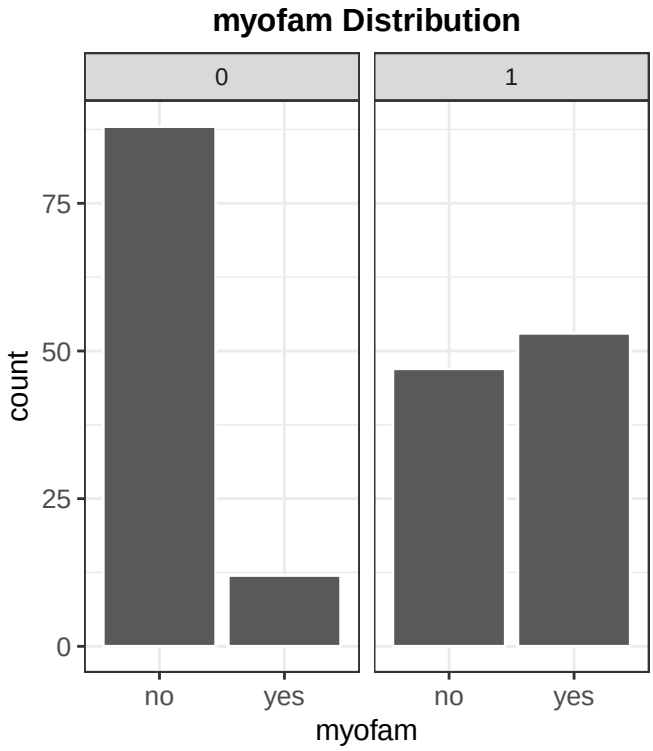
Project II - Angina

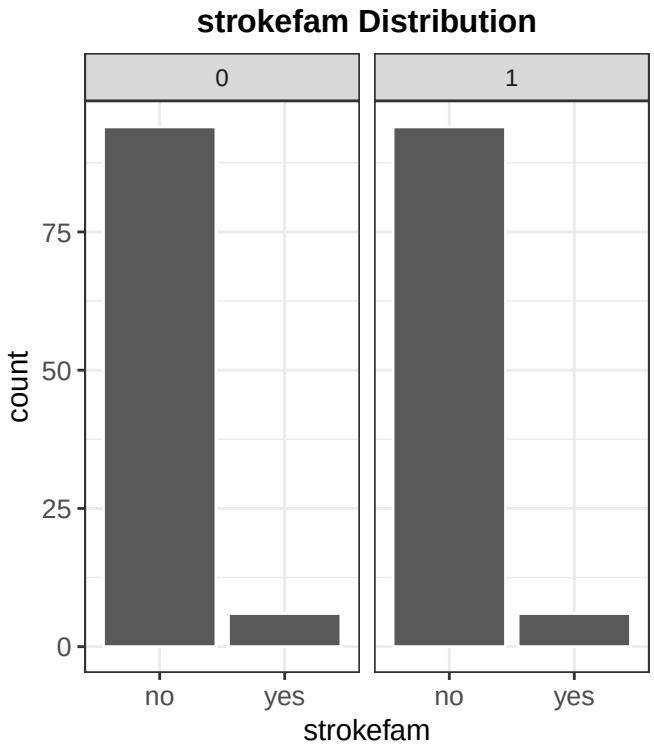
Yuhan Ning 915486450 Dandi Peng 915553480

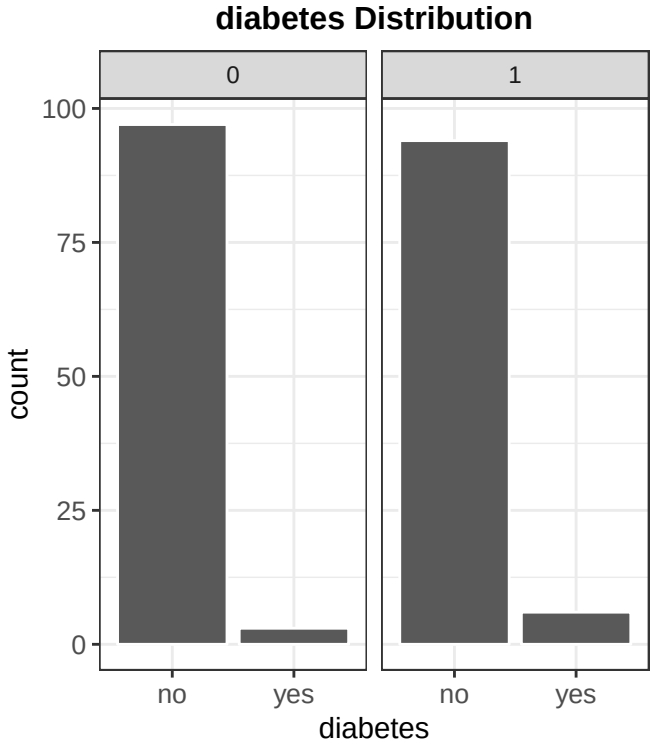
3/1/2019

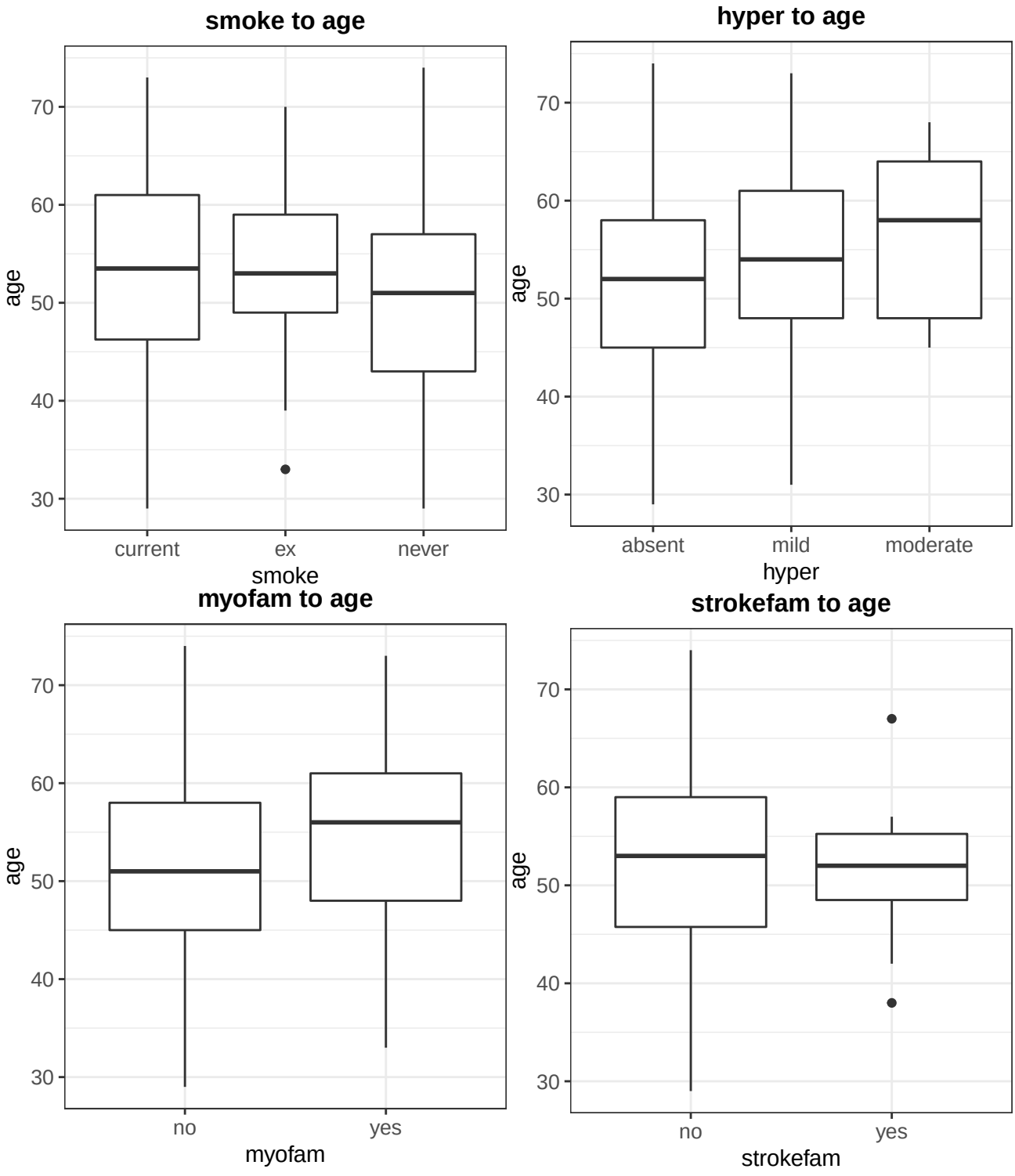
I. Summary

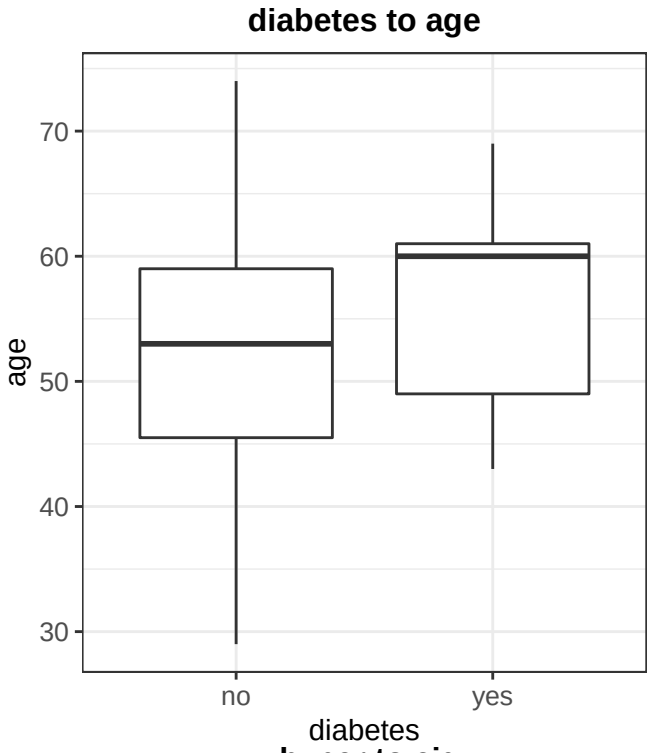

Age Distribution
age
y
Cigrettes Distribution
cig
y
smoke Distribution
count
smoke
current
ex
never
hyper Distribution
count
hyper
absent
mild
moderate

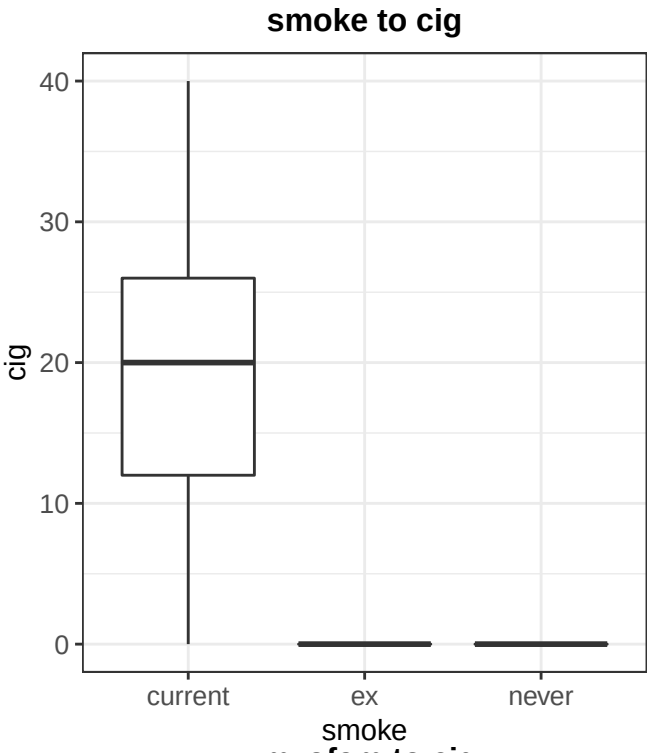

myofam Distribution
0
1
75
50
25
0
count
no
yes
no
yes
myofam

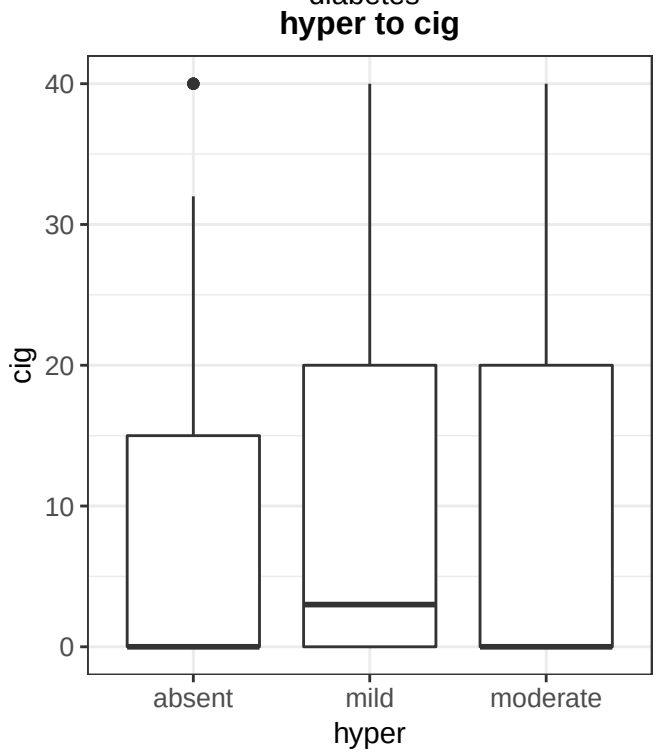

strokefam Distribution
0
1
75
50
25
0
count
no
yes
no
yes
strokefam

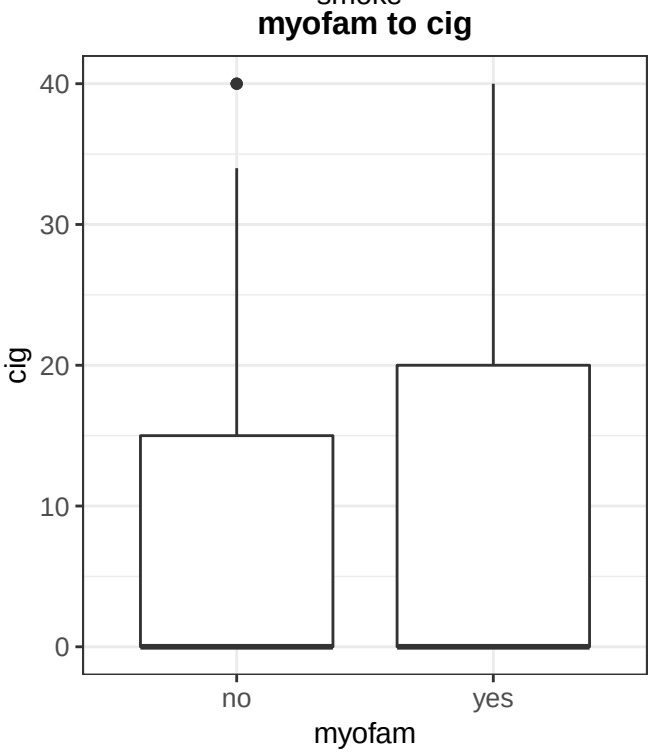

diabetes Distribution
0
1
100
75
50
25
0
count
no
yes
no
yes
diabetes

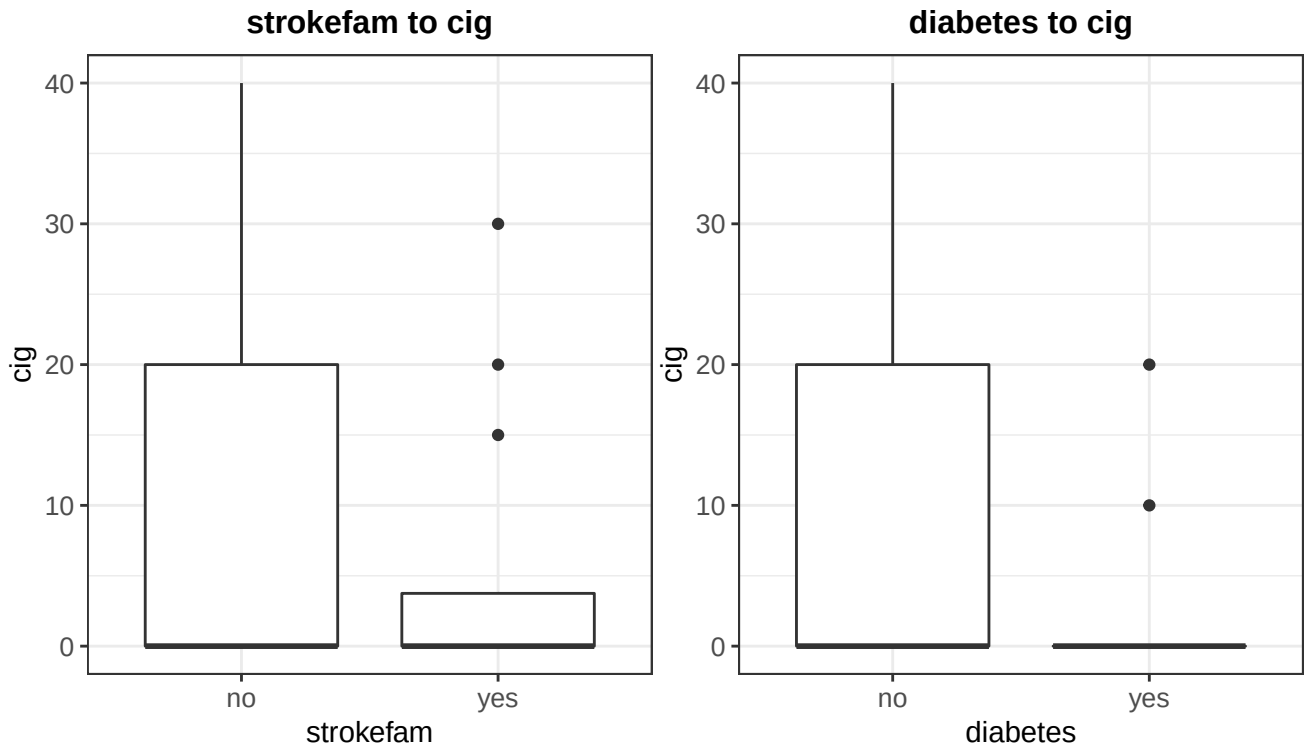

smoke to age
age
70
60
50
40
30
current
ex
never
smoke
hyper to age
age
70
60
50
40
30
absent
mild
moderate
hyper
myofam to age
age
70
60
50
40
30
no
yes
myofam
strokefam to age
age
70
60
50
40
30
no
yes
strokefam


diabetes to age
age
70
60
50
40
30
no
yes
diabetes

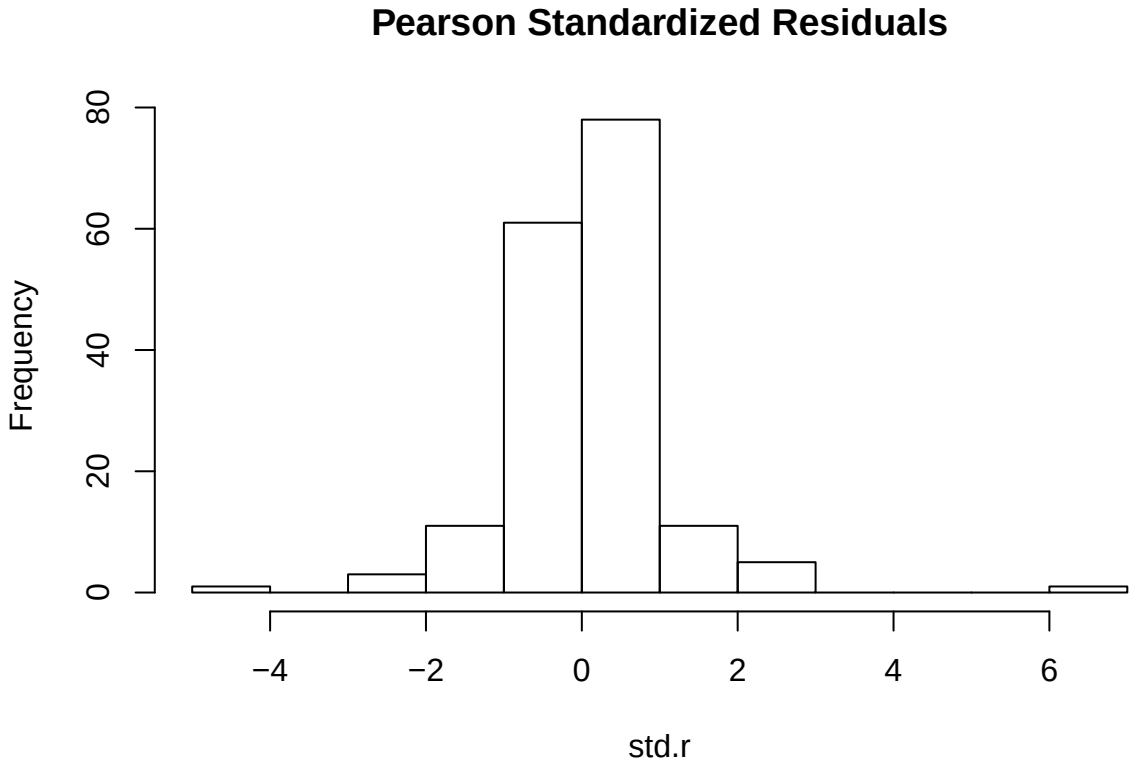

smoke to cig
cig
40
30
20
10
0
current
ex
never
smoke


hyper to cig
cig
40
30
20
10
0
absent
mild
moderate
hyper


myofam to cig
cig
40
30
20
10
0
no
yes
myofam

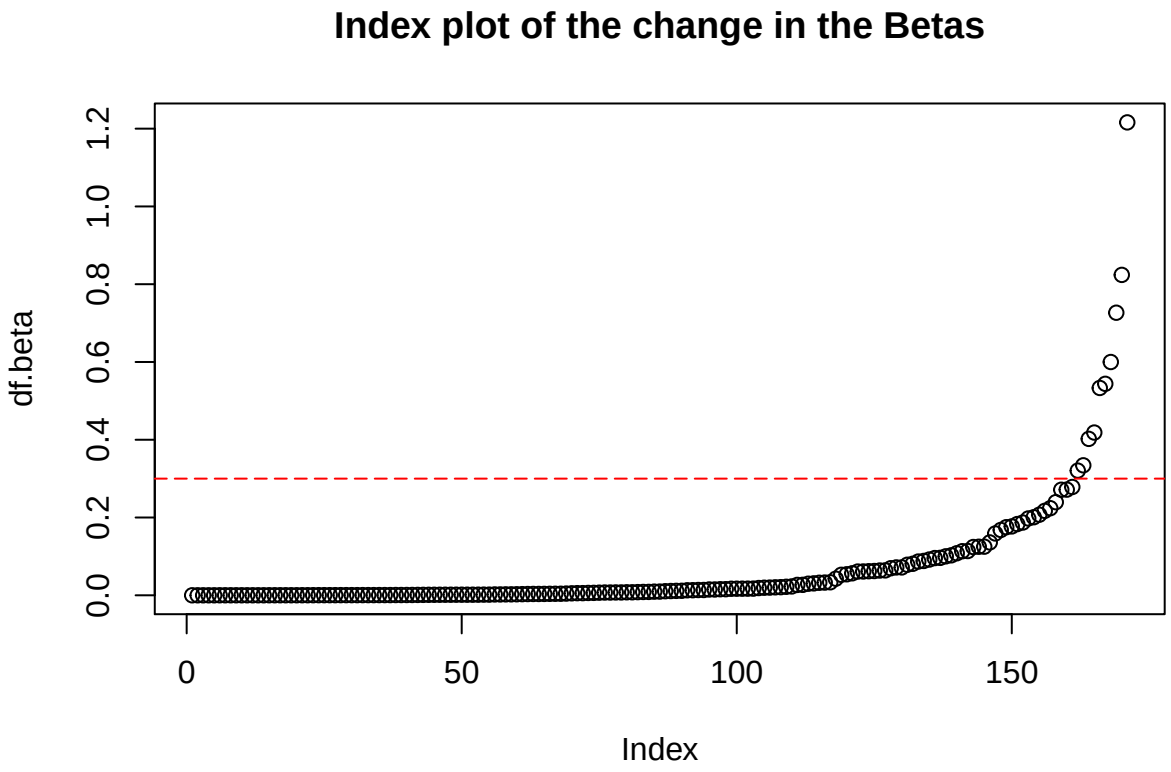


II. Data Preparation




```
##     (Intercept) age smokeex smokenever cig hypermild hypermoderate
## 158           1  38       0          1   0         0             0
```

```
##     myofamyes strokefamyes diabetesyes y       Pr
## 158         0            0           0 1 6.484827
```




```
##     (Intercept) age smokeex smokenever cig hypermild hypermoderate
## 162           1  54       0          0  30         0             0
## 163           1  47       1          0   0         1             0
## 164           1  29       0          0  20         0             0
## 165           1  42       0          0  30         0             0
## 166           1  45       0          0  30         0             0
## 167           1  45       1          0   0         0             1
## 168           1  48       0          1   0         0             1
## 169           1  59       1          0   0         0             0
## 170           1  58       1          0   0         0             1
## 171           1  43       0          0  20         0             0
##     myofamyes strokefamyes diabetesyes y     dBhat
## 162         0            0           0 0 0.3205503
## 163         0            1           0 1 0.3341963
## 164         0            0           0 1 0.4022270
## 165         0            1           0 0 0.4184973
## 166         1            0           0 0 0.5331769
## 167         0            0           0 0 0.5438178
## 168         0            0           0 1 0.5999283
## 169         0            0           1 1 0.7266942
## 170         0            0           0 0 0.8239320
## 171         1            0           1 0 1.2161134
```

III. Model Selection/Analysis

```
##   (Intercept)       smokeex    smokenever           age     myofamyes
```

```
##    -7.2767960     0.6571343    -1.3360134     0.1091857     2.3901902
##     hypermild hypermoderate           cig
##     1.3132608     2.1142710     0.1025953
```

```
##                                     LL p   n      AIC      BIC
## y~age+myofam+hyper+cig       -79.49318 6 200 170.9864 190.7763
## y~smoke+age+myofam+hyper+cig -72.67029 8 200 161.3406 187.7271
```

$$H_0 : \beta_{2,ex} = 0,\ \beta_{2,nv} = 0$$

$$H_a :\ at\ least\ one\ of\ \beta_{2,i} \neq 0$$

Based on the above output, the test statistics is $G^2 = -2(L_0 - L_1) = -2(-79.49318 - (-72.67029)) = 13.64578$, and the d.f. $= 8 - 6 = 2$.

The corresponding p value $= P(\chi_2^2 > G^2) = 0.00108857$, which is less than any $\alpha's$, therefore, we reject the null hypothesis and cannot drop the smoke variable.

$$H_0 : \beta_{2,ex} = 0, \quad H_a : \beta_{2,ex} \neq 0$$

The Wald test statistics is $\frac{\hat{\beta_{2,ex}} - 0}{SE(\hat{\beta_{2,ex}})} = \frac{0.65713}{0.83649} = 0.786$, and its corresponding p value is $P(Z^2 > 0.786) = 0.43211$, which is large than any $\alpha's$, therefore, we fail to reject the null hypothesis and can drop $\beta_{2,ex}$.

Combined the above two tests, we can conclude that smoke varaible should be contained, but there is no significant difference between "ex" and "current" smoking status and we can merge them to be one level - 'some history smoking' vs. the rest 'never smoked'.

```
##   (Intercept)   smokesmoked           age     myofamyes           cig
##   -8.69170240    1.84164976    0.11041110    2.41749918    0.07969663
##     hypermild hypermoderate
##    1.29134347    2.16431152
```

It is satisfying to find out that all variables are significant.

Now let's go to the interaction check.

```
##                                                    LL p   n      AIC
## y~smoke+age+myofam+hyper+cig                   -72.98329 7 200 159.9666
## y~smoke+age+myofam+hyper+cig+smoke*age+smoke*cig -72.44906 9 200 162.8981
##                                                     BIC
## y~smoke+age+myofam+hyper+cig                   183.0548
## y~smoke+age+myofam+hyper+cig+smoke*age+smoke*cig 192.5830
```

Based on above output, it is unnecessary to include the interactions.

Therefore, our final best "model correction" regression function is:

$$In(\frac{\pi}{1-\pi}) = -8.69170 + 0.11041X_1 + 1.84165X_{2,smoked} + 0.0797X_3 + 1.29134X_{4,mild} + 2.16431X_{4,mod} + 2.41750X_{5,yes}$$

where X_1 : age, $X_{2,i}$: smoke, X_3 : cig, $X_{4,i}$: hyper, and $X_{5,i}$: myofam.

IV. Interpretation

```
## Waiting for profiling to be done...
```

```
##                       2.5 %     97.5 %
## (Intercept)   -12.04681314 -5.8318438
## smokesmoked     0.82715688  2.9332006
## age             0.06262450  0.1638984
## myofamyes       1.48378884  3.4553161
## cig             0.03471212  0.1307326
## hypermild       0.24284545  2.3953017
## hypermoderate   0.50582853  4.0255687
```

$exp(0.11041) = 1.1167$: When the age of a subject increases 1 unit, the estimated odds that the subject had angina is multiplied by 1.1167, holding the other variables constant.

$exp(1.84165) = 6.3069$: The estimated odds that a subject who has some smoking history for certain hypertension history and certain myocardial infarction history had angina is 6.3069 times that who has never smoked was.

age and **myofam** (myocardial infarction)

V. Prediction

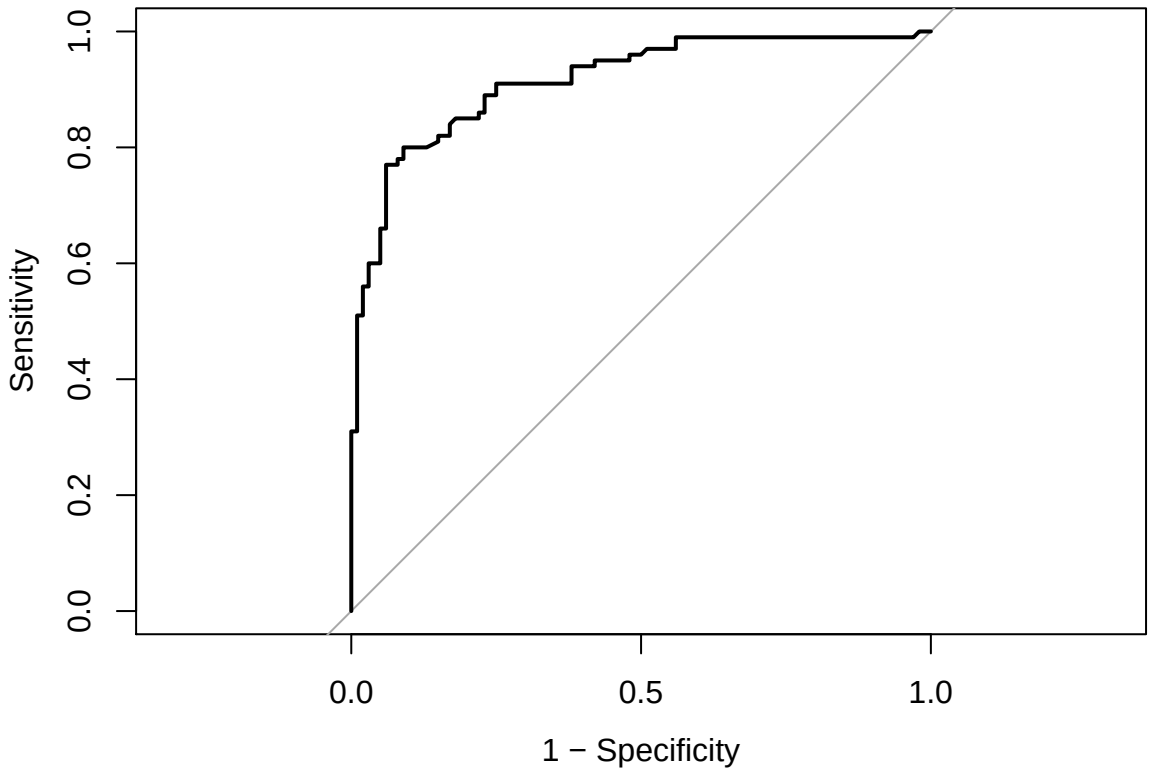
```
##      predicted
## truth  0  1
##     0 83 17
##     1 17 83
```

```
## Sensitivity Specificity  Error-Rate
##        0.83        0.83        0.17
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```




```
## Area under the curve: 0.915
## 95% CI: 0.8762-0.9539 (DeLong)
##          1
## 0.04026606
```

VI. Conclusion