LDA final project

Kaitlin Maciejewski, Morgan de Ferrante, Bingnan Li, Volha Tryputsen

```r
# save data in separate folder "final_data" one directory up
data <- read.csv("../final_data/frmgham2.csv")

library(psych)
describe(data)[,c(2,3,4,5,8,9,10,13)]
```

```
##                  n       mean         sd     median       min        max
## RANDID       11627 5004740.92 2900877.44 5006008.00   2448.00 9999312.0
## SEX          11627       1.57       0.50       2.00      1.00        2.0
## TOTCHOL      11218     241.16      45.37     238.00    107.00      696.0
## AGE          11627      54.79       9.56      54.00     32.00       81.0
## SYSBP        11627     136.32      22.80     132.00     83.50      295.0
## DIABP        11627      83.04      11.66      82.00     30.00      150.0
## CURSMOKE     11627       0.43       0.50       0.00      0.00        1.0
## CIGPDAY      11548       8.25      12.19       0.00      0.00       90.0
## BMI          11575      25.88       4.10      25.48     14.43       56.8
## DIABETES     11627       0.05       0.21       0.00      0.00        1.0
## BPMEDS       11034       0.09       0.28       0.00      0.00        1.0
## HEARTRTE     11621      76.78      12.46      75.00     37.00      220.0
## GLUCOSE      10187      84.12      24.99      80.00     39.00      478.0
## educ         11332       1.99       1.03       2.00      1.00        4.0
## PREVCHD      11627       0.07       0.26       0.00      0.00        1.0
## PREVAP       11627       0.05       0.23       0.00      0.00        1.0
## PREVMI       11627       0.03       0.18       0.00      0.00        1.0
## PREVSTRK     11627       0.01       0.11       0.00      0.00        1.0
## PREVHYP      11627       0.46       0.50       0.00      0.00        1.0
## TIME         11627    1957.02    1758.78    2156.00      0.00     4854.0
## PERIOD       11627       1.90       0.81       2.00      1.00        3.0
## HDLC          3027      49.36      15.63      48.00     10.00      189.0
## LDLC          3026     176.47      46.86     173.00     20.00      565.0
## DEATH        11627       0.30       0.46       0.00      0.00        1.0
## ANGINA       11627       0.16       0.37       0.00      0.00        1.0
## HOSPMI       11627       0.10       0.30       0.00      0.00        1.0
## MI_FCHD      11627       0.15       0.36       0.00      0.00        1.0
## ANYCHD       11627       0.27       0.44       0.00      0.00        1.0
## STROKE       11627       0.09       0.29       0.00      0.00        1.0
## CVD          11627       0.25       0.43       0.00      0.00        1.0
## HYPERTEN     11627       0.74       0.44       1.00      0.00        1.0
## TIMEAP       11627    7241.56    2477.78    8766.00      0.00     8766.0
## TIMEMI       11627    7593.85    2136.73    8766.00      0.00     8766.0
## TIMEMIFC     11627    7543.04    2192.12    8766.00      0.00     8766.0
## TIMECHD      11627    7008.15    2641.34    8766.00      0.00     8766.0
## TIMESTRK     11627    7660.88    2011.08    8766.00      0.00     8766.0
## TIMECVD      11627    7166.08    2541.67    8766.00      0.00     8766.0
## TIMEDTH      11627    7854.10    1788.37    8766.00     26.00     8766.0
## TIMEHYP      11627    3598.96    3464.16    2429.00      0.00     8766.0
##                  range         se
## RANDID      9996864.00   26902.68
## SEX               1.00       0.00
```

```
## TOTCHOL     589.00     0.43
## AGE          49.00     0.09
## SYSBP       211.50     0.21
## DIABP       120.00     0.11
## CURSMOKE      1.00     0.00
## CIGPDAY      90.00     0.11
## BMI          42.37     0.04
## DIABETES      1.00     0.00
## BPMEDS        1.00     0.00
## HEARTRTE    183.00     0.12
## GLUCOSE     439.00     0.25
## educ          3.00     0.01
## PREVCHD       1.00     0.00
## PREVAP        1.00     0.00
## PREVMI        1.00     0.00
## PREVSTRK      1.00     0.00
## PREVHYP       1.00     0.00
## TIME       4854.00    16.31
## PERIOD        2.00     0.01
## HDLC        179.00     0.28
## LDLC        545.00     0.85
## DEATH         1.00     0.00
## ANGINA        1.00     0.00
## HOSPMI        1.00     0.00
## MI_FCHD       1.00     0.00
## ANYCHD        1.00     0.00
## STROKE        1.00     0.00
## CVD           1.00     0.00
## HYPERTEN      1.00     0.00
## TIMEAP     8766.00    22.98
## TIMEMI     8766.00    19.82
## TIMEMIFC   8766.00    20.33
## TIMECHD    8766.00    24.50
## TIMESTRK   8766.00    18.65
## TIMECVD    8766.00    23.57
## TIMEDTH    8740.00    16.59
## TIMEHYP    8766.00    32.13
```