

Longitudinal Data Analysis

Case study of Trenal.XLS using Linear Mixed Effect Model

Group2:Hugo Blain; Oscar Cabanelas;Wanchang Zhang

2023-03-29

Contents

Data description	3
Backgrounds of the data	3
Data preprocess	3
Import and clean up data Trenal.XLS	3
Data Organization	3
Missing Data	4
Univariate summaries	6
Plot histogram of continuous variables	6
Bivariate summaries	7
Plot relationship between pairs of variables	7
Plot the time trend of response	8
Find out a covariate increasing or decreasing the responses time trend	11
Data set analysis to see the age effect	18
Conclusions after exploring data analysis	20
Multilevel Data Analysis	20
Multivariate Linear Model Analysis	20
Two Stage Model Analysis	23
Linear Mixed effects Model Analysis	23
Unconditional Means Model to discover variance distribution	23
Unconditional Growth Model, introducing time in Level One	24
Modeling other trends over time quadratic	26
Piecewise linear time trend	28
Adding subject related variable in level two	30

List of Figures

List of Tables

1	Missing data for each measurement	4
---	---	---

Data description

Backgrounds of the data

The dataset `Trenal.XLS` contains information on patients who received renal graft(kidney transplant). The patients have been followed for at most 10 years.

People with end-stage kidney disease who receive a kidney transplant generally live longer than people with ESRD who are on dialysis. However, kidney transplant recipients must remain on immunosuppressants (medications to suppress the immune system) for the rest of their life to prevent their body from rejecting the new kidney. The long-term immunosuppression puts them at risk for infections and cancer. Haematocrit level is measured for each patient who has received renal graft to see if gender, the age to go through the operation, reject or not, cardio history or not will influence the healthy state of a patient after operation.

Data preprocess

Import and clean up data `Trenal.XLS`

```
trenal <- read_excel("Trenal.XLS") # summary(trenal)

trenal = trenal[,-18] #remove a noninformative column const

# Continuous or discrete variables
trenal$id = as.factor(trenal$id)
trenal$j = as.factor(trenal$j)
#trenal$time = as.factor(trenal$time)
trenal$male = as.factor(trenal$male)
trenal$cardio = as.factor(trenal$cardio)
trenal$reject = as.factor(trenal$reject)

# Change the name of respons
colnames(trenal)[19] <- "Hc"
trenal.long = trenal[,13:20] # long table form

# Remove j
trenal.long = trenal.long[,-6]
trenal.long.unique <- trenal.long[match( unique(trenal.long$id), trenal.long$id),] # meanHc should repla
trenal.long.noNA <- na.omit(trenal.long)

# Wide table form
trenal.wide = as.data.frame(subset(trenal,trenal$j=="1"))[,1:18] # 1160 x 18
```

Data Organization

- The input data:
 - id: total 1160 persons
 - age to perform the operation: from 15 to 76 years old, average is 46.43 years old
 - male: we observe 494 females and 666 males
 - cardio: 953 persons has experienced a cardio-vascular problem during the years preceding the transplant, 207 did not.
 - reject: 793 patients shown symptoms of graft rejection during the first three months after the transportation, 367 has not.
- The response variable Hc level: continous from min 14% to max 65%.The Hc level is dependent on the measured time, individual's age to perform the operation, gender, cardio history and reject history.

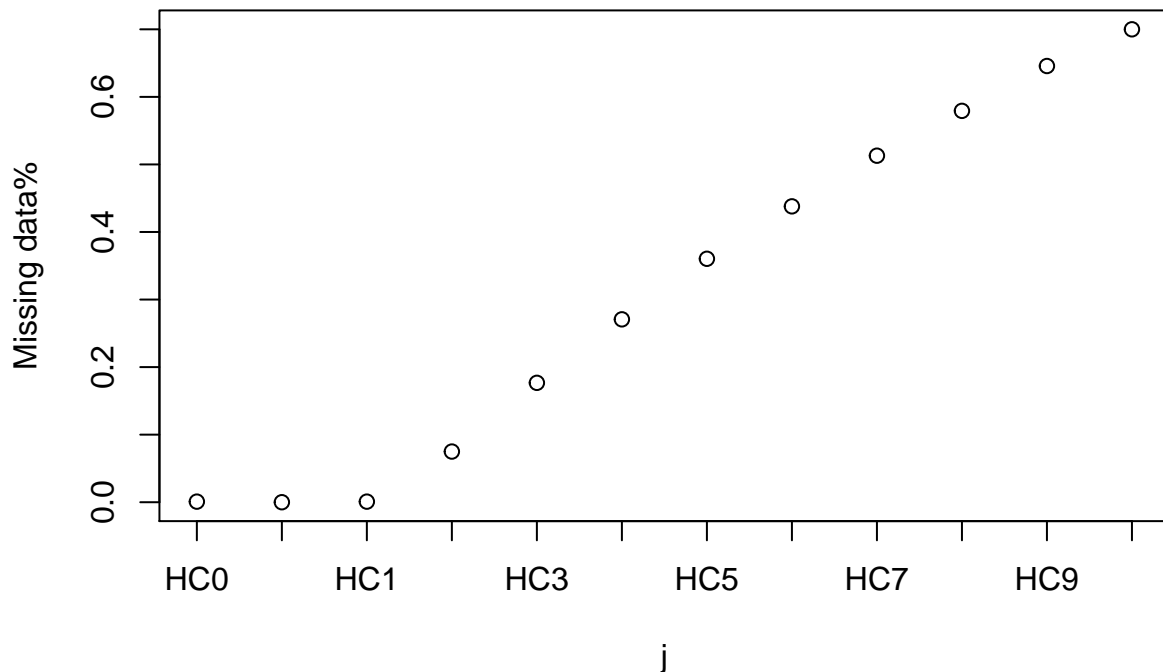
Missing Data

```
# Analyse the NA values descriptively
## First to collect how many NAs are in Hc0, Hc0.5, Hc1, ..., Hc 10
Hc.NA = numeric(12)
for (i in c(1:12)) {
  Hc.NA[i] = sum(is.na(trenal.wide[,i]))
}
# 1  0  1  87 205 314 418 508 595 672 749 812
# The number of missing data / the ideal case have all measurements for everyone
Hc.NA.percentage = Hc.NA/dim(trenal.long.unique)[1]
missing <- data.frame(rbind(Hc.NA,Hc.NA.percentage))
kable(missing, caption= "Missing data for each measurement",
      col.names =colnames(trenal.wide[c(1:12)]),digits = 3)
```

Table 1: Missing data for each measurement

	HC0	HC06	HC1	HC2	HC3	HC4	HC5	HC6	HC7	HC8	HC9	HC10
Hc.NA	1.000	0	1.000	87.000	205.000	314.000	418.00	508.000	595.000	672.000	749.000	812.0
Hc.NA.percentage	0.001	0	0.001	0.075	0.177	0.271	0.36	0.438	0.513	0.579	0.646	0.7

```
plot(Hc.NA.percentage,xaxt="n",xlab="j",ylab="Missing data%")
axis(side=1,at=c(1,2,3,4,5,6,7,8,9,10,11,12),labels=colnames(trenal.wide)[1:12])
```



Conclusion could be at the first three measurements, there are almost full data More people tends to miss

the measurements when time increases. And then we can extract all NA data from the long table to analyse their construction

```
trenal.long.NA = trenal.long[is.na(trenal.long$Hc),]
#t = unique(trenal.long.NA$id) # 821 individuals
trenal.long.NA.unique <- trenal.long.NA[match( unique(trenal.long.NA$id), trenal.long.NA$id),]
summary(trenal.long.NA.unique)
```

```
##          id          age      male      cardio      reject      Hc
##  1      :  1   Min.    :15.00   0:327    0:669    0:588   Min.    : NA
##  2      :  1   1st Qu.:39.00   1:494    1:152    1:233   1st Qu.: NA
##  4      :  1   Median :50.00                      Median : NA
##  7      :  1   Mean    :48.31                      Mean    :NaN
## 14      :  1   3rd Qu.:59.00                      3rd Qu.: NA
## 18      :  1   Max.    :76.00                      Max.    : NA
## (Other):815   NA's    :1                          NA's    :821
```

```
##          time
##  Min.      : 0.000
##  1st Qu.   : 3.000
##  Median    : 5.000
##  Mean      : 5.638
##  3rd Qu.   : 8.000
##  Max.      :10.000
##
```

Conclusion, For the missing data, we can see that

```
png(file="MissingValueAnalysis.png",
    width=600, height=1200)
plot.new()
par(mfrow=c(4,2))
## age
hist(trenal.long.unique$age,title="Age distribution in original data")
hist(trenal.long.NA.unique$age,col="red",title="Age distribution in missing data")
```

```
## male
plot(trenal.long.unique$male)
title(main="Gender distribution in original data")
plot(trenal.long.NA.unique$male,col="red")
title(main="Gender distribution in missing data")
```

```
## cardio
plot(trenal.long.unique$cardio)
title(main="Cardio distribution in original data")
plot(trenal.long.NA.unique$cardio,col="red")
title(main="Cardio distribution in missing data")
```

```
## reject
plot(trenal.long.unique$reject)
title(main="Reject distribution in original data")
plot(trenal.long.NA.unique$reject,col="red")
title(main="Reject distribution in missing data")
dev.off()
```

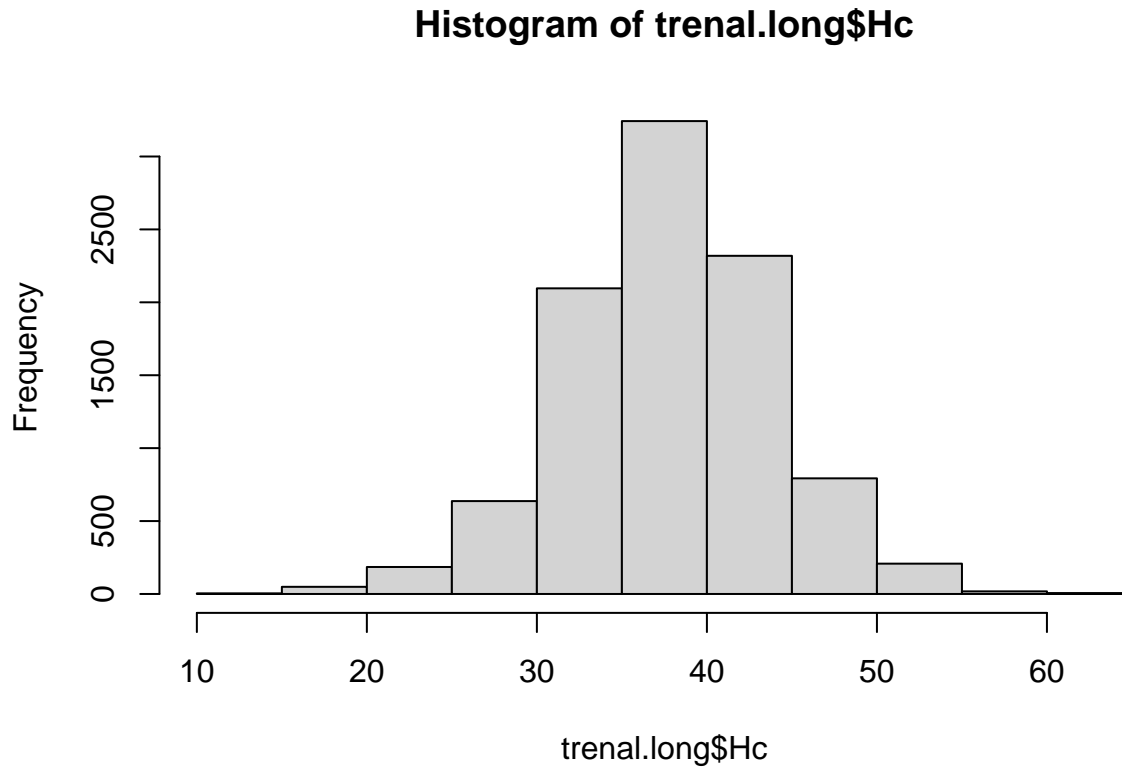
```
## pdf
## 2
```

The missing data has a similar distribution as the ideal full data set, in age, male, cardio and reject plot. So we may conclude that the missing data are random and not depend on any observed predictors or the response. # Exploratory Data Analysis

Univariate summaries

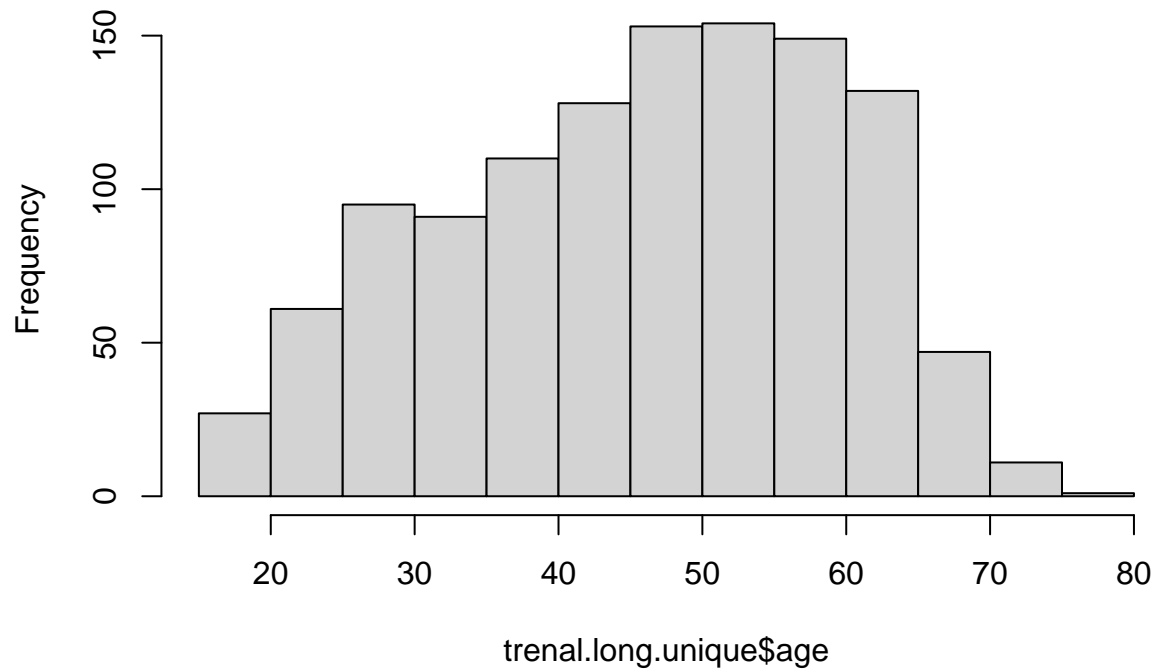
Plot histogram of continuous variables

```
hist(trenal.long$Hc,title="Hc distribution in original data")
```



```
hist(trenal.long.unique$age,title="age distribution in original data")
```

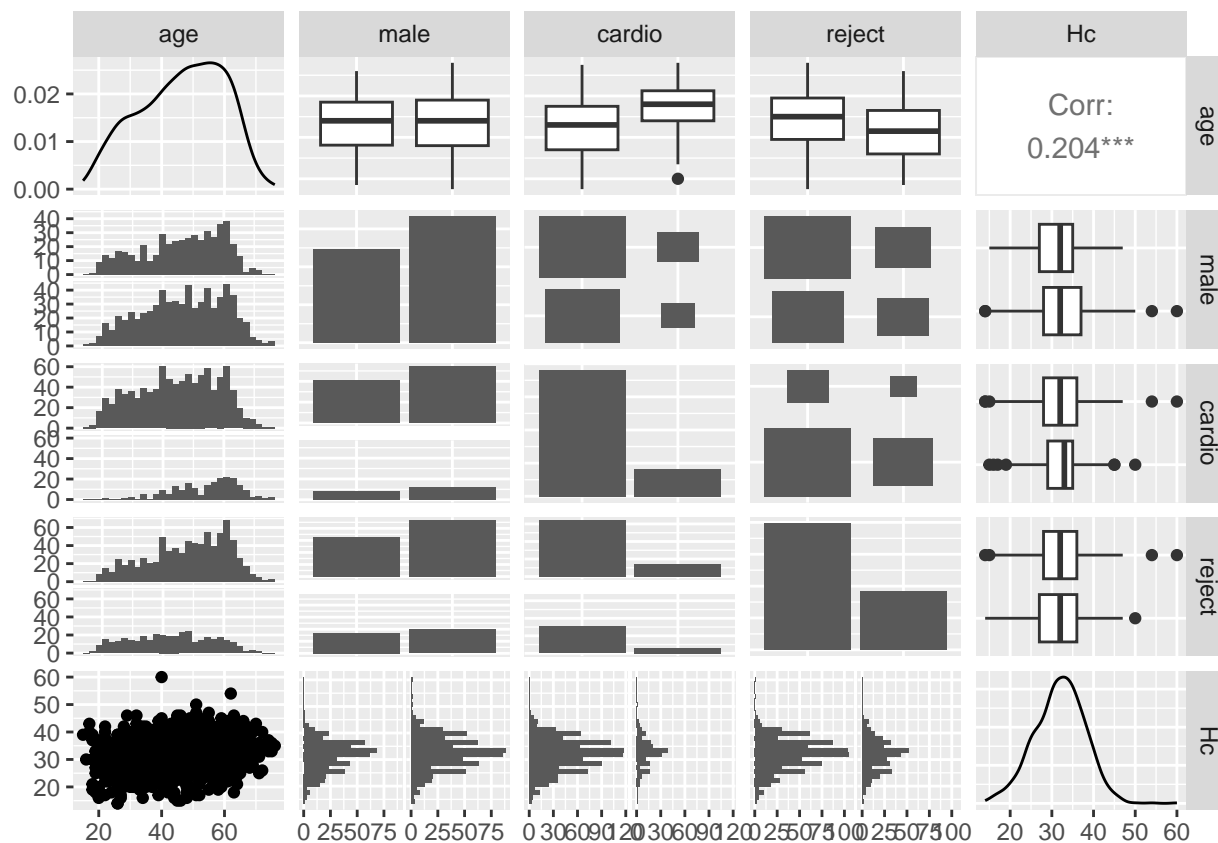
Histogram of trenal.long.unique\$age



Bivariate summaries

Plot relationship between pairs of variables

```
# Bivariate summaries  
gg <- ggpairs(data=trenal.long.unique[,2:6]) # Here Hc is only one value per individual  
gg
```

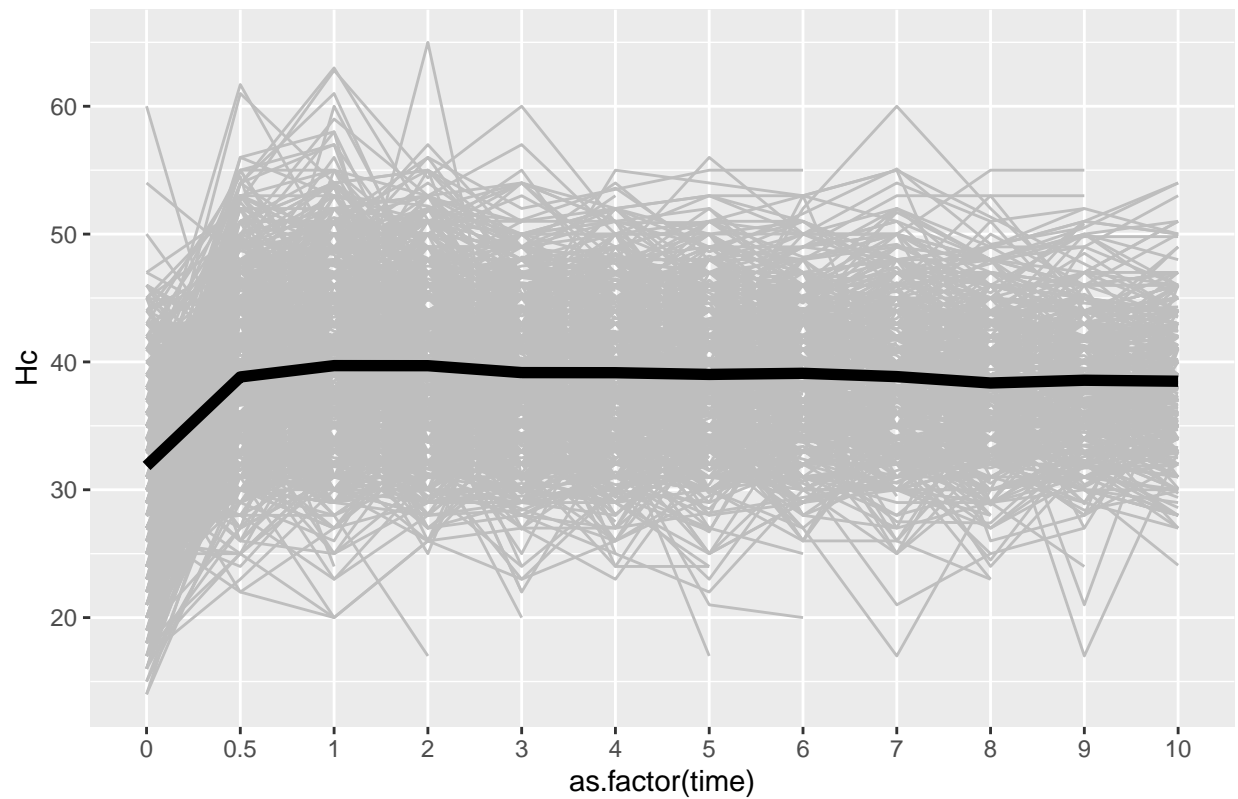


Plot the time trend of response

```
# To view the mean structure of the Hc for all individuals
ggplot(trenal.long.noNA,aes(x=as.factor(time),y=Hc,group=id)) + geom_line(col="grey")+stat_summary(aes(
  labs(title="Line plot of Hc level for all individuals overtime and the mean structure")
```

Mean Structure

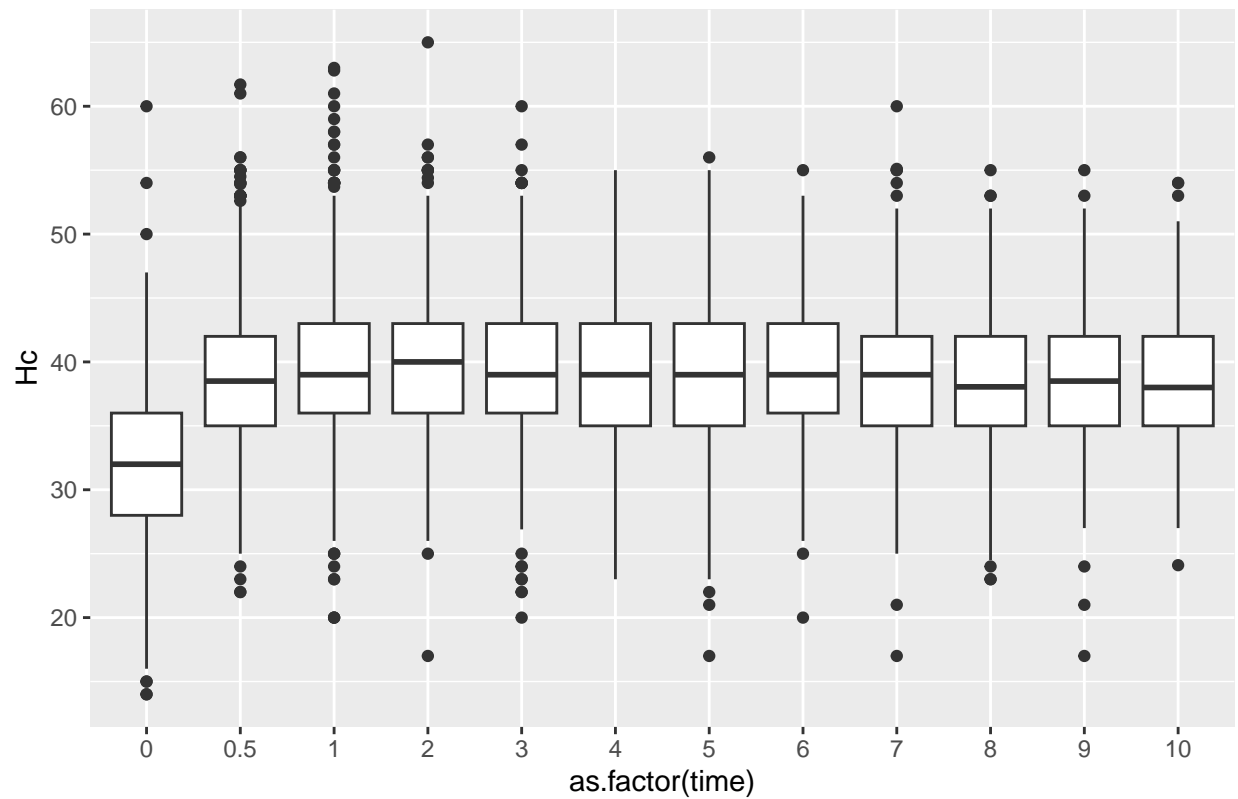
Line plot of Hc level for all individuals overtime and the mean structure



```
# To view to variance structure  
ggplot(trenal.long.noNA,aes(x=as.factor(time),y=Hc))+  
  geom_boxplot(position=position_dodge(1))+  
  labs(title="Box Plot of Hc level for all individuaals over time and the variance structure")
```

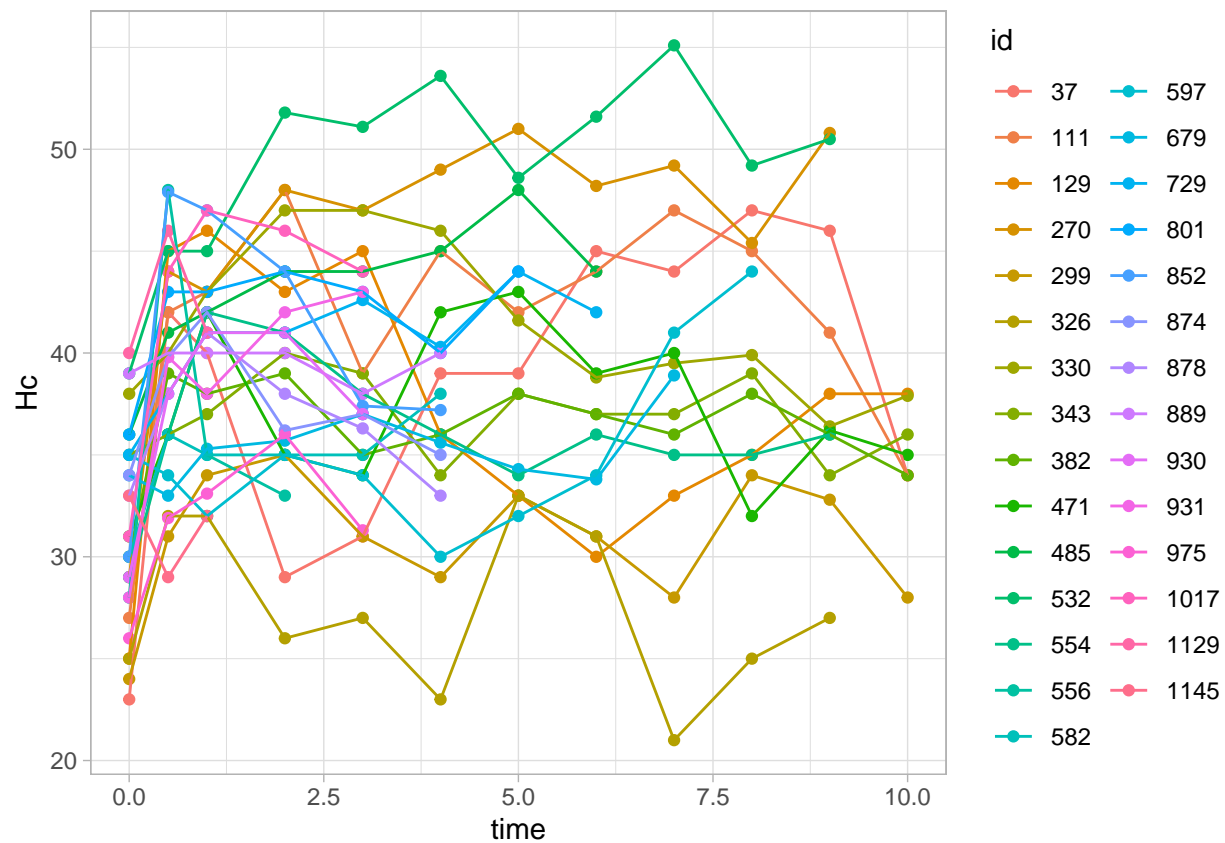
Variance Structure

Box Plot of Hc level for all individuals over time and the variance structure

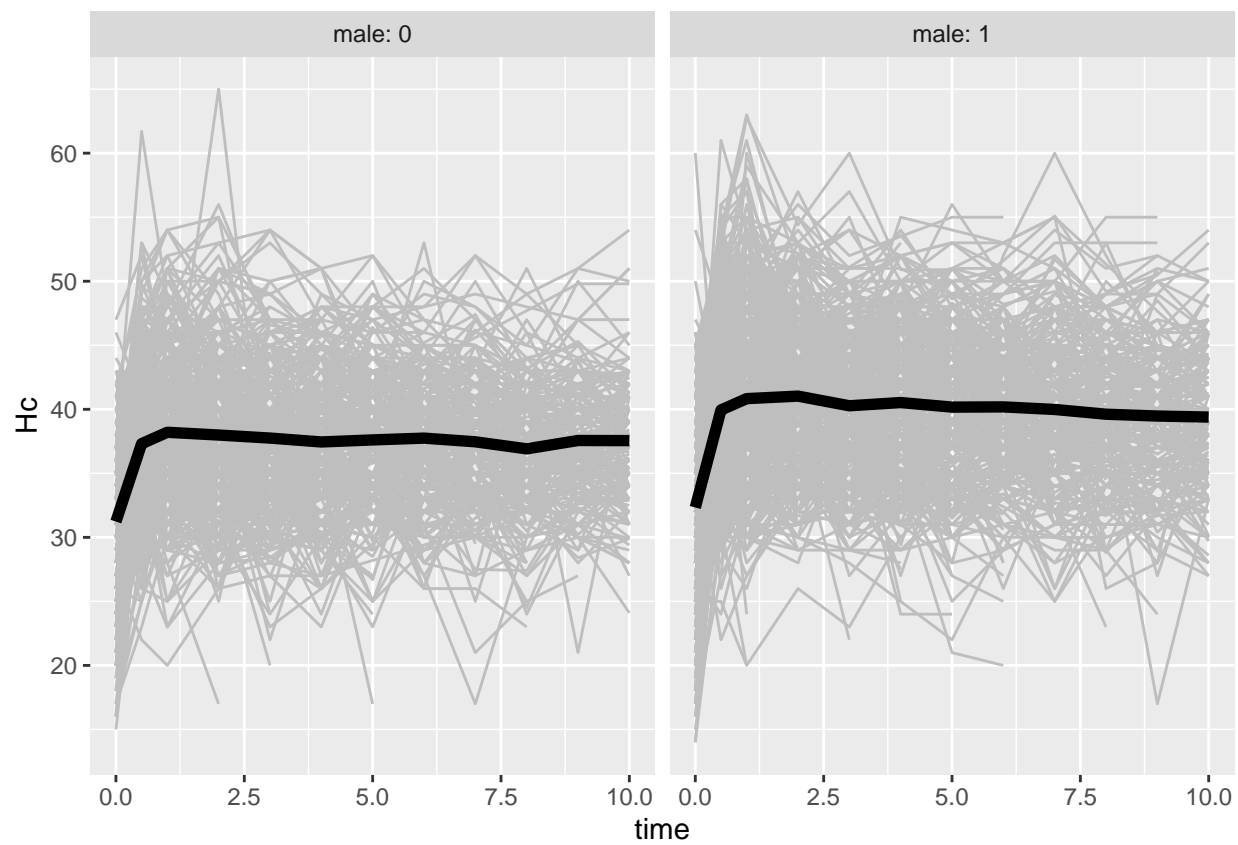


```
HcCorr = trenal.wide[,c(1:12)]
#cor(HcCorr,use="complete.obs" ) # also COV for covariance
chart.Correlation(HcCorr,histogram=TRUE)
```

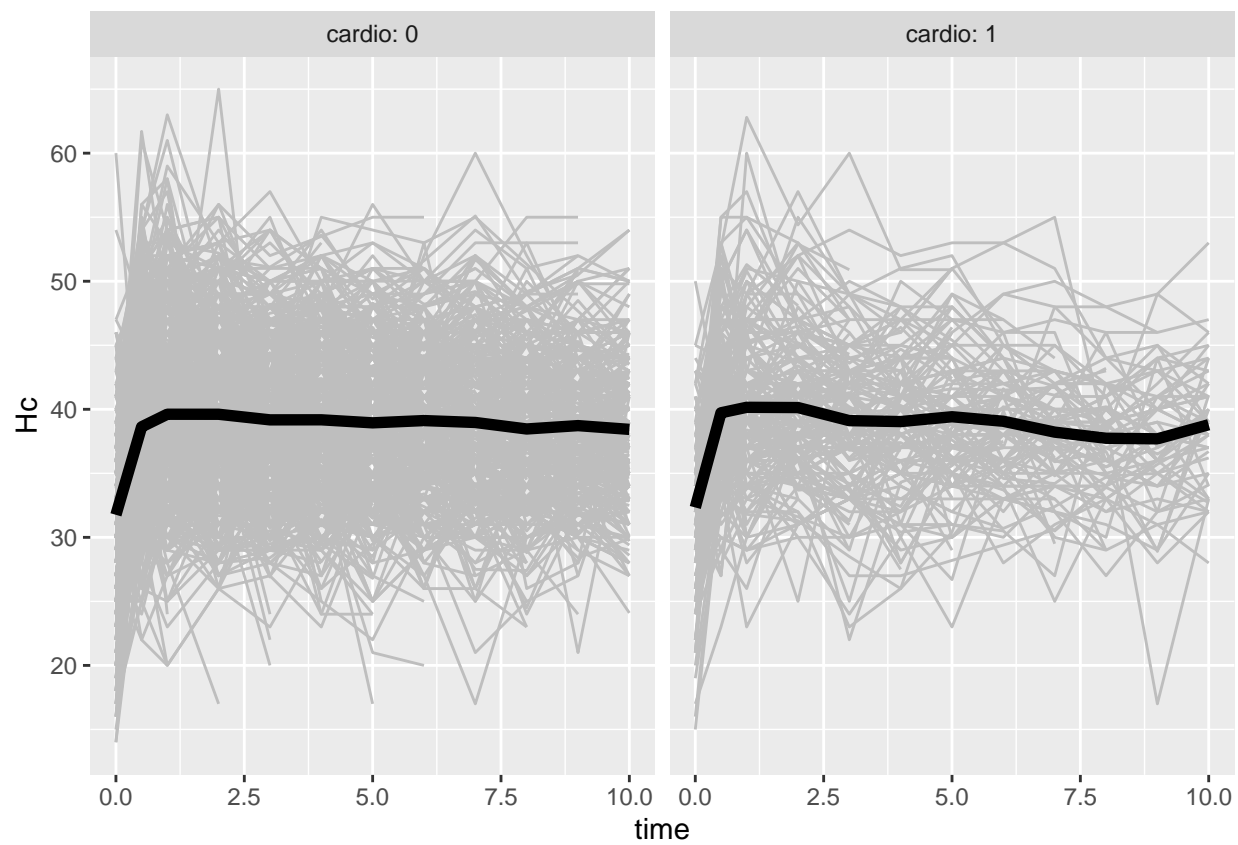
Covariance Structure



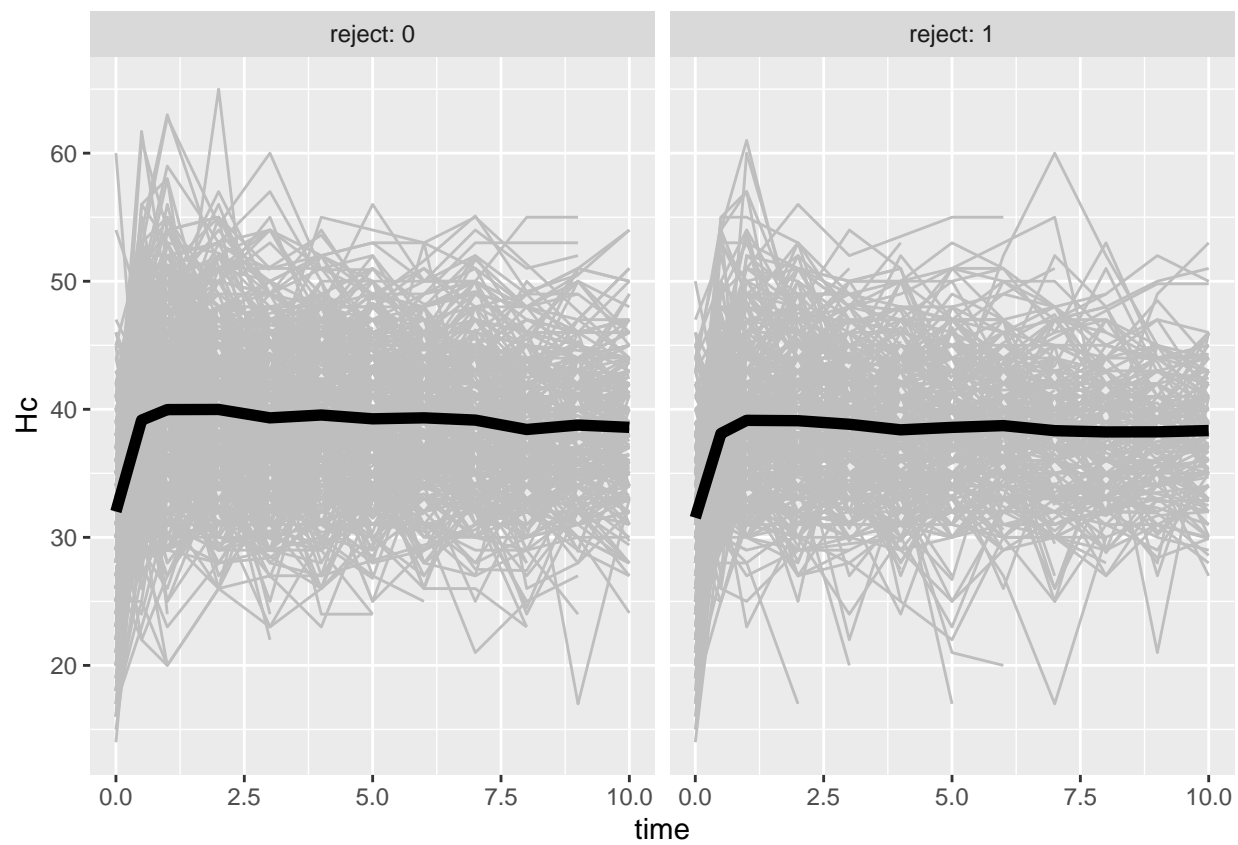
- Spaghetti plot group by male



- Spaghetti plot group by cardio

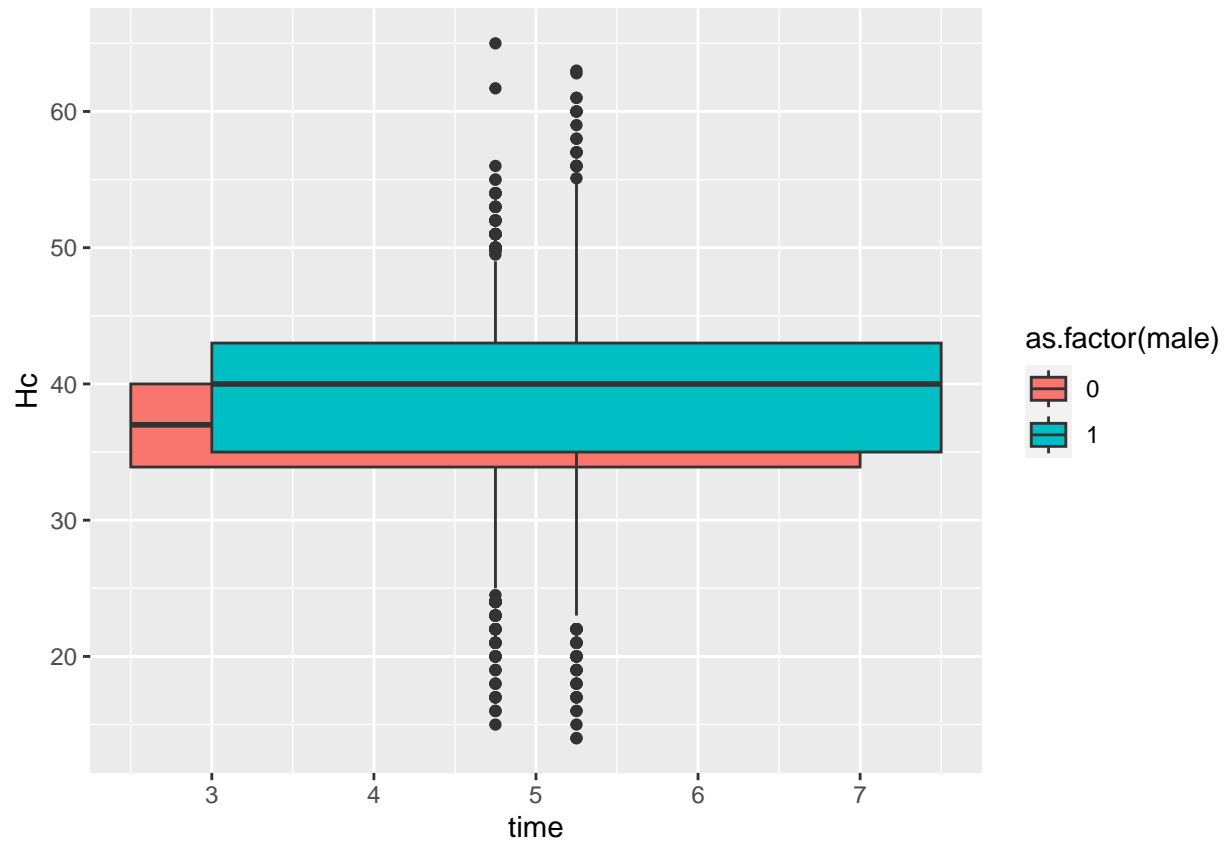


*Spaghetti plot group by reject

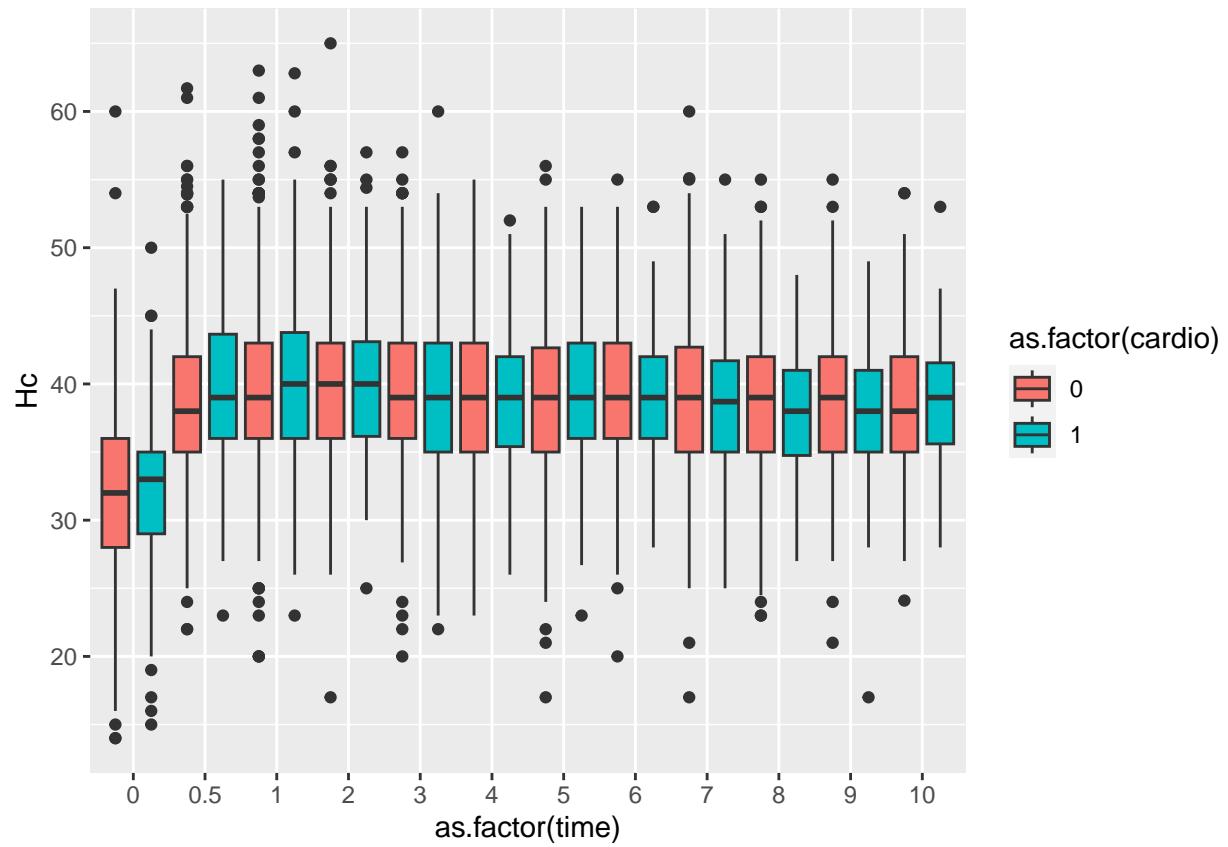


Boxplot

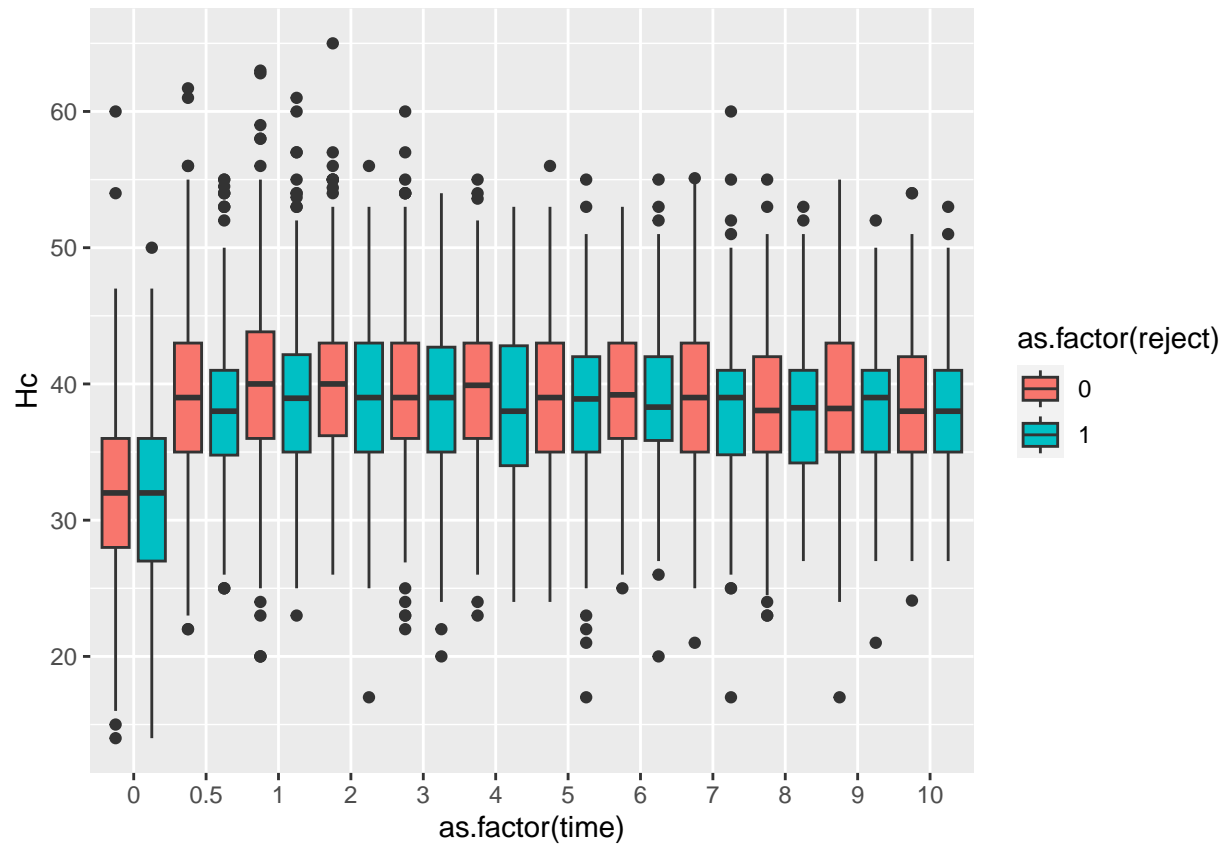
- Box plot by male



- Box plot by cardio

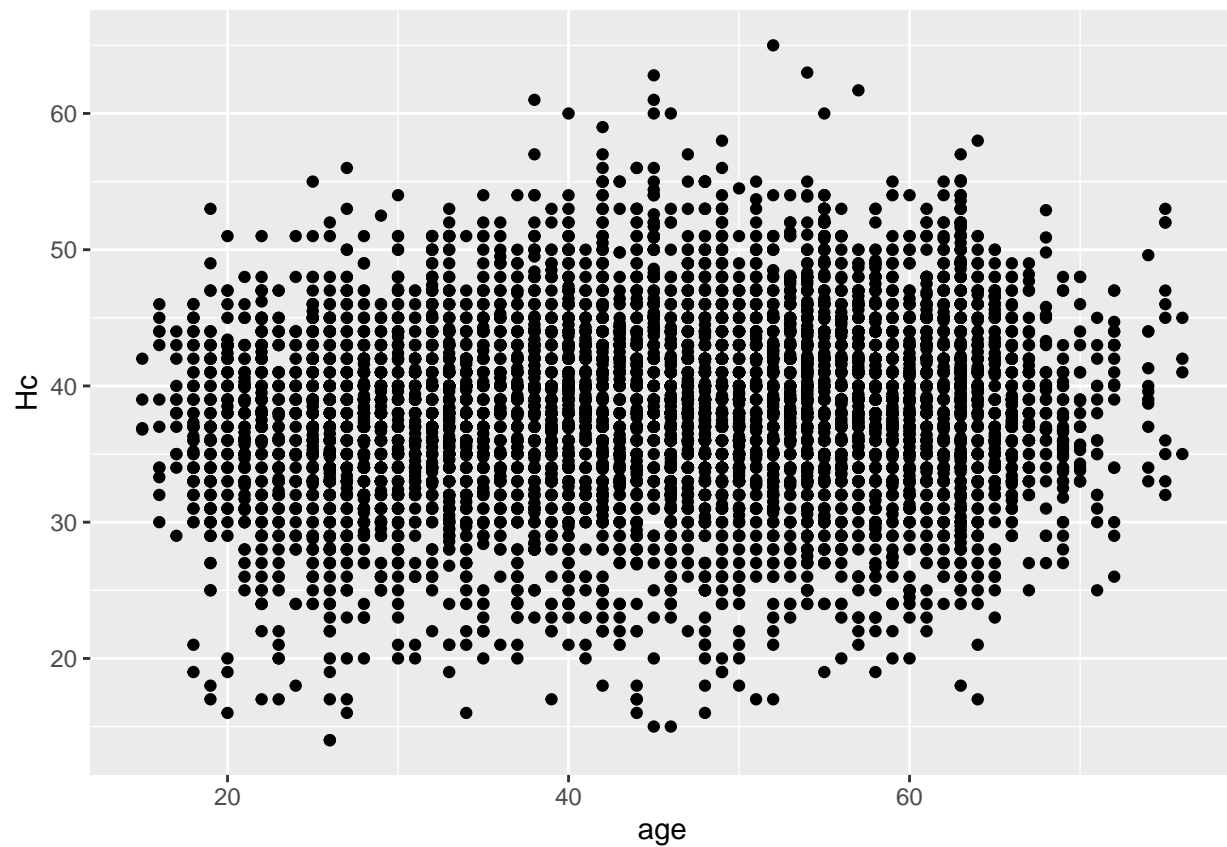


- Box plot by reject

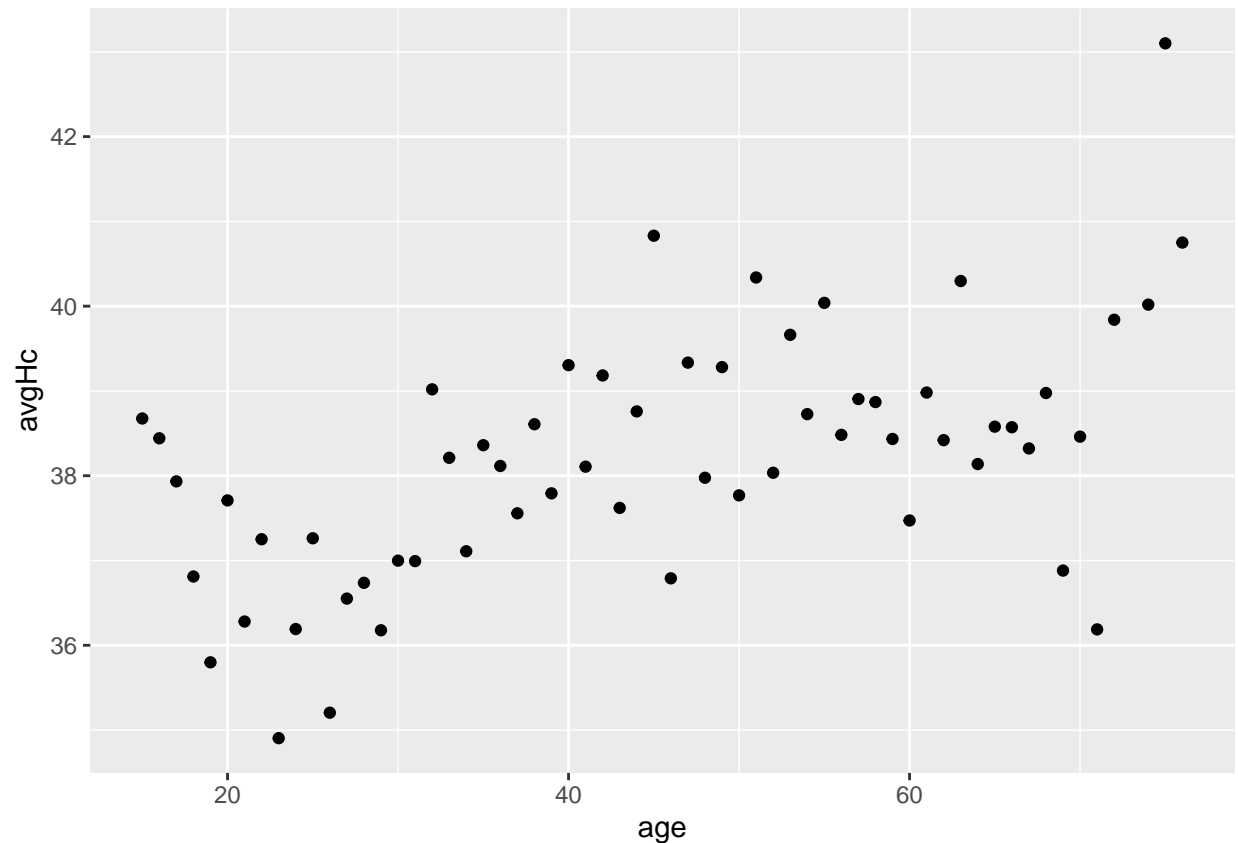


Data set analysis to see the age effect

```
ggplot(data=trenal.long.noNA,aes(y=Hc,x=age))+geom_point()
```



```
data.groupbyage <- trenal.long.noNA %>% group_by(age) %>% summarise(avgHc=mean(Hc))  
ggplot(data=data.groupbyage, aes(y=avgHc, x=age))+geom_point()
```



Conclusions after exploring data analysis

- The Hc time trend tends to increase first from 0 to 0.5 year then keep variated during the rest of the measurements
- The subject related variables age may increase the mean Hc level of a subject
- Male has relative higher Hc level than female
- Cardio or reject play no big difference in the Hc level measurements.

Multilevel Data Analysis

Multivariate Linear Model Analysis

```
lm1 <- lm(Hc ~ time, trenal.long)
summary(lm1)

##
## Call:
## lm(formula = Hc ~ time, data = trenal.long)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.3368  -3.8633   0.0393   3.8206  27.1367
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept) 37.33685    0.09410   396.8   <2e-16 ***
## time         0.26322    0.02073   12.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.023 on 9556 degrees of freedom
## (4362 observations deleted due to missingness)
## Multiple R-squared:  0.01659,    Adjusted R-squared:  0.01648
## F-statistic: 161.2 on 1 and 9556 DF,  p-value: < 2.2e-16

lm2 <- lm(Hc ~ time + age, trenal.long)
summary(lm2)
```

```
##
## Call:
## lm(formula = Hc ~ time + age, data = trenal.long)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.0281  -3.7975  -0.0066   3.8187  26.7575
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.412259   0.238262  144.43   <2e-16 ***
## time         0.290850   0.020655   14.08   <2e-16 ***
## age          0.062472   0.004685   13.33   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.969 on 9548 degrees of freedom
## (4369 observations deleted due to missingness)
## Multiple R-squared:  0.03456,    Adjusted R-squared:  0.03436
## F-statistic: 170.9 on 2 and 9548 DF,  p-value: < 2.2e-16
```

```
lm3 <- lm(Hc ~ time + age + male, trenal.long)
summary(lm3)
```

```
##
## Call:
## lm(formula = Hc ~ time + age + male, data = trenal.long)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.1890  -3.7042   0.1498   3.6798  28.1470
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 32.981301   0.243570  135.41   <2e-16 ***
## time         0.303662   0.020230   15.01   <2e-16 ***
## age          0.062775   0.004586   13.69   <2e-16 ***
## male1        2.457069   0.120470   20.40   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.843 on 9547 degrees of freedom
```

```
## (4369 observations deleted due to missingness)
## Multiple R-squared: 0.07487, Adjusted R-squared: 0.07458
## F-statistic: 257.5 on 3 and 9547 DF, p-value: < 2.2e-16

lm4 <- lm(Hc ~ time + age + male + reject, trenal.long)
summary(lm4)
```

```
##
## Call:
## lm(formula = Hc ~ time + age + male + reject, data = trenal.long)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.2661  -3.6913   0.1555   3.7073  28.0424
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.194768  0.256818 129.254 < 2e-16 ***
## time         0.305407  0.020235  15.093 < 2e-16 ***
## age          0.060616  0.004659  13.011 < 2e-16 ***
## male1        2.443227  0.120550  20.267 < 2e-16 ***
## reject1     -0.336228  0.128591  -2.615  0.00894 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.841 on 9546 degrees of freedom
## (4369 observations deleted due to missingness)
## Multiple R-squared: 0.07553, Adjusted R-squared: 0.07514
## F-statistic: 195 on 4 and 9546 DF, p-value: < 2.2e-16

lm5 <- lm(Hc ~ time + age + male + reject + cardio, trenal.long)
summary(lm5)
```

```
##
## Call:
## lm(formula = Hc ~ time + age + male + reject + cardio, data = trenal.long)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.9829  -3.6915   0.1453   3.6927  27.9570
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.103534  0.259284 127.673 <2e-16 ***
## time         0.305654  0.020229  15.109 <2e-16 ***
## age          0.064002  0.004847  13.204 <2e-16 ***
## male1        2.449916  0.120545  20.324 <2e-16 ***
## reject1     -0.323401  0.128656  -2.514  0.0120 *
## cardio1     -0.417558  0.165619  -2.521  0.0117 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.84 on 9545 degrees of freedom
## (4369 observations deleted due to missingness)
## Multiple R-squared: 0.07615, Adjusted R-squared: 0.07566
```

```
## F-statistic: 157.3 on 5 and 9545 DF, p-value: < 2.2e-16
```

```
anova(lm2,lm3,lm4,lm5)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: Hc ~ time + age
```

```
## Model 2: Hc ~ time + age + male
```

```
## Model 3: Hc ~ time + age + male + reject
```

```
## Model 4: Hc ~ time + age + male + reject + cardio
```

```
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1   9548 340135
```

```
## 2   9547 325933  1  14201.6 416.4704 < 2.2e-16 ***
```

```
## 3   9546 325700  1    233.3   6.8405 0.008925 **
```

```
## 4   9545 325483  1    216.8   6.3564 0.011712 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusions: Variables must keep are: intercept, time, age, male if p value is 0.01, it is better to add reject if p value is 0.05, it is better to add cardio

Two Stage Model Analysis

Linear Mixed effects Model Analysis

This is inspired from the chapter of <https://bookdown.org/roback/bookdown-BeyondMLR/ch-lon.html>
Longitudinal data is a special example of multilevel data, where

- Level One is : time and the response variable e.g. Hc level
- Level Two is : covariates related to each subject, e.g. age, male, reject, cardio

Unconditional Means Model to discover variance distribution

We can first try the unconditional Means Model to explore the variance(within subject and between-subject),
Define Y_{ij} as the Hc level from subject i and measured time j

- Level One:

$$Y_{ij} = a_i + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$

- Level Two:

$$a_i = \alpha_0 + u_i,$$

where $u_i \sim N(0, \sigma_u^2)$

Written in linear mixed effect model is:

$$Y_{ij} = \alpha_0 + u_i + \epsilon_{ij},$$

where $u_i \sim N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$

```
# Model A
```

```
library(lme4)
```

```
model.a <- lmer(Hc ~ 1 + (1|id), REML=T, data=trenal.long)
```

```
summary(model.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: Hc ~ 1 + (1 | id)
```

```
## Data: trenal.long
```

```
##
```

```
## REML criterion at convergence: 59106.7
```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.6999 -0.4371  0.1060  0.5631  6.1641
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   id       (Intercept) 13.60    3.688
##   Residual                23.07    4.803
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  38.1630    0.1206   316.4
##
## AIC = 59112.68 ;BIC = 59134.18
```

From the output of `model.a`, we obtain estimates of three model parameters:

- $\hat{\alpha}_0 = 38.16$: the mean of Hc level μ_{Hc} across all subjects and all years
- $\hat{\sigma}^2 = 23.07$: the variance in within-subjects deviation, between years of measurements Hc_j and the mean μ_{Hc} across all subjects and all years
- $\hat{\sigma}_u^2 = 13.60$: the variance in between-subjects deviation, between subject mean μ_{Hc_i} and the overall mean μ_{Hc} across all subjects and all years.

The intraclass correlation coefficient:

$$\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}^2} = \frac{13.60}{13.60 + 23.07} = 0.371$$

37.1% of the total variation in Hc levels is attributable to differences among subjects rather than changes over time within each subject.

Unconditional Growth Model, introducing time in Level One

- Level One:

$$Y_{ij} = a_i + b_i \times time_{ij} + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$

- Level Two:

$$a_i = \alpha_0 + u_i, \tag{1}$$

$$b_i = \beta_0 + v_i \tag{2}$$

where $\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$

Written in linear mixed effect model is:

$$Y_{ij} = [\alpha_0 + \beta_0 \times time_{ij}] + [u_i + v_i \times time_{ij} + \epsilon_{ij}],$$

where $\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$

```
# model b
model.b <- lmer(Hc ~ time + (time|id), REML=T, data= trenal.long)
summary(model.b)
```



```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time + (time | id)
## Data: trenal.long
##
## REML criterion at convergence: 58690.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.2501 -0.4524  0.0685  0.5468  6.4063
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## id      (Intercept)    13.5731   3.6842
##         time              0.1856   0.4308  -0.15
## Residual                    20.8883   4.5704
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  37.26780    0.13057   285.4
## time          0.31583    0.02375    13.3
##
## Correlation of Fixed Effects:
##      (Intr)
## time -0.394
##
## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
```

From the `model.b`, we obtain estimates of our six model parameters:

- $\hat{\alpha}_0 = 37.2678$: the mean Hc level for the subjects at time 0, Hc_0
- $\hat{\beta}_0 = 0.31583$: the mean change in successively measurements during totally 12 measurements
- $\hat{\sigma}^2 = 20.8883$: the variance in within-subject deviations
- $\hat{\sigma}_u^2 = 13.5731$: the variance between subjects at time 0, Hc_0
- $\hat{\sigma}_v^2 = 0.1856$: the variance between subjects in rate of changes in Hc level
- $\rho_{uv} = -0.15$: the correlation in subject's Hc_0 and the rate of change in Hc level

The estimated within-subject variance $\hat{\sigma}^2$ decreased by about 9% from the unconditional means model implying that 9% of within-subject variability in Hc level can be explained by a linear increase over time:

$$PseudoR_{L1}^2 = \frac{\hat{\sigma}^2(\text{uncond.means}) - \hat{\sigma}^2(\text{uncond.growth})}{\hat{\sigma}^2(\text{uncond.growth})} = \frac{23.07 - 20.8883}{23.07} = 0.0948$$

```
model.b1 <- lmer(Hc ~ time + (1|id),REML=T,data= trenal.long)
summary(model.b1)
```

Unconditional growth but only random intercepts

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time + (1 | id)
## Data: trenal.long
##
## REML criterion at convergence: 58849.2
##
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -5.5197 -0.4741  0.0847  0.5683  6.4131
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   id       (Intercept) 13.69    3.700
##   Residual                22.36    4.729
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  37.2953    0.1317  283.24
## time         0.2913    0.0178   16.36
##
## Correlation of Fixed Effects:
##      (Intr)
## time -0.402
##
## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
```

Modeling other trends over time quadratic

From the spaghetti plots we notice that our Hc level usually increases first very quickly then stays stable. (piecewise linear?) To reduce the correlation between the linear and quadratic components of time effect, we need to center the time variable first:

```
trenal.long.center <- trenal.long%>%
  mutate(timec = time - 5, timec2 = timec^2)
```

- Level One:

$$Y_{ij} = a_i + b_i \times time_{ij} + c_i \times time_{ij}^2 + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$

- Level Two:

$$a_i = \alpha_0 + u_i, \quad (3)$$

$$b_i = \beta_0 + v_i, \quad (4)$$

$$c_i = \gamma_0 + w_i, \quad (5)$$

$$\text{where } \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v & \rho_{uw}\sigma_u\sigma_w \\ & \sigma_v^2 & \rho_{vw}\sigma_v\sigma_w \\ & & \sigma_w^2 \end{bmatrix} \right)$$

Written in linear mixed effect model is:

$$Y_{ij} = [\alpha_0 + \beta_0 \times time_{ij} + \gamma_0 \times time_{ij}^2] + [u_i + v_i \times time_{ij} + w_i \times time_{ij}^2 + \epsilon_{ij}],$$

$$\text{where } \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v & \rho_{uw}\sigma_u\sigma_w \\ & \sigma_v^2 & \rho_{vw}\sigma_v\sigma_w \\ & & \sigma_w^2 \end{bmatrix} \right) \text{ and } \epsilon_{ij} \sim N(0, \sigma^2)$$

```
model.c <- lmer(Hc ~ timec + timec2 + (timec + timec2|id),REML=T, data=trenal.long.center)
summary(model.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ timec + timec2 + (timec + timec2 | id)
## Data: trenal.long.center
##
## REML criterion at convergence: 57865.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.7256 -0.4767  0.0361  0.5308  6.7119
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## id (Intercept) 21.436916 4.63000
## timec 0.197651 0.44458 0.27
## timec2 0.007685 0.08767 -0.82 -0.44
## Residual 18.231536 4.26984
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 40.201415 0.162783 246.963
## timec 0.055012 0.024108 2.282
## timec2 -0.161860 0.006341 -25.526
##
## Correlation of Fixed Effects:
## (Intr) timec
## timec 0.244
## timec2 -0.590 0.198
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0353379 (tol = 0.002, component 1)
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?

## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
```

```
model.c1 <- lmer(Hc ~ timec + timec2 + (1|id),REML=T, data=trenal.long.center)
summary(model.c1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ timec + timec2 + (1 | id)
## Data: trenal.long.center
##
## REML criterion at convergence: 58237
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.4559 -0.4917  0.0467  0.5568  6.4954
```

```

##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   id       (Intercept) 13.91    3.730
##   Residual                20.77    4.558
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) 40.179999   0.137545 292.122
## timec       0.103646    0.018717   5.538
## timec2      -0.149817    0.005903 -25.379
##
## Correlation of Fixed Effects:
##      (Intr) timec
## timec  0.071
## timec2 -0.409  0.396
##
## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83

```

Piecewise linear time trend

In the **piecewise linear model**, the complete time span of the study is divided into two segments, with a separate slope relating time to the response in each segment.

- Level One:

$$Y_{ij} = a_i + b_i \times time_{1ij} + c_i \times time_{2ij} + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$

- Level Two:

$$a_i = \alpha_0 + u_i, \tag{6}$$

$$b_i = \beta_0 + v_i, \tag{7}$$

$$c_i = \gamma_0 + w_i, \tag{8}$$

$$\text{where } \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v & \rho_{uw}\sigma_u\sigma_w \\ & \sigma_v^2 & \rho_{vw}\sigma_v\sigma_w \\ & & \sigma_w^2 \end{bmatrix} \right)$$

Written in linear mixed effect model is:

$$Y_{ij} = [\alpha_0 + \beta_0 \times time_{1ij} + \gamma_0 \times time_{2ij}] + [u_i + v_i \times time_{1ij} + w_i \times time_{2ij} + \epsilon_{ij}],$$

$$\text{where } \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v & \rho_{uw}\sigma_u\sigma_w \\ & \sigma_v^2 & \rho_{vw}\sigma_v\sigma_w \\ & & \sigma_w^2 \end{bmatrix} \right) \text{ and } \epsilon_{ij} \sim N(0, \sigma^2)$$

In our case study, we can fit separate slope in time 0 – 0.5 and 0.5 – 10

```

# Modeling piecewise linear time trend with two intervals
time1 = trenal.long$time
time1[time1>0.5] = 0
time2 = trenal.long$time
time2[time2<1] = 0

trenal.long.piecewise = trenal.long
trenal.long.piecewise['time1'] <- time1
trenal.long.piecewise['time2'] <- time2

model.b.piecewise <- lmer(Hc ~ time1 + time2 + (1|id),REML=T,data=trenal.long.piecewise)

summary(model.b.piecewise)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time1 + time2 + (1 | id)
## Data: trenal.long.piecewise
##
## REML criterion at convergence: 58716.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.4595 -0.4903  0.0741  0.5791  6.5397
##
## Random effects:
## Groups Name Variance Std.Dev.
## id      (Intercept) 13.76  3.709
## Residual                22.01  4.692
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 36.81383    0.13799 266.78
## time1        4.02855    0.32348  12.45
## time2        0.36597    0.01881  19.46
##
## Correlation of Fixed Effects:
##      (Intr) time1
## time1 -0.322
## time2 -0.461  0.393

## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83
## model.b.piecewise: AIC = 58726.86 ;BIC = 58762.69
model.b.piecewise1 <- lmer(Hc ~ time1 + time2 + (time2|id),REML=T,data=trenal.long.piecewise)
summary(model.b.piecewise1)

## Linear mixed model fit by REML ['lmerMod']

```

```
## Formula: Hc ~ time1 + time2 + (time2 | id)
## Data: trenal.long.piecewise
##
## REML criterion at convergence: 58545.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.1940 -0.4729  0.0625  0.5607  6.5532
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## id      (Intercept) 13.6532   3.6950
##         time2      0.1867   0.4321  -0.15
## Residual                20.4690   4.5243
## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 36.76665    0.13678  268.79
## time1       4.12291    0.31360   13.15
## time2       0.40031    0.02467   16.22
##
## Correlation of Fixed Effects:
##      (Intr) time1
## time1 -0.324
## time2 -0.439  0.330
##
## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83
## model.b.piecewise: AIC = 58726.86 ;BIC = 58762.69
## model.b.piecewise1: AIC = 58559.93 ;BIC = 58610.09
```

From the AIC and BIC value, we can see the quadratic model `model.c` outperforms the piecewise linear model `model.b.piecewise1`. However, we believe that in reality the *Hc* level of a person will not change quadratically. It makes more sense that the *Hc* level of a patient will increase faster the first half year of his or her operation, and then keep stable with probably some random effects to change over the following years.

With the level one model fixed, we can consider adding level two variables sequentially.

Adding subject related variable in level two

```
model.b.piecewise.age <- lmer(Hc ~ time1 + time2 + age + (time2|id), REML=T, data=trenal.long.piecewise)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time1 + time2 + age + (time2 | id)
## Data: trenal.long.piecewise
##
## REML criterion at convergence: 58467.8
##
```

```

## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.1368 -0.4701  0.0627  0.5628  6.5372
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   id       (Intercept) 13.2327  3.6377
##           time2        0.1872  0.4327  -0.18
##   Residual                20.4668  4.5240
## Number of obs: 9551, groups: id, 1159
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 34.009541  0.434914  78.198
## time1       4.116852  0.313703  13.123
## time2       0.404186  0.024665  16.387
## age         0.059373  0.008894   6.676
##
## Correlation of Fixed Effects:
##      (Intr) time1  time2
## time1 -0.104
## time2 -0.172  0.330
## age   -0.950  0.002  0.032
## optimizer (nlptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00365004 (tol = 0.002, component 1)

## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83
## model.b.piecewise: AIC = 58726.86 ;BIC = 58762.69
## model.b.piecewise1: AIC = 58559.93 ;BIC = 58610.09
## model.b.piecewise,age: AIC = 58483.85 ;BIC = 58541.16
model.b.piecewise.male <- lmer(Hc ~ time1 + time2 + male + (time2|id),REML=T,data=trenal.long.piecewise)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time1 + time2 + male + (time2 | id)
##   Data: trenal.long.piecewise
##
## REML criterion at convergence: 58448.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.2294 -0.4679  0.0612  0.5625  6.5926
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   id       (Intercept) 12.4340  3.5262
##           time2        0.1865  0.4319  -0.18
##   Residual                20.4734  4.5248

```

```

## Number of obs: 9558, groups: id, 1160
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept) 35.41528    0.18827  188.10
## time1       4.11679    0.31361   13.13
## time2       0.40188    0.02462   16.32
## male1       2.35907    0.23253   10.14
##
## Correlation of Fixed Effects:
##      (Intr) time1  time2
## time1 -0.234
## time2 -0.334  0.331
## male1 -0.708 -0.001  0.010
##
## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83
## model.b.piecewise: AIC = 58726.86 ;BIC = 58762.69
## model.b.piecewise1: AIC = 58559.93 ;BIC = 58610.09
## model.b.piecewise,age: AIC = 58483.85 ;BIC = 58541.16
## model.b.piecewise,male: AIC = 58464.65 ;BIC = 58521.97
model.b.piecewise.maleage <- lmer(Hc ~ time1 + time2 + male +age +(time2|id),REML=T,data=trenal.long.pi

## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time1 + time2 + male + age + (time2 | id)
## Data: trenal.long.piecewise
##
## REML criterion at convergence: 58368
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.1683 -0.4680  0.0637  0.5619  6.5758
##
## Random effects:
## Groups   Name      Variance Std.Dev. Corr
## id      (Intercept) 12.0382  3.4696
##         time2       0.1874  0.4329  -0.21
## Residual          20.4704  4.5244
## Number of obs: 9551, groups: id, 1159
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept) 32.694557    0.435712  75.037
## time1       4.112153    0.313713  13.108
## time2       0.406601    0.024616  16.518
## male1       2.346642    0.228052  10.290
## age         0.058747    0.008512   6.902

```



```

##
## Correlation of Fixed Effects:
##      (Intr) time1  time2  male1
## time1 -0.104
## time2 -0.180  0.331
## male1 -0.290 -0.001  0.011
## age   -0.905  0.003  0.035 -0.011
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00794242 (tol = 0.002, component 1)

## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83
## model.b.piecewise: AIC = 58726.86 ;BIC = 58762.69
## model.b.piecewise1: AIC = 58559.93 ;BIC = 58610.09
## model.b.piecewise,age: AIC = 58483.85 ;BIC = 58541.16
## model.b.piecewise,male: AIC = 58464.65 ;BIC = 58521.97
## model.b.piecewise,maleage: AIC = 58386 ;BIC = 58450.48
model.b.piecewise.maleagereject <- lmer(Hc ~ time1 + time2 + male + age + reject +(time2|id),REML=T,data=

## Linear mixed model fit by REML ['lmerMod']
## Formula: Hc ~ time1 + time2 + male + age + (time2 | id)
## Data: trenal.long.piecewise
##
## REML criterion at convergence: 58368
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.1683 -0.4680  0.0637  0.5619  6.5758
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## id      (Intercept) 12.0382 3.4696
##         time2      0.1874 0.4329 -0.21
## Residual      20.4704 4.5244
## Number of obs: 9551, groups: id, 1159
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 32.694557  0.435712  75.037
## time1       4.112153  0.313713  13.108
## time2       0.406601  0.024616  16.518
## male1       2.346642  0.228052  10.290
## age         0.058747  0.008512   6.902
##
## Correlation of Fixed Effects:
##      (Intr) time1  time2  male1
## time1 -0.104

```

```
## time2 -0.180 0.331
## male1 -0.290 -0.001 0.011
## age -0.905 0.003 0.035 -0.011
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00794242 (tol = 0.002, component 1)

## model.a: AIC = 59112.68 ;BIC = 59134.18
## model.b: AIC = 58702.37 ;BIC = 58745.36
## model.b1: AIC = 58857.21 ;BIC = 58885.87
## model.c: AIC = 57885.46 ;BIC = 57957.11
## model.c1: AIC = 58247.01 ;BIC = 58282.83
## model.b.piecewise: AIC = 58726.86 ;BIC = 58762.69
## model.b.piecewise1: AIC = 58559.93 ;BIC = 58610.09
## model.b.piecewise,age: AIC = 58483.85 ;BIC = 58541.16
## model.b.piecewise,male: AIC = 58464.65 ;BIC = 58521.97
## model.b.piecewise,maleage: AIC = 58386 ;BIC = 58450.48
## model.b.piecewise,maleagereject: AIC = 58386.74 ;BIC = 58458.38
```

```
drop_in_dev <- anova(model.b.piecewise1,model.b.piecewise.male,test="Chisq")
drop_in_dev
```

Comparing nested model using anova

```
## Data: trenal.long.piecewise
## Models:
## model.b.piecewise1: Hc ~ time1 + time2 + (time2 | id)
## model.b.piecewise.male: Hc ~ time1 + time2 + male + (time2 | id)
##
##          npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## model.b.piecewise1      7 58551 58602 -29269    58537
## model.b.piecewise.male    8 58455 58512 -29220    58439 98.453   1 < 2.2e-16
##
## model.b.piecewise1
## model.b.piecewise.male ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
drop_in_dev <- anova(model.b.piecewise.age,model.b.piecewise.maleage,model.b.piecewise.maleagereject,test="Chisq")
drop_in_dev
```

```
## Data: trenal.long.piecewise
## Models:
## model.b.piecewise.age: Hc ~ time1 + time2 + age + (time2 | id)
## model.b.piecewise.maleage: Hc ~ time1 + time2 + male + age + (time2 | id)
## model.b.piecewise.maleagereject: Hc ~ time1 + time2 + male + age + reject + (time2 | id)
##
##          npar   AIC   BIC logLik deviance  Chisq Df
## model.b.piecewise.age      8 58468 58525 -29226    58452
## model.b.piecewise.maleage    9 58368 58433 -29175    58350 101.1485 1
## model.b.piecewise.maleagereject 10 58368 58440 -29174    58348 2.2337 1
##
##          Pr(>Chisq)
## model.b.piecewise.age
```

```
## model.b.piecewise.maleage <2e-16 ***
## model.b.piecewise.maleagereject 0.135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finally, our optimal model would be the piecewise linear model with age gender as the level two variables `model.b.piecewise.maleage`.

- Level One:

$$Y_{ij} = a_i + b_i \times time_{1ij} + c_i \times time_{2ij} + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$

- Level Two:

$$a_i = \alpha_0 + \alpha_1 \times male_i + \alpha_2 \times age_i + u_i, \quad (9)$$

$$b_i = \beta_0 + \beta_1 \times male_i + \beta_2 \times age_i + v_i, \quad (10)$$

$$c_i = \gamma_0 + \gamma_1 \times male_i + \gamma_2 \times age_i + w_i, \quad (11)$$

$$\text{where } \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v & \rho_{uw}\sigma_u\sigma_w \\ & \sigma_v^2 & \rho_{vw}\sigma_v\sigma_w \\ & & \sigma_w^2 \end{bmatrix} \right)$$

The simplified linear mixed effect model is:

$$Y_{ij} = 32.69 + 4.112 \times time_{1ij} + 0.4066 \times time_{2ij} + 2.347 \times male_i + 0.058747 \times age_i + u_i + v_i$$

$$Y_{ij} = [\alpha_0 + \beta_0 \times time_{1ij} + \gamma_0 \times time_{2ij}] + [u_i + v_i \times time_{1ij} + w_i \times time_{2ij} + \epsilon_{ij}],$$

$$\text{where } \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v & \rho_{uw}\sigma_u\sigma_w \\ & \sigma_v^2 & \rho_{vw}\sigma_v\sigma_w \\ & & \sigma_w^2 \end{bmatrix} \right) \text{ and } \epsilon_{ij} \sim N(0, \sigma^2)$$