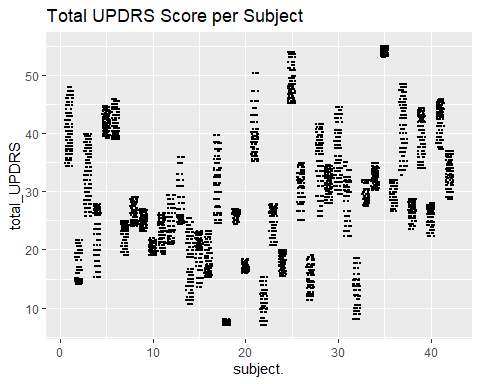
project\_option2

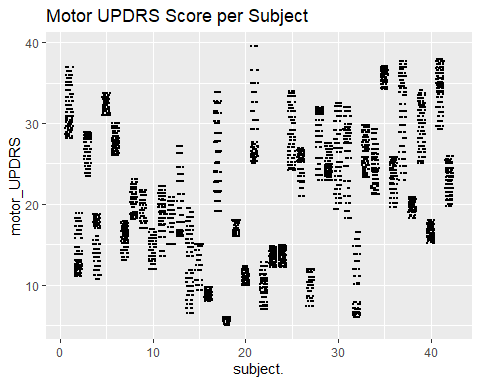
categorical\_col = c(1,2,3)  
full\_data = read.csv("D:/R/data\_467/data467\_project/parkinsons\_updrs.data")  
data = full\_data[-categorical\_col]  
for (i in 1:length(data)){  
 data[,i] = scale(data[,i], center = TRUE, scale = TRUE)  
}  
subject\_data = full\_data  
head(data)

## test\_time motor\_UPDRS total\_UPDRS Jitter... Jitter.Abs. Jitter.RAP  
## 1 -1.6319513 0.8491244 0.5027024 0.08289818 -0.2842180 0.3274246  
## 2 -1.5005486 0.8796314 0.5490563 -0.56074568 -0.7566586 -0.5337008  
## 3 -1.3692936 0.9101384 0.5953167 -0.23892375 -0.5393359 -0.3000125  
## 4 -1.2576661 0.9359710 0.6346615 -0.15535673 -0.4851442 -0.3448294  
## 5 -1.1080747 0.9706604 0.6874638 -0.49851492 -0.6638379 -0.6585480  
## 6 -0.9769133 1.0011674 0.7337243 -0.46651053 -0.5871358 -0.5753165  
## Jitter.PPQ5 Jitter.DDP Shimmer Shimmer.dB. Shimmer.APQ3 Shimmer.APQ5  
## 1 -0.0286346 0.3284775 -0.3245661 -0.3516122 -0.2096910 -0.4233205  
## 2 -0.4761714 -0.5347790 -0.5339707 -0.5731071 -0.5451114 -0.5655438  
## 3 -0.3207395 -0.2989574 -0.6690579 -0.5644210 -0.7415288 -0.7023662  
## 4 -0.1706673 -0.3448413 -0.4236559 0.0696623 -0.4605008 -0.4497248  
## 5 -0.5297687 -0.6596258 -0.6582200 -0.5861362 -0.7830786 -0.6513579  
## 6 -0.4520527 -0.5753275 -0.4553956 -0.4211008 -0.5360460 -0.4065177  
## Shimmer.APQ11 Shimmer.DDA NHR HNR RPDE DFA  
## 1 -0.5434195 -0.2096865 -0.29869541 -0.009203976 -1.21396224 -1.478374  
## 2 -0.5299101 -0.5451066 -0.35193510 1.282540523 -1.05502898 -1.247774  
## 3 -0.6454902 -0.7415238 -0.19935262 0.318684263 -0.78479294 -1.540008  
## 4 -0.3928152 -0.4607479 -0.07174823 0.644475136 -0.53644116 -1.062024  
## 5 -0.4648651 -0.7833254 -0.34334103 1.036216523 -0.68913591 -1.297843  
## 6 -0.2427112 -0.5357894 -0.37997892 0.295147154 -0.01963576 -1.139737  
## PPE  
## 1 -0.6506028  
## 2 -1.2184810  
## 3 -0.1032714  
## 4 1.2369692  
## 5 -0.2839301  
## 6 -0.2687386

#TOTAL UPDRS  
ggplot(full\_data, aes(subject., total\_UPDRS)) +  
 geom\_jitter(size = 0.01)+  
 ggtitle('Total UPDRS Score per Subject')



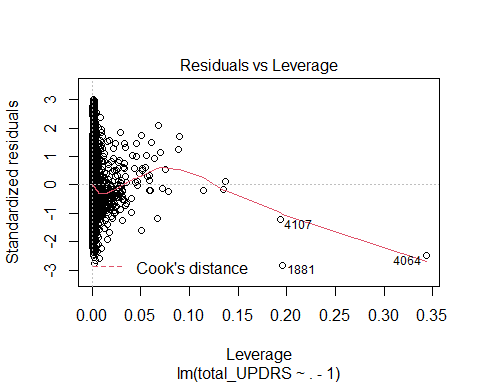
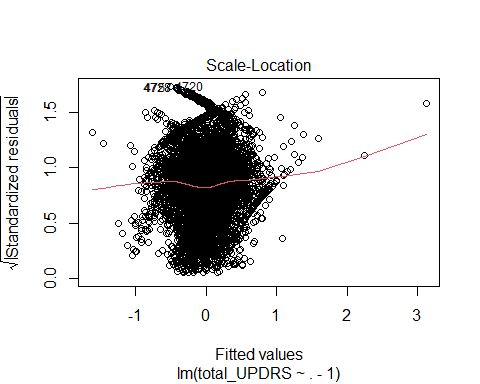
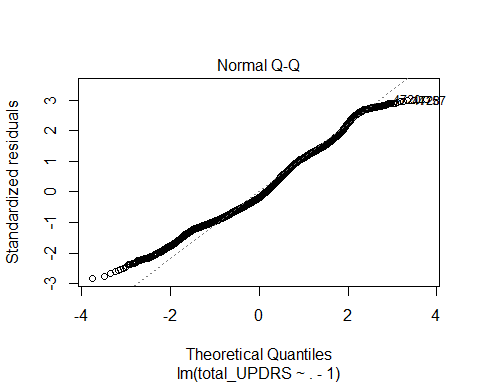
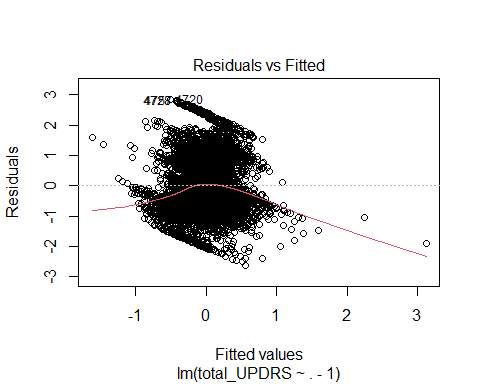
#MOTOR UPDRS  
ggplot(full\_data, aes(subject., motor\_UPDRS)) +  
 geom\_jitter(size = 0.01)+  
 ggtitle('Motor UPDRS Score per Subject')



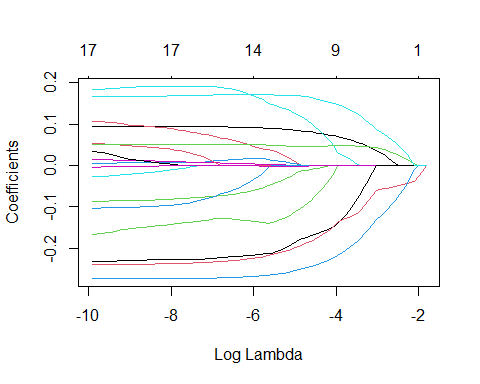
set.seed(7)  
  
tot\_df = data[-2] #Removing motor\_UPDRS  
  
#FULL MODEL  
full\_model = lm(total\_UPDRS ~.-1, data = tot\_df)  
summary(full\_model)

##   
## Call:  
## lm(formula = total\_UPDRS ~ . - 1, data = tot\_df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.6178 -0.6973 -0.1820 0.7077 2.8463   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## test\_time 0.08829 0.01240 7.121 1.20e-12 \*\*\*  
## Jitter... 0.06863 0.11639 0.590 0.555426   
## Jitter.Abs. -0.10601 0.03220 -3.293 0.000999 \*\*\*  
## Jitter.RAP -13.80989 14.19452 -0.973 0.330642   
## Jitter.PPQ5 -0.02475 0.06819 -0.363 0.716636   
## Jitter.DDP 13.88979 14.19602 0.978 0.327903   
## Shimmer 0.28441 0.16238 1.751 0.079917 .   
## Shimmer.dB. -0.10831 0.10810 -1.002 0.316407   
## Shimmer.APQ3 -25.52941 60.42833 -0.422 0.672694   
## Shimmer.APQ5 -0.20346 0.08931 -2.278 0.022755 \*   
## Shimmer.APQ11 0.20057 0.04799 4.179 2.97e-05 \*\*\*  
## Shimmer.DDA 25.32614 60.42825 0.419 0.675150   
## NHR -0.23069 0.03502 -6.587 4.87e-11 \*\*\*  
## HNR -0.24593 0.02859 -8.602 < 2e-16 \*\*\*  
## RPDE 0.05478 0.01778 3.082 0.002068 \*\*   
## DFA -0.27176 0.01561 -17.414 < 2e-16 \*\*\*  
## PPE 0.17956 0.02588 6.937 4.43e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9456 on 5858 degrees of freedom  
## Multiple R-squared: 0.1083, Adjusted R-squared: 0.1057   
## F-statistic: 41.85 on 17 and 5858 DF, p-value: < 2.2e-16

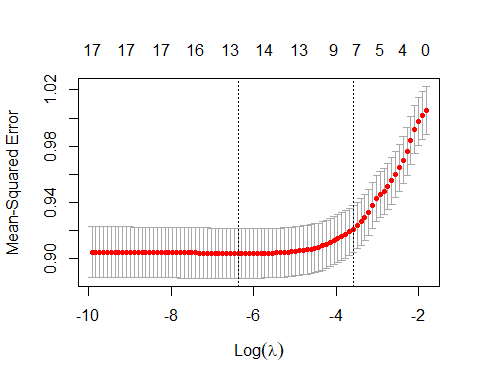
plot(full\_model)



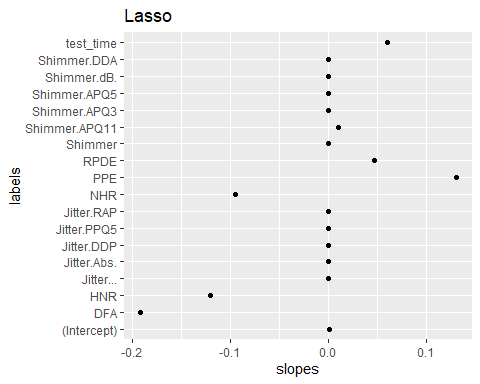
#CROSS VALIDATION  
tot\_split = initial\_split(tot\_df , prop = 0.7, strata = 'total\_UPDRS')  
tot\_train = training(tot\_split); tot\_test = testing(tot\_split)  
tot\_train\_y = tot\_train$total\_UPDRS  
tot\_train\_x = model.matrix(lm(total\_UPDRS ~ .-1, data = tot\_train))  
  
  
# #RIDGE  
# tot\_ridge = glmnet(x = tot\_train\_x, y = tot\_train\_y, alpha = 0)  
# plot(tot\_ridge, xvar = 'lambda')  
#   
# tot\_ridge\_cv = cv.glmnet(x = tot\_train\_x, y = tot\_train\_y, alpha = 0)  
# plot(tot\_ridge\_cv)  
#   
# ridge\_coef = coef(tot\_ridge, tot\_ridge\_cv$lambda.1se)  
# ridge\_coef\_df = as.data.frame(as.matrix(ridge\_coef))  
# colnames(ridge\_coef\_df) = c('slopes')  
# ridge\_coef\_df$labels = rownames(ridge\_coef\_df)  
# ggplot(data = ridge\_coef\_df, mapping = aes(x = slopes, y = labels)) + geom\_point()+ggtitle('Ridge')  
# ridge\_coef\_df  
  
#LASSO  
tot\_lasso = glmnet(x = tot\_train\_x, y = tot\_train\_y, alpha = 1)  
plot(tot\_lasso, xvar = 'lambda')



tot\_lasso\_cv = cv.glmnet(x = tot\_train\_x, y = tot\_train\_y, alpha = 1)  
plot(tot\_lasso\_cv)



lasso\_coef = coef(tot\_lasso, tot\_lasso\_cv$lambda.1se)  
lasso\_coef\_df = as.data.frame(as.matrix(lasso\_coef))  
colnames(lasso\_coef\_df) = c('slopes')  
lasso\_coef\_df$labels = rownames(lasso\_coef\_df)  
ggplot(data = lasso\_coef\_df, mapping = aes(x = slopes, y = labels)) + geom\_point() +ggtitle('Lasso')



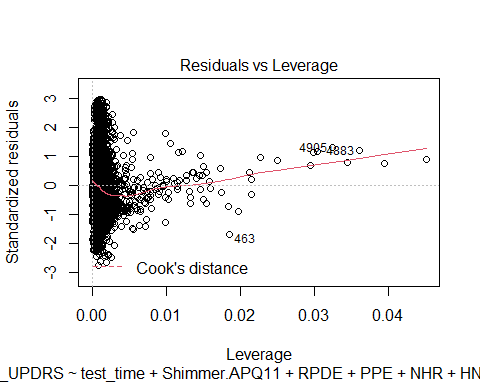
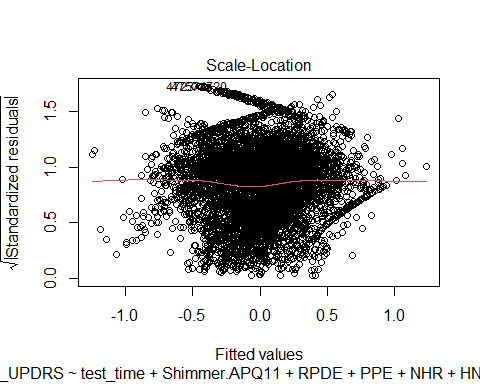
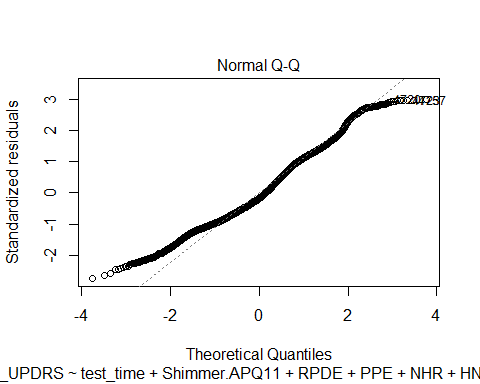
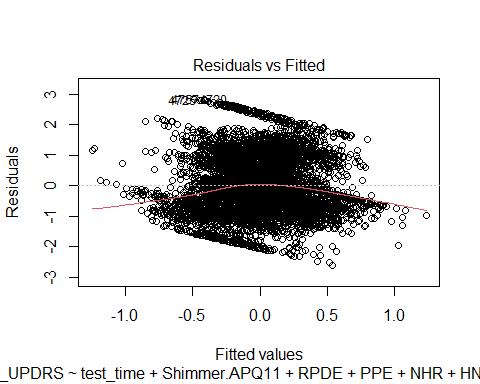
lasso\_coef\_df

## slopes labels  
## (Intercept) 0.0009700013 (Intercept)  
## test\_time 0.0596532999 test\_time  
## Jitter... 0.0000000000 Jitter...  
## Jitter.Abs. 0.0000000000 Jitter.Abs.  
## Jitter.RAP 0.0000000000 Jitter.RAP  
## Jitter.PPQ5 0.0000000000 Jitter.PPQ5  
## Jitter.DDP 0.0000000000 Jitter.DDP  
## Shimmer 0.0000000000 Shimmer  
## Shimmer.dB. 0.0000000000 Shimmer.dB.  
## Shimmer.APQ3 0.0000000000 Shimmer.APQ3  
## Shimmer.APQ5 0.0000000000 Shimmer.APQ5  
## Shimmer.APQ11 0.0105209293 Shimmer.APQ11  
## Shimmer.DDA 0.0000000000 Shimmer.DDA  
## NHR -0.0952388334 NHR  
## HNR -0.1200046925 HNR  
## RPDE 0.0473549429 RPDE  
## DFA -0.1917464936 DFA  
## PPE 0.1308031817 PPE

red\_model = lm(total\_UPDRS~ test\_time+ Shimmer.APQ11 + RPDE + PPE + NHR + HNR + DFA -1, data = tot\_df)  
summary(red\_model)

##   
## Call:  
## lm(formula = total\_UPDRS ~ test\_time + Shimmer.APQ11 + RPDE +   
## PPE + NHR + HNR + DFA - 1, data = tot\_df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5971 -0.7127 -0.1682 0.7171 2.8278   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## test\_time 0.08513 0.01243 6.848 8.24e-12 \*\*\*  
## Shimmer.APQ11 0.05281 0.02152 2.454 0.014156 \*   
## RPDE 0.05668 0.01676 3.382 0.000725 \*\*\*  
## PPE 0.19601 0.02053 9.547 < 2e-16 \*\*\*  
## NHR -0.23690 0.01981 -11.959 < 2e-16 \*\*\*  
## HNR -0.17864 0.02678 -6.670 2.79e-11 \*\*\*  
## DFA -0.26997 0.01451 -18.601 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9505 on 5868 degrees of freedom  
## Multiple R-squared: 0.09745, Adjusted R-squared: 0.09638   
## F-statistic: 90.51 on 7 and 5868 DF, p-value: < 2.2e-16

plot(red\_model)



anova1 = anova(red\_model, full\_model)  
anova1

## Analysis of Variance Table  
##   
## Model 1: total\_UPDRS ~ test\_time + Shimmer.APQ11 + RPDE + PPE + NHR +   
## HNR + DFA - 1  
## Model 2: total\_UPDRS ~ (test\_time + Jitter... + Jitter.Abs. + Jitter.RAP +   
## Jitter.PPQ5 + Jitter.DDP + Shimmer + Shimmer.dB. + Shimmer.APQ3 +   
## Shimmer.APQ5 + Shimmer.APQ11 + Shimmer.DDA + NHR + HNR +   
## RPDE + DFA + PPE) - 1  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 5868 5301.6   
## 2 5858 5237.9 10 63.642 7.1176 3.113e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

qf(0.95, anova1$Df[2], anova1$Res.Df[1])

## [1] 1.832312

dt(2.454,5686)

## [1] 0.01966379

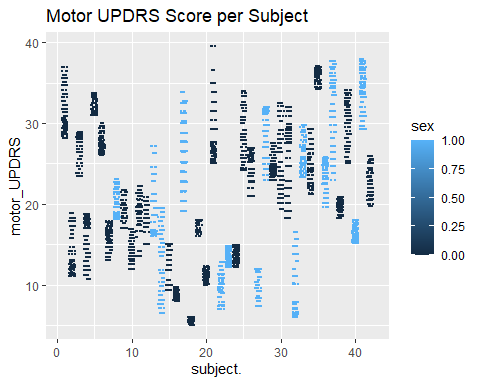
grouped\_subject = subject\_data %>%   
 group\_by(subject.)  
  
head(grouped\_subject)

## # A tibble: 6 x 22  
## # Groups: subject. [1]  
## subject. age sex test\_time motor\_UPDRS total\_UPDRS Jitter... Jitter.Abs.  
## <int> <int> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 72 0 5.64 28.2 34.4 0.00662 0.0000338  
## 2 1 72 0 12.7 28.4 34.9 0.003 0.0000168  
## 3 1 72 0 19.7 28.7 35.4 0.00481 0.0000246  
## 4 1 72 0 25.6 28.9 35.8 0.00528 0.0000266  
## 5 1 72 0 33.6 29.2 36.4 0.00335 0.0000201  
## 6 1 72 0 40.7 29.4 36.9 0.00353 0.0000229  
## # ... with 14 more variables: Jitter.RAP <dbl>, Jitter.PPQ5 <dbl>,  
## # Jitter.DDP <dbl>, Shimmer <dbl>, Shimmer.dB. <dbl>, Shimmer.APQ3 <dbl>,  
## # Shimmer.APQ5 <dbl>, Shimmer.APQ11 <dbl>, Shimmer.DDA <dbl>, NHR <dbl>,  
## # HNR <dbl>, RPDE <dbl>, DFA <dbl>, PPE <dbl>

gs\_lmer = lmer(formula = total\_UPDRS ~ sex + (1 | subject.), data = grouped\_subject)  
summary(gs\_lmer)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: total\_UPDRS ~ sex + (1 | subject.)  
## Data: grouped\_subject  
##   
## REML criterion at convergence: 28948.8  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.2925 -0.5928 0.0396 0.5703 3.7142   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## subject. (Intercept) 110.064 10.491   
## Residual 7.666 2.769   
## Number of obs: 5875, groups: subject., 42  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 29.431 1.983 14.84  
## sex -2.679 3.435 -0.78  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sex -0.577

ggplot(grouped\_subject, aes(subject., motor\_UPDRS, color = sex)) +  
 geom\_jitter(size = 0.01)+  
 ggtitle('Motor UPDRS Score per Subject')



gs\_lm = lm(total\_UPDRS ~ sex, data = grouped\_subject)  
gs\_me = lmer(total\_UPDRS ~ (1 | subject.), data = grouped\_subject)  
anova(gs\_lmer, gs\_me)

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

## Data: grouped\_subject  
## Models:  
## gs\_me: total\_UPDRS ~ (1 | subject.)  
## gs\_lmer: total\_UPDRS ~ sex + (1 | subject.)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)  
## gs\_me 3 28963 28983 -14478 28957   
## gs\_lmer 4 28964 28991 -14478 28956 0.634 1 0.4259

anova(gs\_lmer, gs\_lm)

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

## Data: grouped\_subject  
## Models:  
## gs\_lm: total\_UPDRS ~ sex  
## gs\_lmer: total\_UPDRS ~ sex + (1 | subject.)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## gs\_lm 3 44473 44493 -22234 44467   
## gs\_lmer 4 28964 28991 -14478 28956 15511 1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1