Main Analysis

Group C

Data Description

The UCLA Nurse Blood Pressure Study collected information from registered nurses in the Los Angeles area between 24 and 50 years of age on blood pressure (BP) and potential factors that contribute to hypertension. This information includes family history, and whether the subject had one or two hypertensive parents, as well as a wide range of measures of the physical and emotional condition of each nurse throughout the day. Researchers sought to study the links between BP and family history, personality, mood changes, working status, and menstrual phase. The first BP measurement was taken half an hour before the subject's normal start of work, and BP was then measured approximately every 20 minutes for the rest of the day. At each BP reading, the nurses also rate their mood on several dimensions, including how stressed they feel at the moment the BP is taken. In addition, the activity of each subject during the 10 minutes before each reading was measured using an actigraph worn on the waist.

Research Question

What are the factors that are associated with higher systolic blood pressure?

Load necessary packages

```
library(here)
library(tidyverse)
library(dplyr)
library(car)
library(mice)
```

```
## Warning: package 'mice' was built under R version 4.4.3
```

```
library(lme4)
library(nlme)
library(lmerTest)
```

```
## Warning: package 'lmerTest' was built under R version 4.4.3
```

```
library(ggplot2)
```

Data Inspection and Manipulation

```
# Load data
nursebp <- read.csv(here("Data/nursebp.csv"), header = TRUE)

# Check structure
str(nursebp)
```

```
## 'data.frame':    9573 obs. of  16 variables:
##  $ SNUM    : int  1002 1002 1002 1002 1002 1002 1002 1002 1002 1002 ...
##  $ SYS     : int  136 114 130 120 117 143 118 117 91 102 ...
##  $ DIA     : int  76 63 72 68 57 64 79 79 75 65 ...
##  $ HRT     : int  77 83 80 86 85 84 79 82 75 76 ...
##  $ MNACT5  : num  NA 230 189 229 213 ...
##  $ PHASE   : chr  "L" "L" "L" "L" ...
##  $ DAY     : chr  "W" "W" "W" "W" ...
##  $ POSTURE : chr  "SIT" "STAND" "STAND" "STAND" ...
##  $ STR     : int  1 2 1 1 1 1 2 1 2 1 ...
##  $ HAP     : int  5 4 5 5 5 5 5 5 4 5 ...
##  $ TIR     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ AGE     : int  49 49 49 49 49 49 49 49 49 49 ...
##  $ FH123   : chr  "YES" "YES" "YES" "YES" ...
##  $ time    : int  427 450 468 489 506 527 531 549 565 569 ...
##  $ timept  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ timepass: int  0 23 41 62 79 100 104 122 138 142 ...
```

```
# Summary of key variables
summary(nursebp)
```

```
##      SNUM           SYS             DIA             HRT
##  Min.   :1002   Min.   : 75.0   Min.   : 40.00   Min.   : 35.00
##  1st Qu.:1091   1st Qu.:108.0   1st Qu.: 63.00   1st Qu.: 71.00
##  Median :1154   Median :117.0   Median : 71.00   Median : 80.00
##  Mean   :1156   Mean   :118.2   Mean   : 71.38   Mean   : 80.03
##  3rd Qu.:1231   3rd Qu.:127.0   3rd Qu.: 79.00   3rd Qu.: 88.00
##  Max.   :1307   Max.   :200.0   Max.   :120.00   Max.   :144.00
##
##      MNACT5         PHASE               DAY              POSTURE
##  Min.   :  0.0   Length:9573        Length:9573        Length:9573
##  1st Qu.:160.2   Class :character   Class :character   Class :character
##  Median :207.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :190.4
##  3rd Qu.:236.4
##  Max.   :359.4
##  NA's   :985
##      STR             HAP             TIR             AGE
##  Min.   :1.00    Min.   :1.000   Min.   :1.000   Min.   :24.00
##  1st Qu.:1.00    1st Qu.:2.000   1st Qu.:1.000   1st Qu.:33.00
##  Median :1.00    Median :3.000   Median :2.000   Median :38.00
##  Mean   :1.51    Mean   :3.099   Mean   :1.954   Mean   :37.82
##  3rd Qu.:2.00    3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:43.00
##  Max.   :5.00    Max.   :5.000   Max.   :5.000   Max.   :50.00
##  NA's   :754     NA's   :755     NA's   :755
##      FH123                time            timept           timepass
```

```
##   Length:9573         Min.   : 300   Min.   : 1.00   Min.   :    0
##   Class :character    1st Qu.: 665   1st Qu.:12.00   1st Qu.: 217
##   Mode  :character    Median : 891   Median :24.00   Median : 443
##                       Mean   : 890   Mean   :24.41   Mean   : 447
##                       3rd Qu.:1118   3rd Qu.:36.00   3rd Qu.: 671
##                       Max.   :1439   Max.   :60.00   Max.   :1450
##
```

```
# Count missingness
colSums(is.na(nursebp))
```

```
##     SNUM      SYS      DIA      HRT   MNACT5    PHASE      DAY  POSTURE
##        0        0        0        0      985        0        0        0
##      STR      HAP      TIR      AGE    FH123     time   timept timepass
##      754      755      755        0        0        0        0        0
```

```
# Number of unique subjects
length(unique(nursebp$SNUM))
```

```
## [1] 203
```

```
nursebp <- nursebp %>%
  mutate(
    DAY = as.factor(DAY),
    FH123 = as.factor(FH123),
    PHASE = as.factor(PHASE),
    POSTURE = as.factor(POSTURE),
    SNUM = as.factor(SNUM),
    STR = as.factor(STR),
    HAP = as.factor(HAP)
  )
```

Exploratory Analysis

```
ggplot(nursebp, aes(x = SYS)) +
  geom_histogram(binwidth = 5, fill = "blue", alpha = 0.6) +
  labs(title = "Distribution of Systolic Blood Pressure",
       x = "SYS (mmHg)", y = "Count")
```

Distribution of Systolic Blood Pressure
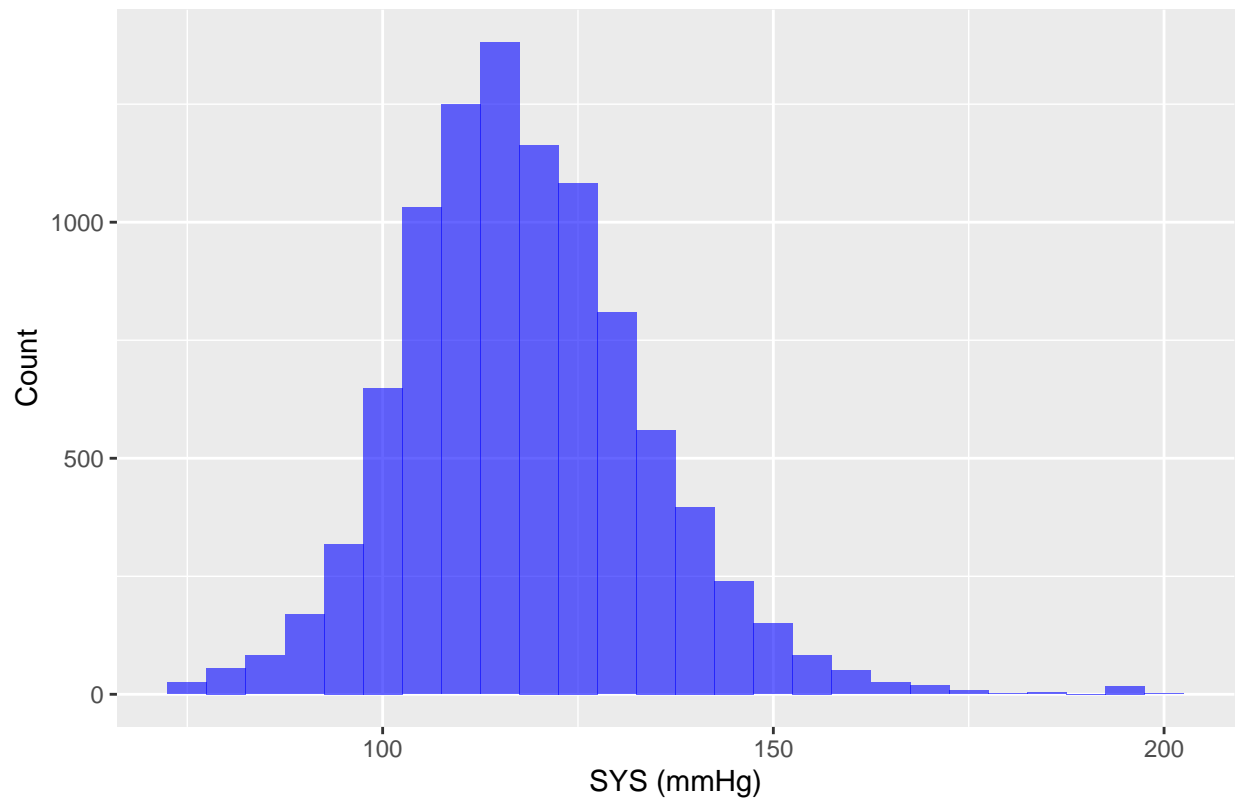

```r
ggplot(nursebp, aes(x = time, y = SYS, group = SNUM, color = SNUM)) +
  geom_line(alpha = 0.4) +
  labs(title = "Individual SBP Trajectory Over Time",
       x = "Time (minutes from midnight)", y = "SYS (mmHg)") +
  theme(legend.position = "none")
```

## Individual SBP Trajectory Over Time
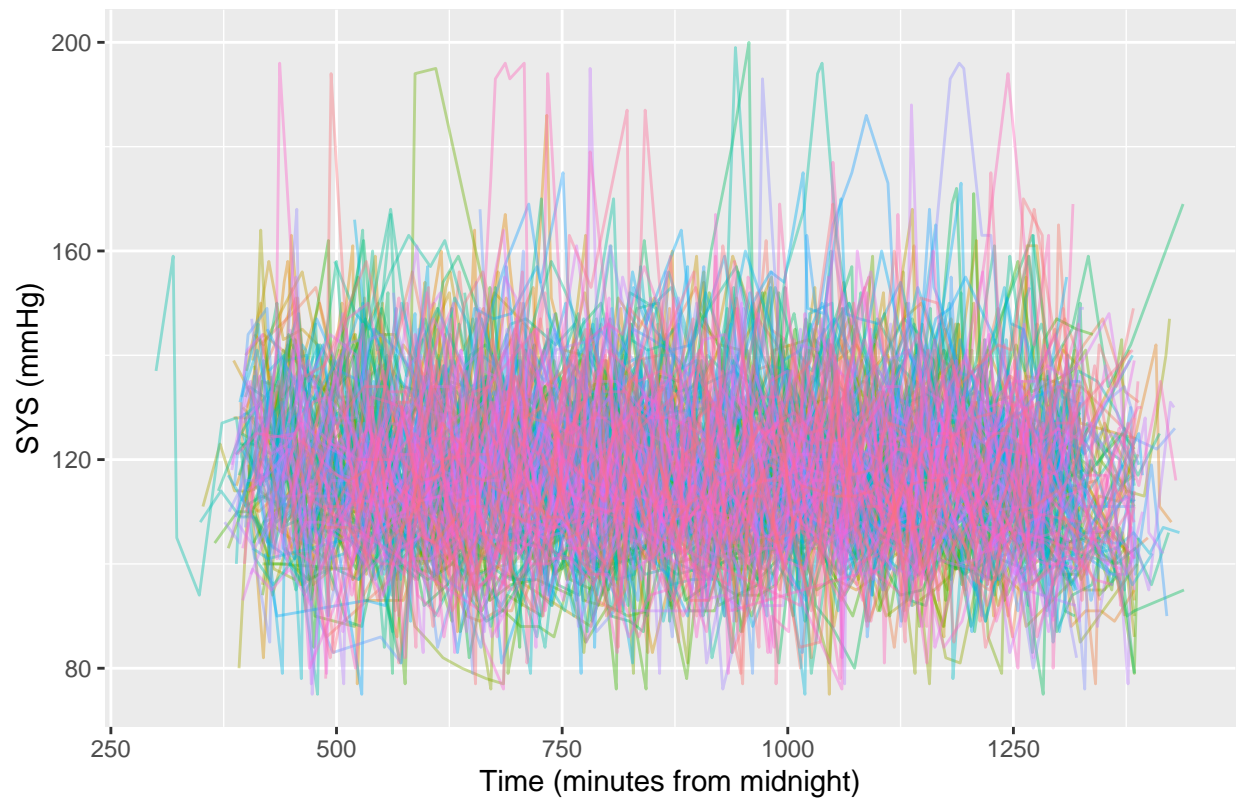

```r
ggplot(nursebp, aes(x = FH123, y = SYS, fill = FH123)) +
  geom_boxplot() +
  labs(title = "SBP by Family History of Hypertension",
       x = "Family History", y = "SYS (mmHg)")
```

SBP by Family History of Hypertension
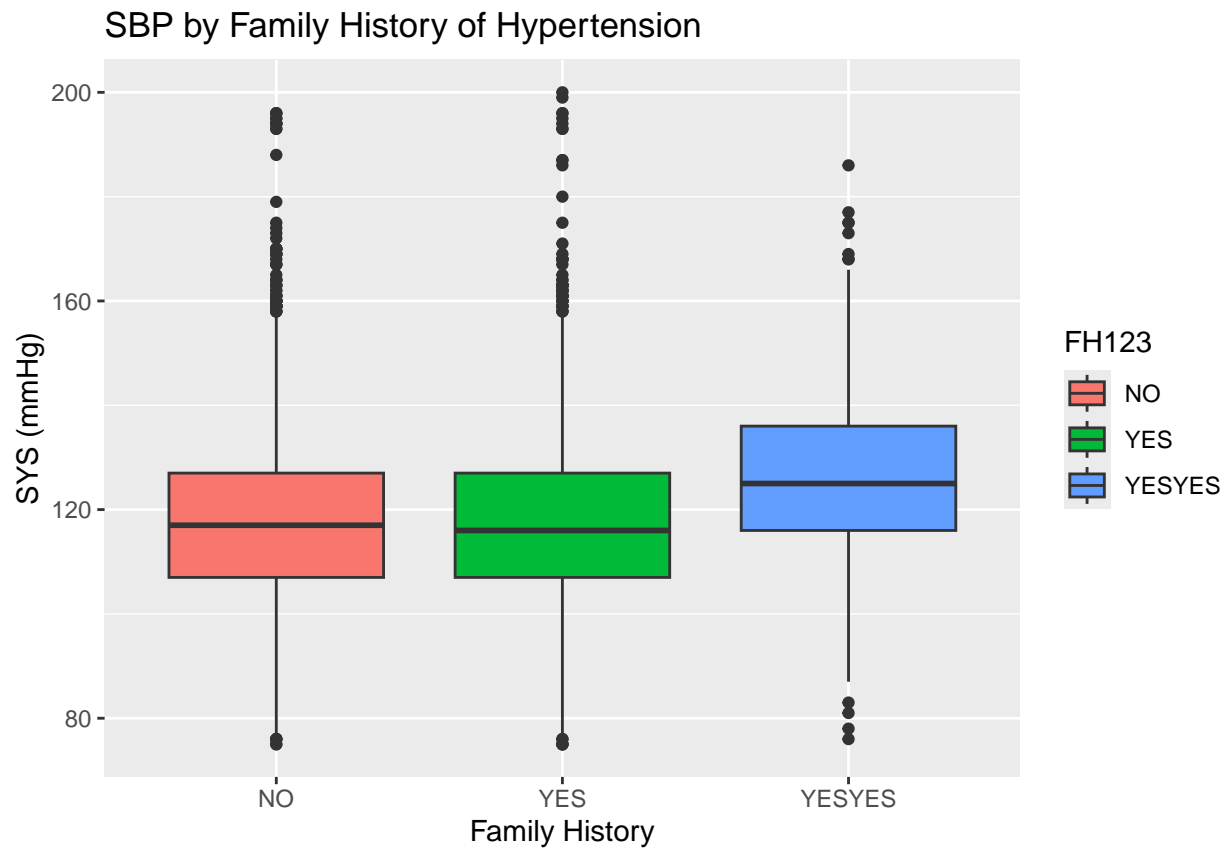
```
ggplot(nursebp, aes(x = PHASE, y = SYS, fill = PHASE)) +
  geom_boxplot() +
  labs(title = "SBP by Menstrual Phase",
       x = "Menstrual Phase", y = "SYS (mmHg)")
```

SBP by Menstrual Phase


```
ggplot(nursebp, aes(x = POSTURE, y = SYS, fill = POSTURE)) +
  geom_boxplot() +
  labs(title = "SBP by Posture",
       x = "Position during BP measurement", y = "SYS (mmHg)")
```

SBP by Posture
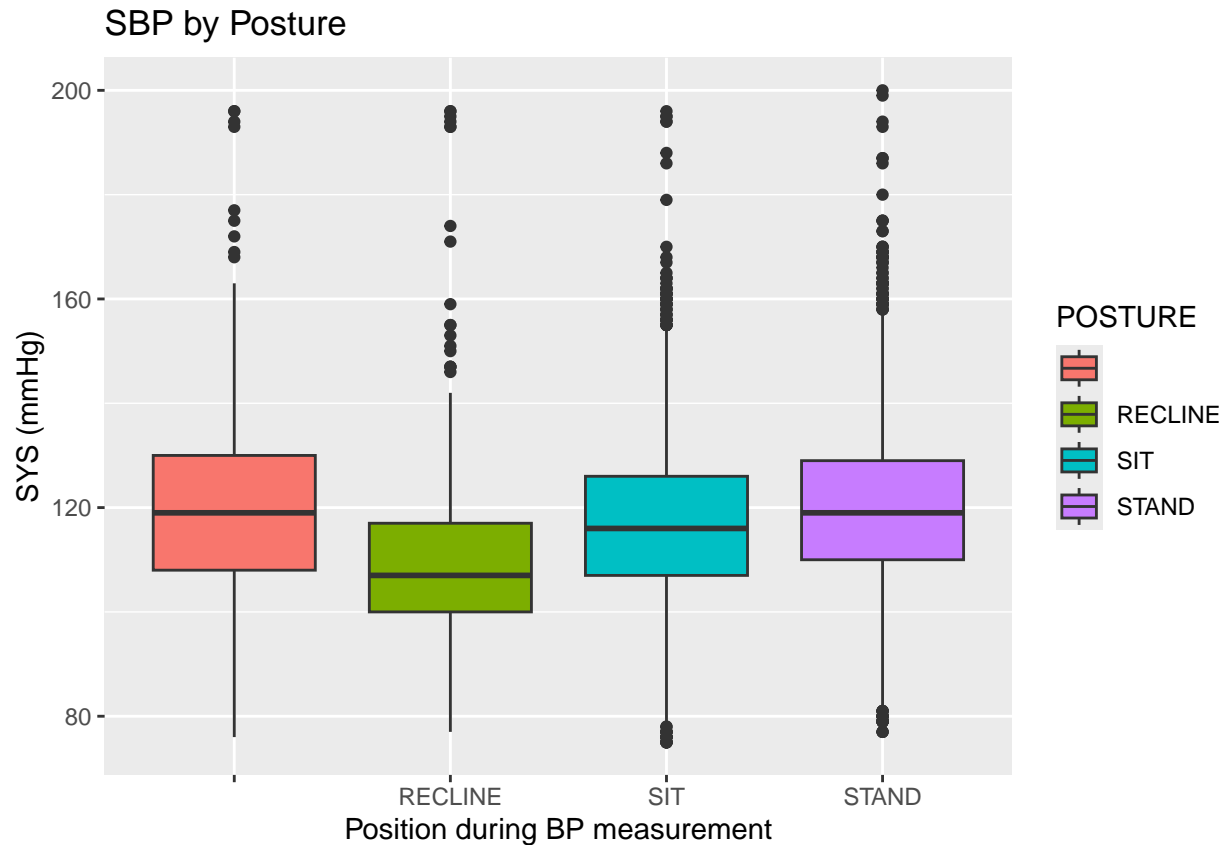
Correlation among variables

```
# DIA should be correlated with SYS, and it is with moderate correlation. The other variables seem no o
nursebp %>%
  select(AGE, SYS, DIA, HRT, time) %>%
  cor(use = "complete.obs")
```

```
##              AGE         SYS         DIA         HRT        time
## AGE   1.00000000  0.03479532  0.11380282 -0.04977327 -0.01273649
## SYS   0.03479532  1.00000000  0.53557471  0.18468989 -0.04091748
## DIA   0.11380282  0.53557471  1.00000000  0.20380942 -0.08098117
## HRT  -0.04977327  0.18468989  0.20380942  1.00000000 -0.03003795
## time -0.01273649 -0.04091748 -0.08098117 -0.03003795  1.00000000
```

Modelling

Missingness examination

```
nursebp$STR_missing <- ifelse(is.na(nursebp$STR), 1, 0)
table(nursebp$AGE, nursebp$STR_missing)
```

```
##
##        0   1
##    24 42   1
```

```
##    25 152    1
##    26 355    8
##    27 268    7
##    28 219   14
##    29 391   52
##    30 157   19
##    31 183   60
##    32 346   25
##    33 260   26
##    34 262   12
##    35 560   34
##    36 355   42
##    37 457   18
##    38 480   41
##    39 750   97
##    40 386   17
##    41 411   12
##    42 269   25
##    43 423   48
##    44 320   20
##    45 457   41
##    46 402   28
##    47 557   14
##    48 148    6
##    49 166   24
##    50  43   62
```

```r
# Density plot of age by missingness indicator
ggplot(nursebp, aes(x = AGE, fill = factor(STR_missing))) +
  geom_density(alpha = 0.5) +
  labs(x = "Age",
       y = "Density",
       fill = "STR Missing (1 = Yes, 0 = No)",
       title = "Age Distribution by STR Missingness")
```

Age Distribution by STR Missingness

still need to decide whether impute the data

```
fixed_model <- lm(SYS ~ AGE + DIA + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE + POSTURE + tim
vif(fixed_model)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## AGE      1.092867  1        1.045403
## DIA      1.121460  1        1.058990
## FH123    1.108143  2        1.026004
## HAP      1.247107  4        1.027988
## HRT      1.216603  1        1.102997
## MNACT5   1.317044  1        1.147625
## STR      1.292430  4        1.032585
## TIR      1.264703  1        1.124590
## DAY      1.095818  1        1.046813
## PHASE    1.039525  1        1.019571
## POSTURE  1.465908  3        1.065821
## time    54.908396  1        7.410020
## I(time^2) 56.284715 1        7.502314
```

high multicollinearity among time and time^2, could consider centering

1. Random intercept,iInclude all variables

```r
# Center and scale time
nursebp$time_c <- scale(nursebp$time, center = TRUE, scale = TRUE)
nursebp$time_c2 <-(nursebp$time_c)^2
# Linear mixed model: random intercept model
model0 <- lmer(SYS ~ AGE + DIA + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE + POSTURE + time_c
summary(model0)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: SYS ~ AGE + DIA + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY +
##      PHASE + POSTURE + time_c + time_c2 + (1 | SNUM)
##    Data: nursebp
##
## REML criterion at convergence: 61252.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7682 -0.5618 -0.0535  0.4820  7.9503
##
## Random effects:
##  Groups    Name         Variance Std.Dev.
##  SNUM     (Intercept)  38.16    6.178
##  Residual              129.08   11.361
## Number of obs: 7896, groups:  SNUM, 182
##
## Fixed effects:
##                   Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)      7.316e+01  4.219e+00 7.305e+02  17.340   <2e-16 ***
## AGE             -4.093e-02  7.416e-02 1.723e+02  -0.552   0.5817
## DIA              4.998e-01  1.193e-02 7.866e+03  41.904   <2e-16 ***
## FH123YES        -2.394e-01  1.021e+00 1.715e+02  -0.234   0.8150
## FH123YESYES      4.435e+00  1.930e+00 1.721e+02   2.298   0.0228 *
## HAP2             1.089e-01  6.242e-01 7.671e+03   0.174   0.8615
## HAP3            -4.653e-02  6.305e-01 7.058e+03  -0.074   0.9412
## HAP4            -2.722e-01  6.970e-01 6.560e+03  -0.391   0.6961
## HAP5            -9.751e-01  8.524e-01 5.474e+03  -1.144   0.2527
## HRT              4.084e-02  1.426e-02 7.236e+03   2.864   0.0042 **
## MNACT5           3.192e-02  2.433e-03 7.860e+03  13.123   <2e-16 ***
## STR2             2.188e-01  3.551e-01 7.868e+03   0.616   0.5378
## STR3            -5.028e-01  5.722e-01 7.869e+03  -0.879   0.3796
## STR4             1.173e+00  9.629e-01 7.849e+03   1.218   0.2232
## STR5             7.068e-01  1.671e+00 7.821e+03   0.423   0.6724
## TIR             -1.097e-01  1.715e-01 7.594e+03  -0.640   0.5224
## DAYW             1.183e+00  9.656e-01 1.738e+02   1.225   0.2222
## PHASEL           3.437e-01  9.647e-01 1.713e+02   0.356   0.7220
## POSTURERECLINE   3.339e-01  2.737e+00 7.758e+03   0.122   0.9029
## POSTURESIT       8.421e-01  2.679e+00 7.754e+03   0.314   0.7532
## POSTURESTAND     7.930e-01  2.678e+00 7.755e+03   0.296   0.7672
## time_c           1.816e-01  1.468e-01 7.828e+03   1.237   0.2160
## time_c2          1.200e-01  1.477e-01 7.774e+03   0.812   0.4167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
```

```
## Correlation matrix not shown by default, as p = 23 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)         if you need it
```

```
##################DO NOT RUN THIS YET, FOR MULTICOLINEARITY ONLY##################
# Correlation matrix
#cor_matrix <- cor(nursebp[, c("HAP", "STR", "TIR")], use = "complete.obs")
#print(cor_matrix)
# VIF calculation
#library(car)
#vif(model0)
```

2. Random Intercept and slope model

```
# Linear mixd model: random intercept and slope
model1 <- lmer(SYS ~ AGE + DIA + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE + POSTURE + time_c
summary(model1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: SYS ~ AGE + DIA + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY +
##     PHASE + POSTURE + time_c + time_c2 + (1 + time_c | SNUM)
##    Data: nursebp
##
## REML criterion at convergence: 61193.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8960 -0.5496 -0.0611  0.4761  7.4605
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SNUM     (Intercept)  38.783   6.228
##           time_c        3.197   1.788   0.25
##  Residual             126.042  11.227
## Number of obs: 7896, groups:  SNUM, 182
##
## Fixed effects:
##                 Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)    7.286e+01  4.207e+00  7.421e+02  17.319  < 2e-16 ***
## AGE           -3.058e-02  7.368e-02  1.736e+02  -0.415  0.67858
## DIA            4.920e-01  1.195e-02  7.853e+03  41.193  < 2e-16 ***
## FH123YES      -7.288e-01  1.015e+00  1.729e+02  -0.718  0.47385
## FH123YESYES    4.548e+00  1.921e+00  1.747e+02   2.367  0.01902 *
## HAP2           2.627e-01  6.284e-01  7.536e+03   0.418  0.67593
## HAP3           9.260e-02  6.420e-01  6.410e+03   0.144  0.88532
## HAP4          -7.669e-02  7.097e-01  5.931e+03  -0.108  0.91395
## HAP5          -4.921e-01  8.742e-01  4.744e+03  -0.563  0.57353
## HRT            3.979e-02  1.445e-02  6.825e+03   2.753  0.00593 **
## MNACT5         3.157e-02  2.442e-03  7.839e+03  12.927  < 2e-16 ***
## STR2           2.174e-01  3.585e-01  7.683e+03   0.606  0.54423
## STR3          -5.426e-01  5.787e-01  7.514e+03  -0.938  0.34849
## STR4           1.224e+00  9.699e-01  7.751e+03   1.262  0.20684
```

```
## STR5            5.907e-01  1.681e+00  7.765e+03   0.351  0.72533
## TIR            -5.932e-02  1.852e-01  3.478e+03  -0.320  0.74880
## DAYW            1.166e+00  9.597e-01  1.753e+02   1.215  0.22590
## PHASEL          6.657e-01  9.591e-01  1.729e+02   0.694  0.48859
## POSTURERECLINE  5.743e-01  2.731e+00  7.755e+03   0.210  0.83348
## POSTURESIT      1.158e+00  2.670e+00  7.743e+03   0.434  0.66456
## POSTURESTAND    1.186e+00  2.669e+00  7.744e+03   0.444  0.65671
## time_c          1.920e-01  2.001e-01  2.079e+02   0.960  0.33837
## time_c2         1.369e-01  1.519e-01  6.566e+03   0.902  0.36734
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##
## Correlation matrix not shown by default, as p = 23 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

3. Remove DIA

```
model3 <- lmer(SYS ~ AGE + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE + POSTURE + time_c + tir

summary(model3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: SYS ~ AGE + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE +
##     POSTURE + time_c + time_c2 + (1 + time_c | SNUM)
##    Data: nursebp
##
## REML criterion at convergence: 62712.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.3190 -0.5430 -0.0366  0.5131  6.5224
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  SNUM     (Intercept)  61.035   7.812
##           time_c        5.799   2.408   0.21
##  Residual             151.320  12.301
## Number of obs: 7896, groups:  SNUM, 182
##
## Fixed effects:
##                 Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)    9.948e+01  4.873e+00  5.776e+02  20.415  < 2e-16 ***
## AGE            6.988e-02  9.170e-02  1.750e+02   0.762  0.44707
## FH123YES      -7.872e-01  1.265e+00  1.748e+02  -0.622  0.53445
## FH123YESYES    7.108e+00  2.391e+00  1.758e+02   2.973  0.00336 **
## HAP2           1.582e-01  6.951e-01  7.744e+03   0.228  0.82002
## HAP3           9.745e-02  7.145e-01  7.027e+03   0.136  0.89152
## HAP4           2.447e-01  7.912e-01  6.659e+03   0.309  0.75715
## HAP5          -3.271e-01  9.795e-01  5.675e+03  -0.334  0.73845
## HRT            9.898e-02  1.597e-02  7.309e+03   6.199     6e-10 ***
```

```
## MNACT5         3.279e-02  2.688e-03  7.839e+03  12.195  < 2e-16 ***
## STR2           5.121e-01  3.958e-01  7.802e+03   1.294  0.19570
## STR3          -2.934e-01  6.395e-01  7.695e+03  -0.459  0.64642
## STR4           2.044e+00  1.069e+00  7.803e+03   1.912  0.05588 .
## STR5           3.004e+00  1.850e+00  7.789e+03   1.623  0.10454
## TIR           -3.188e-01  2.081e-01  4.362e+03  -1.532  0.12559
## DAYW           2.703e+00  1.194e+00  1.762e+02   2.264  0.02479 *
## PHASEL         4.117e-01  1.195e+00  1.748e+02   0.345  0.73078
## POSTURERECLINE -3.629e+00 2.999e+00  7.733e+03  -1.210  0.22624
## POSTURESIT     1.332e-01  2.932e+00  7.718e+03   0.045  0.96376
## POSTURESTAND   5.115e-01  2.931e+00  7.719e+03   0.174  0.86149
## time_c         6.301e-02  2.436e-01  2.093e+02   0.259  0.79615
## time_c2        3.536e-01  1.680e-01  6.992e+03   2.105  0.03536 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
anova(model0, model3)
```

```
## refitting model(s) with ML (instead of REML)


## Data: nursebp
## Models:
## model0: SYS ~ AGE + DIA + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE + POSTURE + time_c +
## model3: SYS ~ AGE + FH123 + HAP + HRT + MNACT5 + STR + TIR + DAY + PHASE + POSTURE + time_c + time_c2
##        npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## model0   25 61288 61463 -30619    61238
## model3   26 62765 62946 -31356    62713     0  1          1
```

```
# Based on likelihood ratio test, we should not remove DIA
# Just a test for model comparison
```