

# hw5

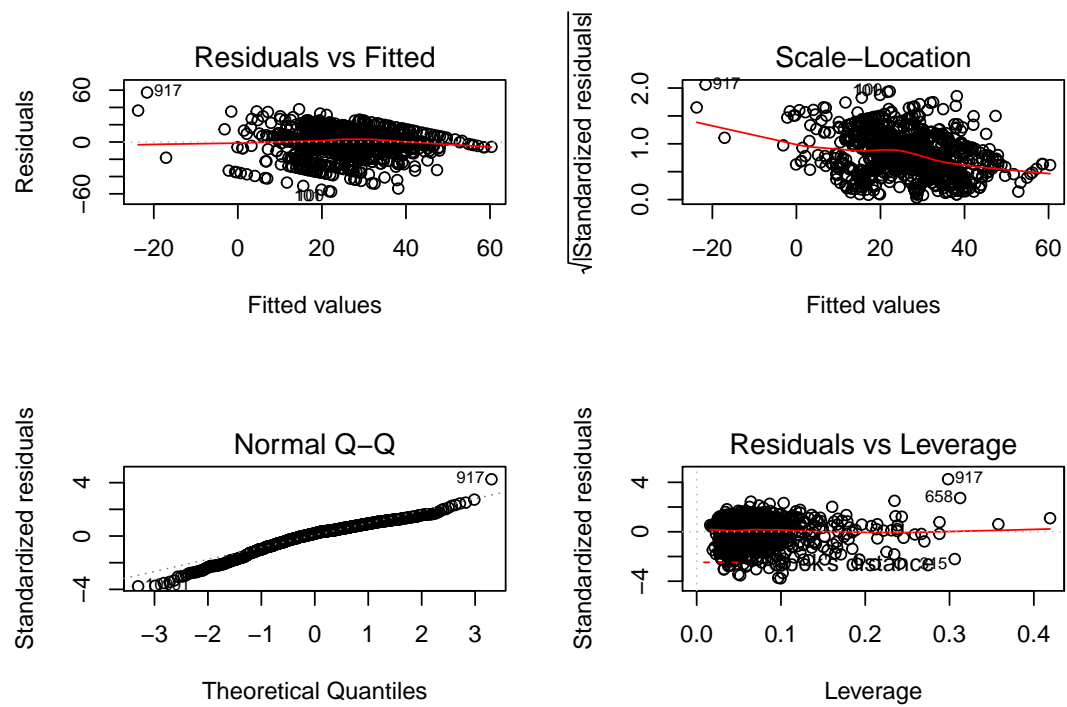
*Huamin Zhang*

*Mar 19, 2017*

## 1.1

```
set.seed(1)
library(MASS)
library(glmnet)

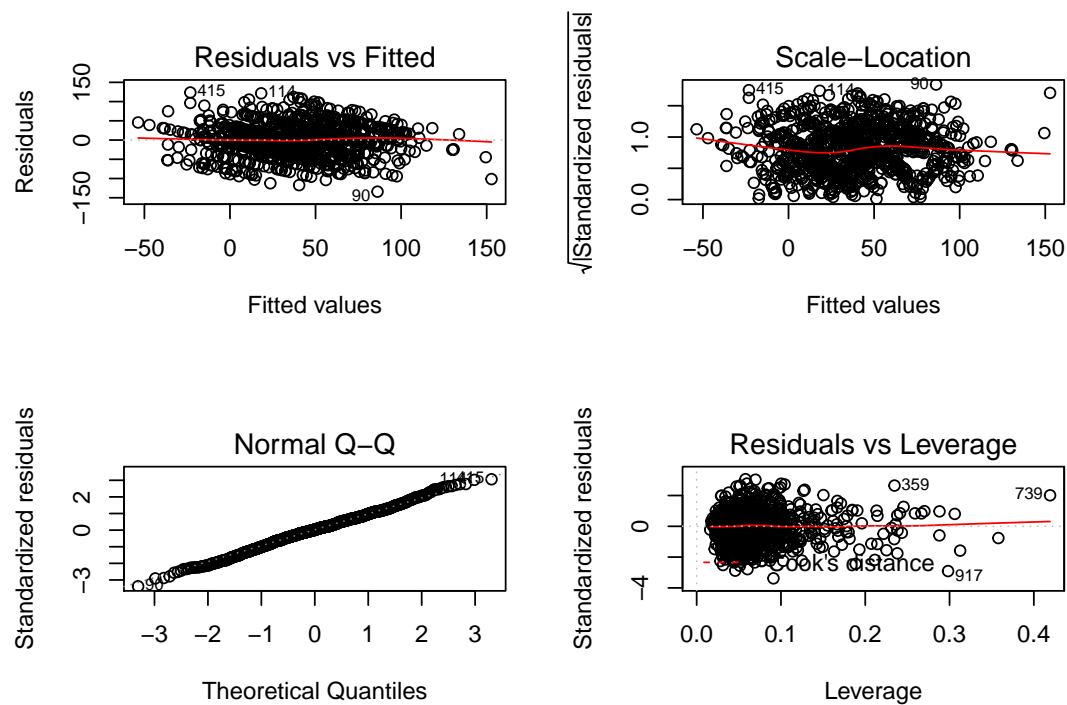
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following objects are masked from 'package:base':
##
##      crossprod, tcrossprod
## Loading required package: foreach
## Loaded glmnet 2.0-5
setwd("C:/Users/98302/Desktop/hw5")
read.csv("default_plus_chromatic_features_1059_tracks.txt", header=F) -> data
latitude = as.matrix(data[, dim(data)[2]-1])
longitude = as.matrix(data[, dim(data)[2]])
data = as.matrix(data[, -c(dim(data)[2]-1, dim(data)[2])])
latitude_lm = lm(latitude~data)
mat<-matrix(1:4, 2, 2)
layout(mat)
plot(latitude_lm)
```



```
latitude_r2 = summary(latitude_lm)$adj.r.squared
latitude_r2
```

```
## [1] 0.2411685
```

```
longitude_lm = lm(longitude~data)
layout(mat)
plot(longitude_lm)
```



```
longitude_r2 = summary(longitude_lm)$adj.r.squared
longitude_r2
```

```
## [1] 0.3181766
```

## 1.2

```
R2 = NULL
mse = NULL
model = NULL
summary(latitude)
```

```
##           V1
##  Min.      :-35.30
##  1st Qu.: 14.66
##  Median : 33.66
##  Mean   : 26.65
##  3rd Qu.: 39.91
##  Max.    : 54.68
```

```
summary(longitude)
```

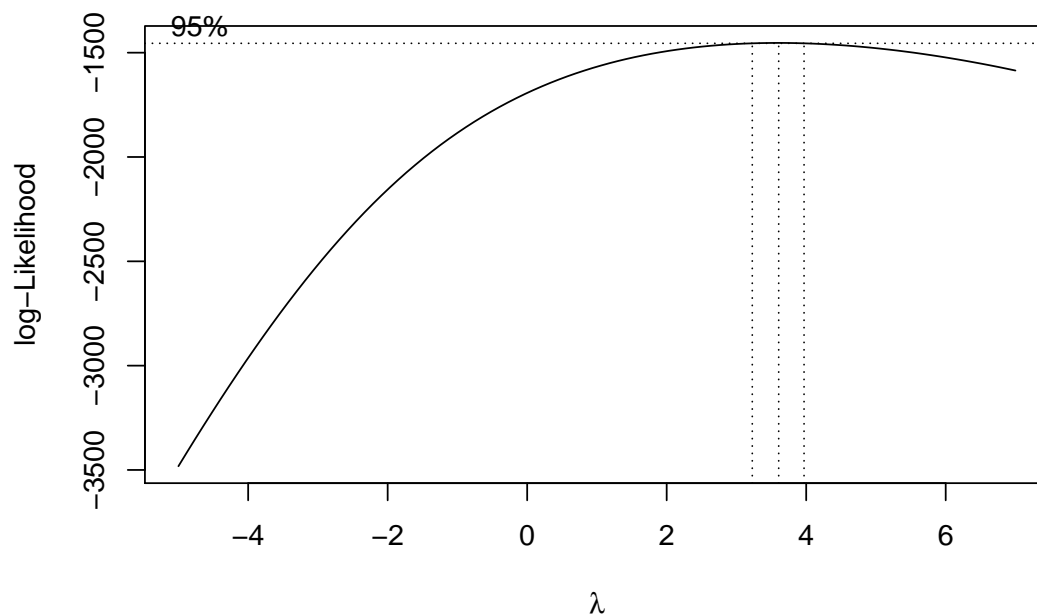
```
##           V1
##  Min.      :-88.76
##  1st Qu.:  3.21
##  Median : 32.83
##  Mean   : 38.41
##  3rd Qu.: 74.60
```

```
## Max. :149.12
latitude_orignal <- latitude
longitude_orignal <- longitude

#negative = which(latitude<0)
#latitude[negative] = latitude[negative] + 90
#negative = which(longitude<0)
#longitude[negative] = longitude[negative] + 180
#latitude_new <- latitude
#longitude_new <- longitude

latitude_new <- latitude + 90
longitude_new <- longitude + 180

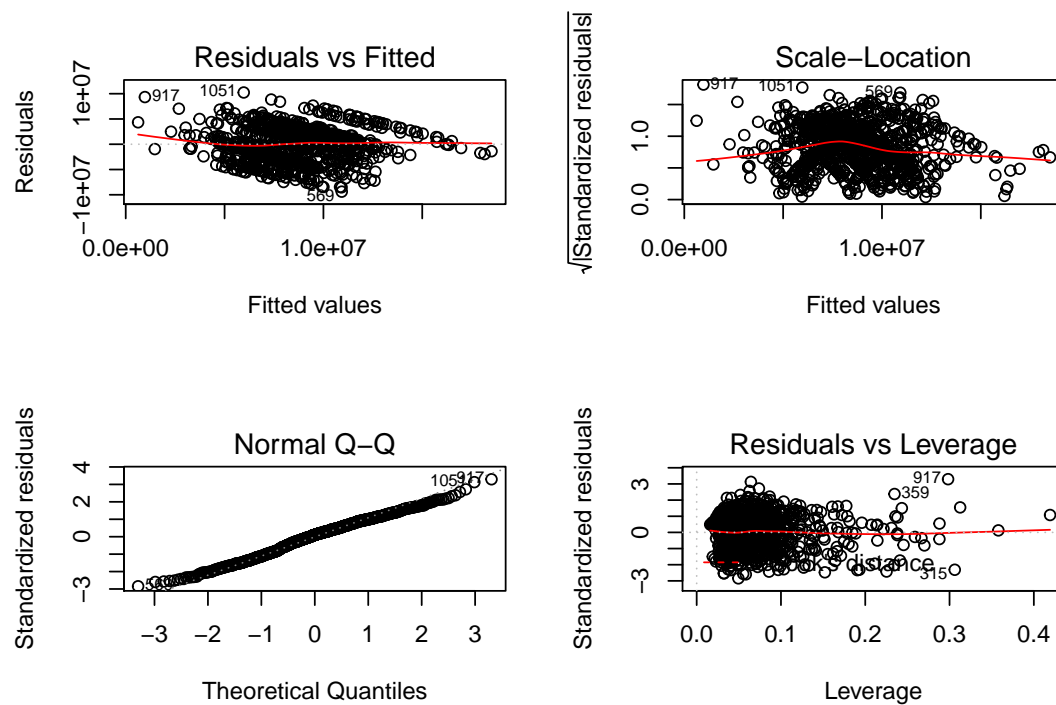
#boxcox for new
layout(1)
la_boxcox_new = boxcox(lm(latitude_new~data),lambda = seq(-5, 7, length = 100))
```



```
la_lambda_new = la_boxcox_new$x[which.max(la_boxcox_new$y)]
latitude_new_trans=(latitude_new~la_lambda_new-1)/la_lambda_new
latitude_new_trans.lm = lm(latitude_new_trans~data)
latitude_new_trans.r2 = summary(latitude_new_trans.lm)$adj.r.squared
latitude_new_trans.r2
```

```
## [1] 0.2782052
```

```
layout(mat)
plot(latitude_new_trans.lm)
```



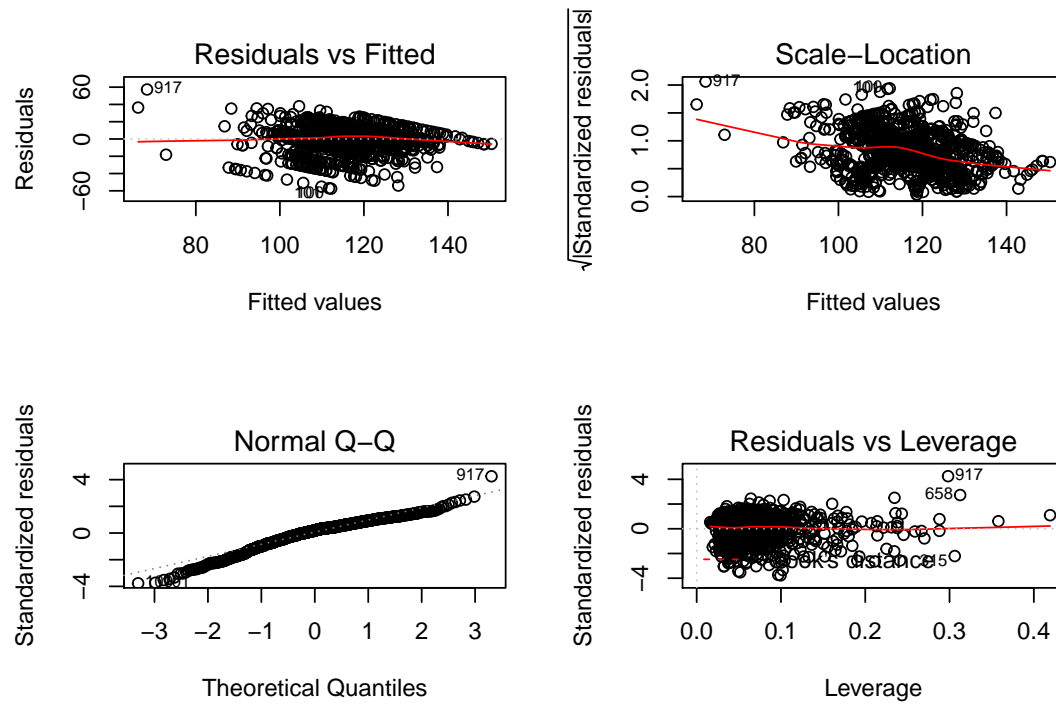
```
latitude_new_trans.mse = mean((latitude_new_trans.lm$fitted.values - latitude_new_trans)^2)
latitude_new_trans.mse
```

```
## [1] 1.06593e+13
```

```
latitude_new_without_trans.lm = lm(latitude_new~data)
latitude_new_without_trans.r2 = summary(latitude_new_without_trans.lm)$adj.r.squared
latitude_new_without_trans.r2
```

```
## [1] 0.2411685
```

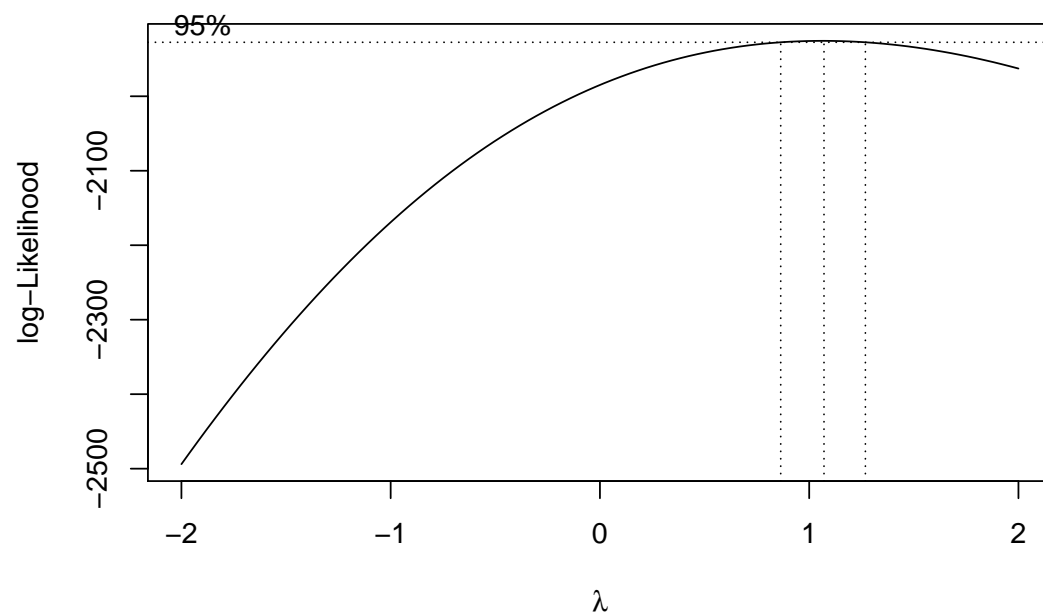
```
plot(latitude_new_without_trans.lm)
```



```
latitude_new_without_trans.pred = (latitude_new_without_trans.lm$fitted.values^
                                   la_lambda_new-1)/la_lambda_new
latitude_new_without_trans.mse = mean((latitude_new_without_trans.pred - latitude_new_trans)^2)
latitude_new_without_trans.mse

## [1] 1.123119e+13

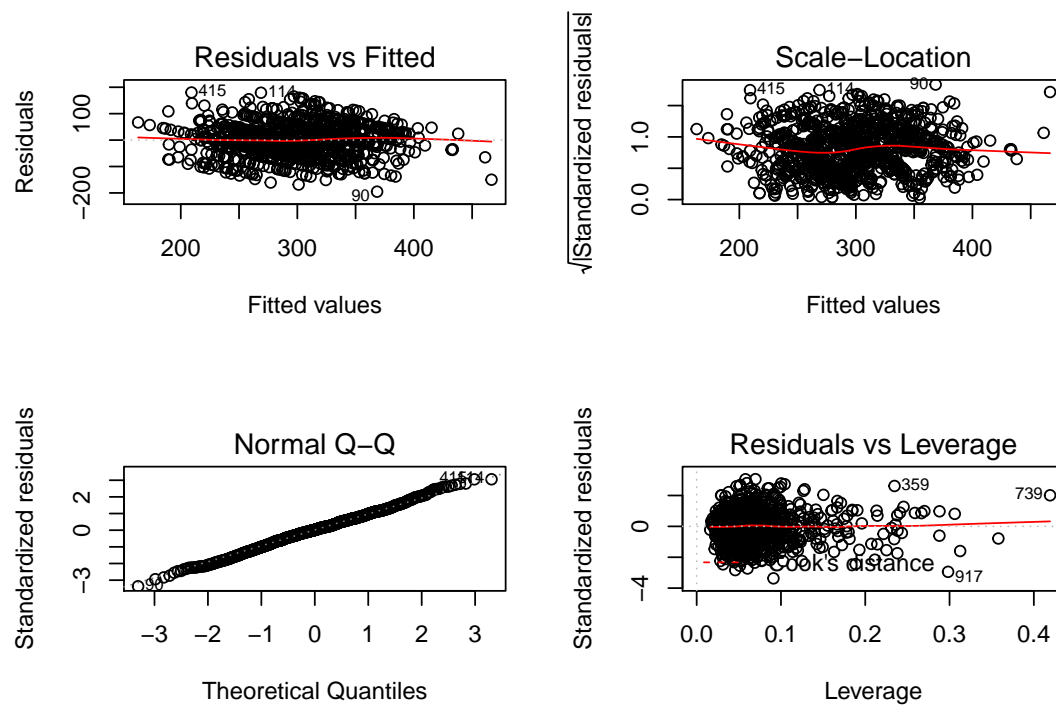
layout(1)
lo_boxcox_new = boxcox(lm(longitude_new~data))
```



```
lo_lambda_new = lo_boxcox_new$x[which.max(lo_boxcox_new$y)]
longitude_new_trans=(longitude_new^lo_lambda_new-1)/lo_lambda_new
longitude_new_trans.lm = lm(longitude_new_trans~data)
longitude_new_trans.r2 = summary(longitude_new_trans.lm)$adj.r.squared
longitude_new_trans.r2
```

```
## [1] 0.3187052
```

```
layout(mat)
plot(longitude_new_trans.lm)
```



```
longitude_new_trans.mse = mean((longitude_new_trans.lm$fitted.values - longitude_new_trans)^2)
longitude_new_trans.mse
```

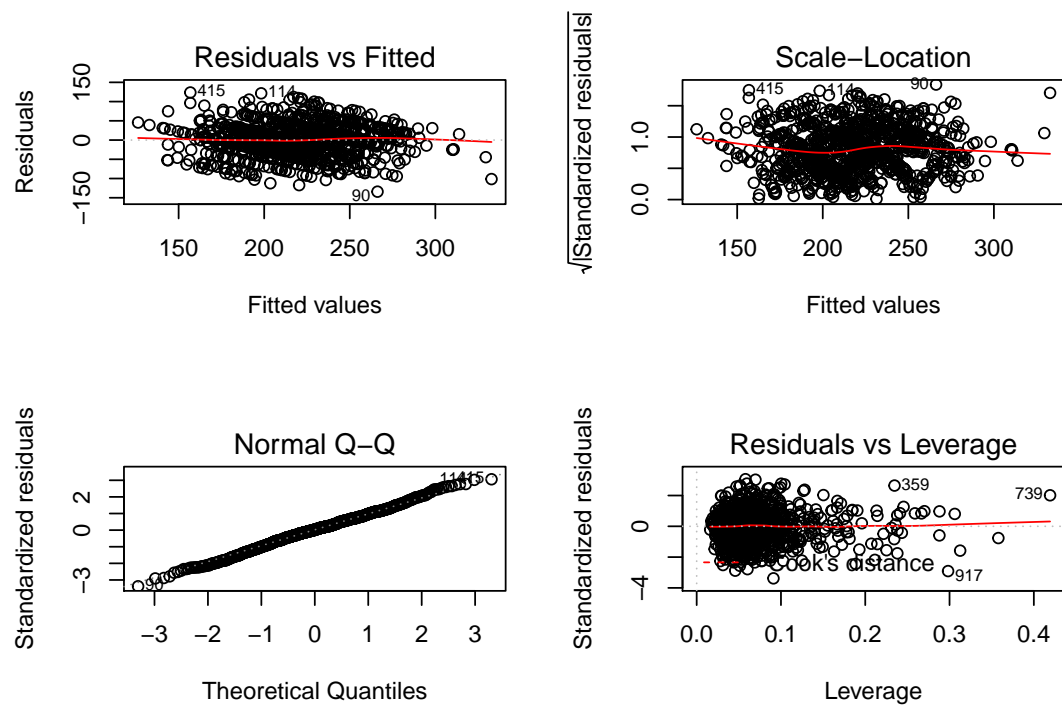
```
## [1] 3441.23
```

```
longitude_new_without_trans.lm = lm(longitude_new~data)
longitude_new_without_trans.r2 = summary(longitude_new_without_trans.lm)$adj.r.squared
longitude_new_without_trans.r2
```

```
## [1] 0.3181766
```

```
plot(longitude_new_without_trans.lm)
```





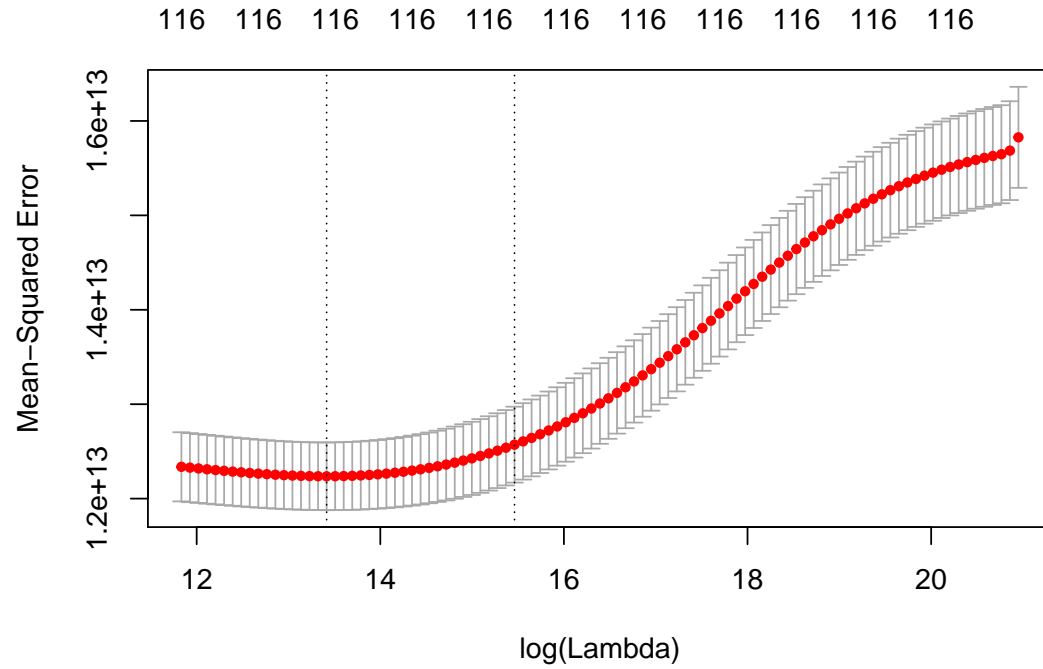
```
longitude_new_without_trans.pred = (longitude_new_without_trans.lm$fitted.values^
                                   lo_lambda_new-1)/lo_lambda_new
longitude_new_without_trans.mse = mean((longitude_new_without_trans.pred - longitude_new_trans)^2)
longitude_new_without_trans.mse

## [1] 3442.134
```

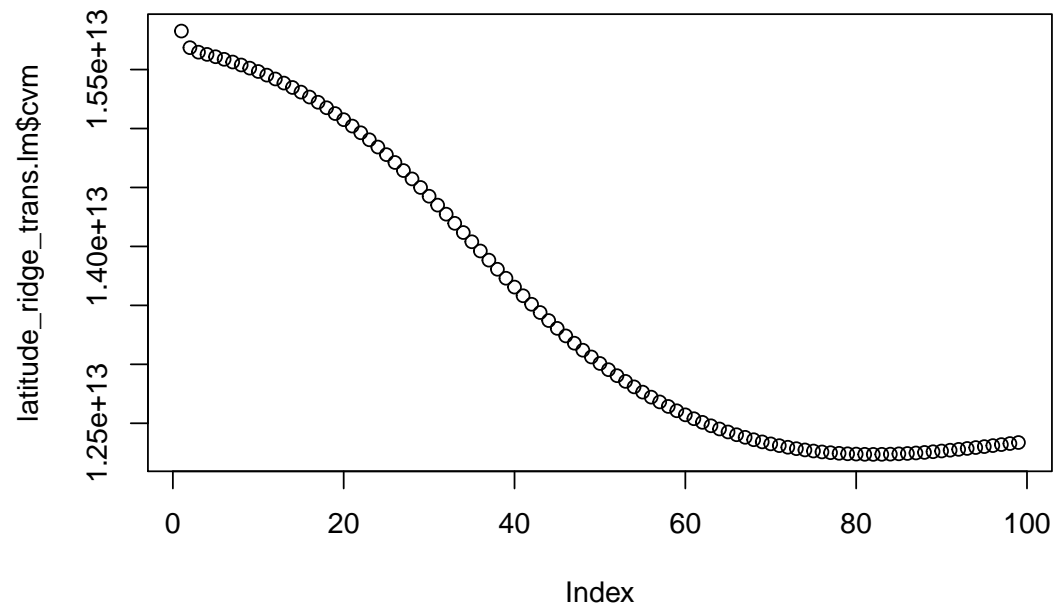
## 1.3a ridge

### latitude

```
layout(1)
latitude_ridge_trans.lm = cv.glmnet(x=data,y=latitude_new_trans,alpha=0,
                                   nfold = 10,family = "gaussian")
plot(latitude_ridge_trans.lm)
```



```
plot(latitude_ridge_trans.lm$cvm)
```



```
latitude_ridge_trans.pred <- predict(latitude_ridge_trans.lm, s =  
  latitude_ridge_trans.lm$lambda.min, newx = data)
```

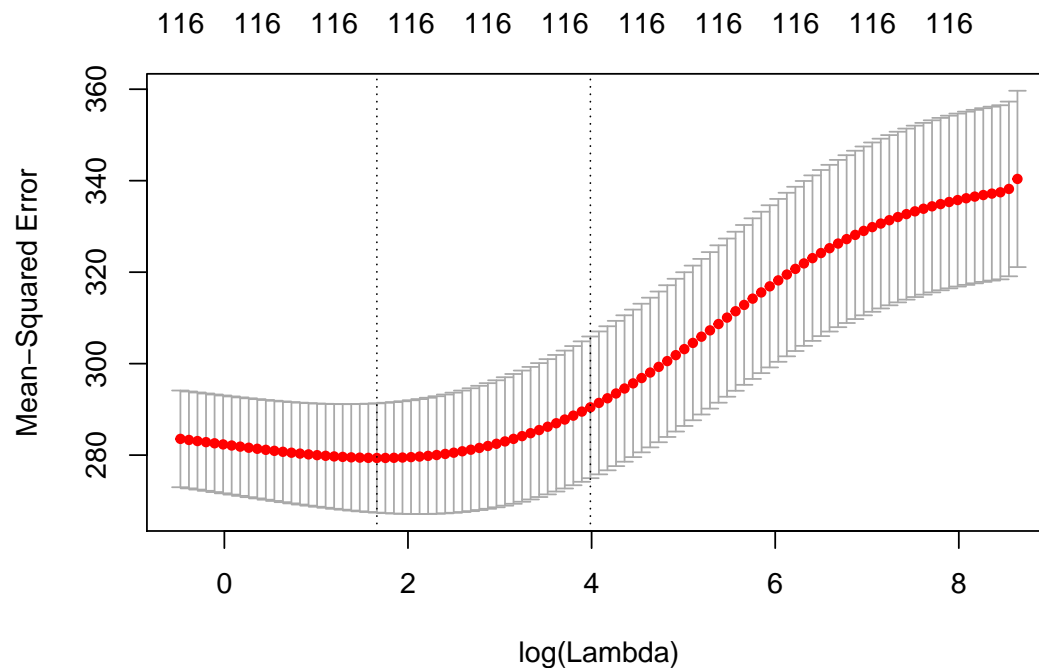
```
#latitude_ridge_trans.pred.original = (latitude_ridge.pred * la_lambda_new + 1)^(1/la_lambda_new)
latitude_ridge_trans.R2 = var(latitude_ridge_trans.pred)/var(latitude_new_trans)
latitude_ridge_trans.R2
```

```
##          1
## 1 0.2409245
```

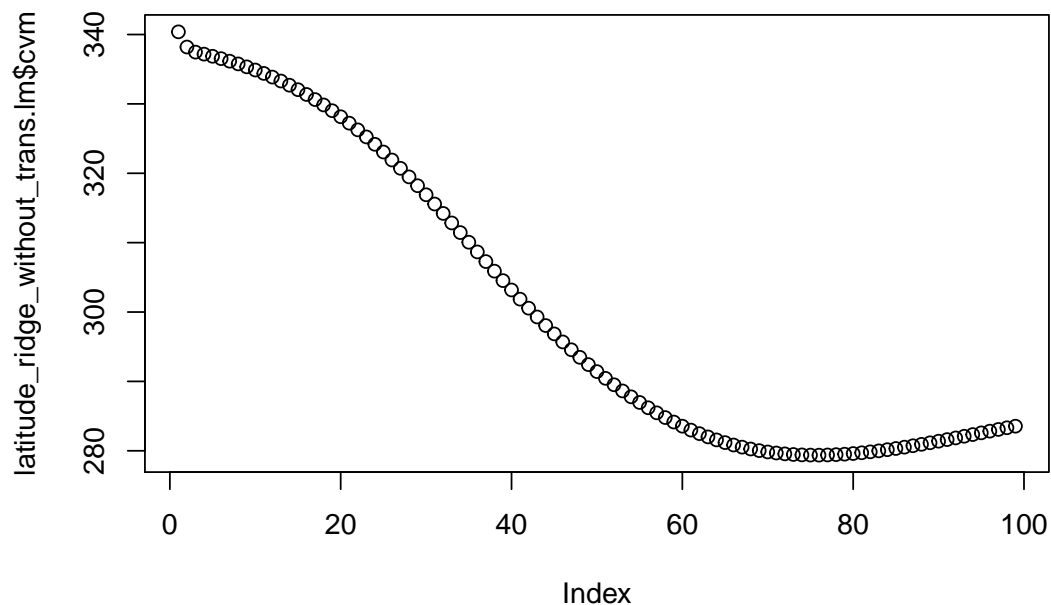
```
latitude_ridge_trans.mse = mean((latitude_ridge_trans.pred - latitude_new_trans)^2)
latitude_ridge_trans.mse
```

```
## [1] 1.1095e+13
```

```
latitude_ridge_without_trans.lm = cv.glmnet(x=data,y=latitude_new,alpha=0,
                                             nfold = 10,family = "gaussian")
plot(latitude_ridge_without_trans.lm )
```



```
plot(latitude_ridge_without_trans.lm$cvm)
```

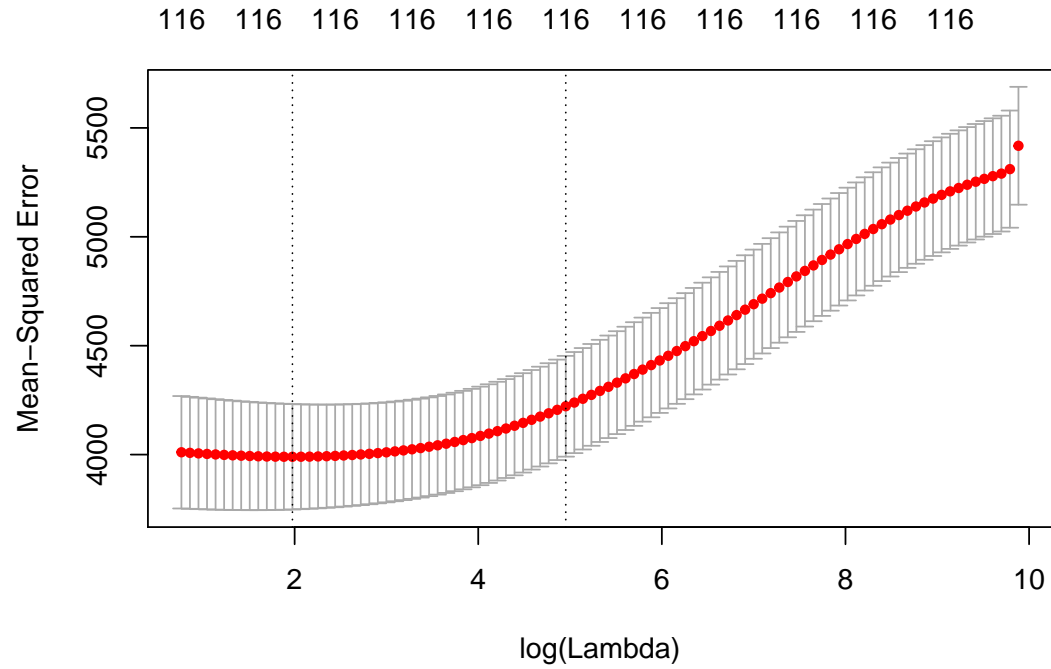


```
latitude_ridge_without_trans.pred <- predict(latitude_ridge_without_trans.lm,
      s = latitude_ridge_without_trans.lm$lambda.min,
      newx = data)
#latitude_ridge.pred.original = (latitude_ridge.pred * la_lambda_new + 1)^(1/la_lambda_new)
latitude_ridge_without_trans.pred <- (latitude_ridge_without_trans.pred
      ^la_lambda_new-1)/la_lambda_new
latitude_ridge_without_trans.R2 = var(latitude_ridge_without_trans.pred)/var(latitude_new_trans)
latitude_ridge_without_trans.R2

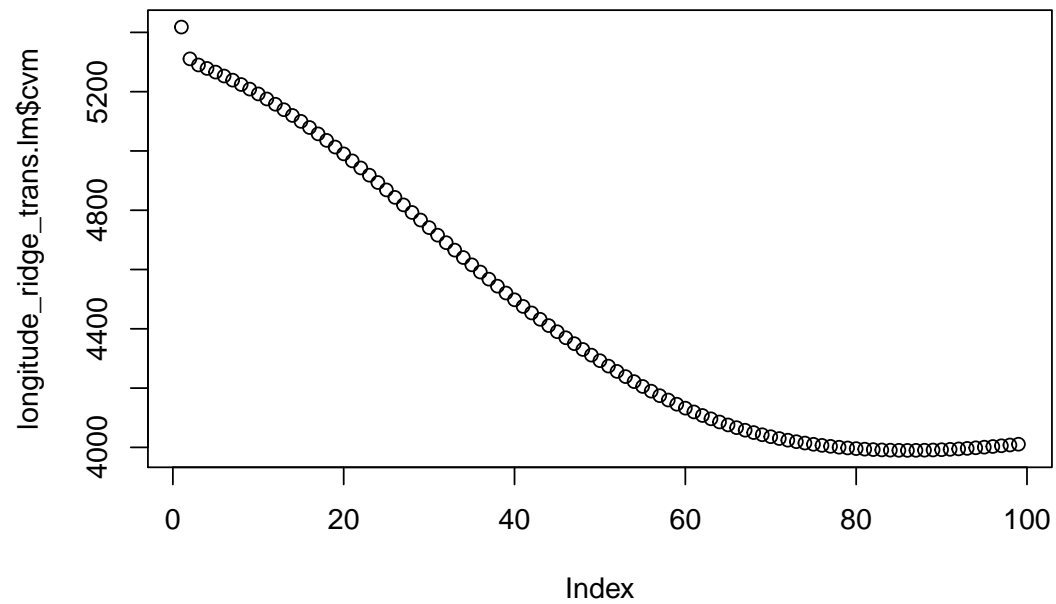
##          1
## 1 0.2462596

latitude_ridge_without_trans.mse = mean((latitude_ridge_without_trans.pred -
      latitude_new_trans)^2)
latitude_ridge_without_trans.mse

## [1] 1.173027e+13
#longitude
longitude_ridge_trans.lm = cv.glmnet(x=data,y=longitude_new_trans,alpha=0,
      nfold = 10,family = "gaussian")
plot(longitude_ridge_trans.lm)
```



```
plot(longitude_ridge_trans.lm$cvm)
```



```
longitude_ridge_trans.pred <- predict(longitude_ridge_trans.lm, s =  
  longitude_ridge_trans.lm$lambda.min,
```

```

newx = data)
longitude_ridge_trans.R2 = var(longitude_ridge_trans.pred)/var(longitude_new_trans)
longitude_ridge_trans.R2

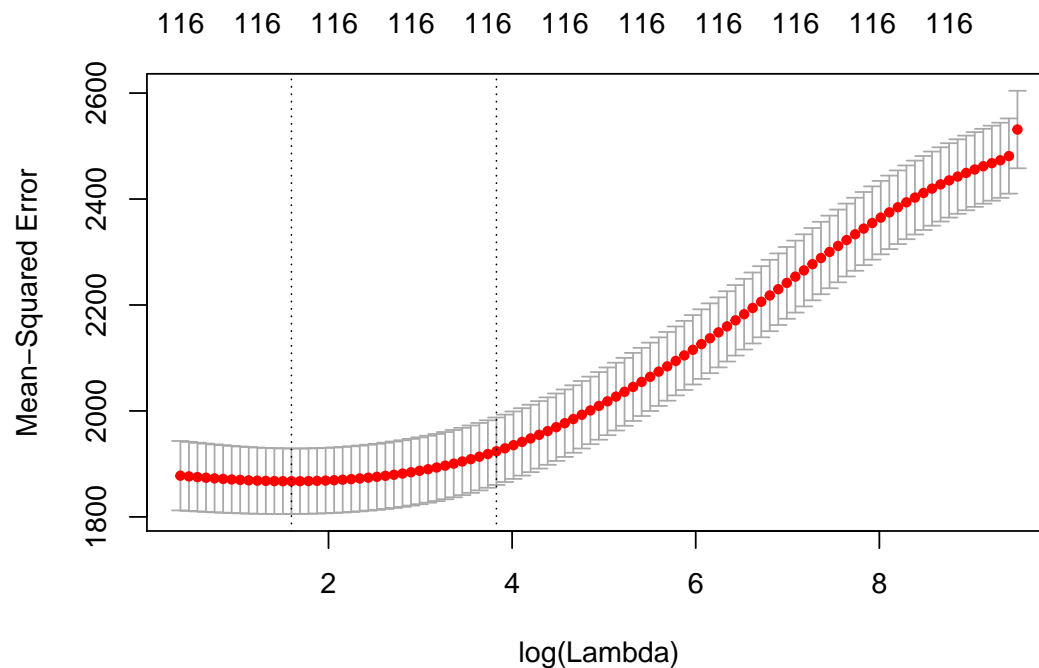
##           1
## 1 0.2930358

longitude_ridge_trans.mse = mean((longitude_ridge_trans.pred - longitude_new_trans)^2)
longitude_ridge_trans.mse

## [1] 3533.256

longitude_ridge_without_trans.lm = cv.glmnet(x=data,y=longitude_new,alpha=0,
                                             nfold = 10,family = "gaussian")
plot(longitude_ridge_without_trans.lm )

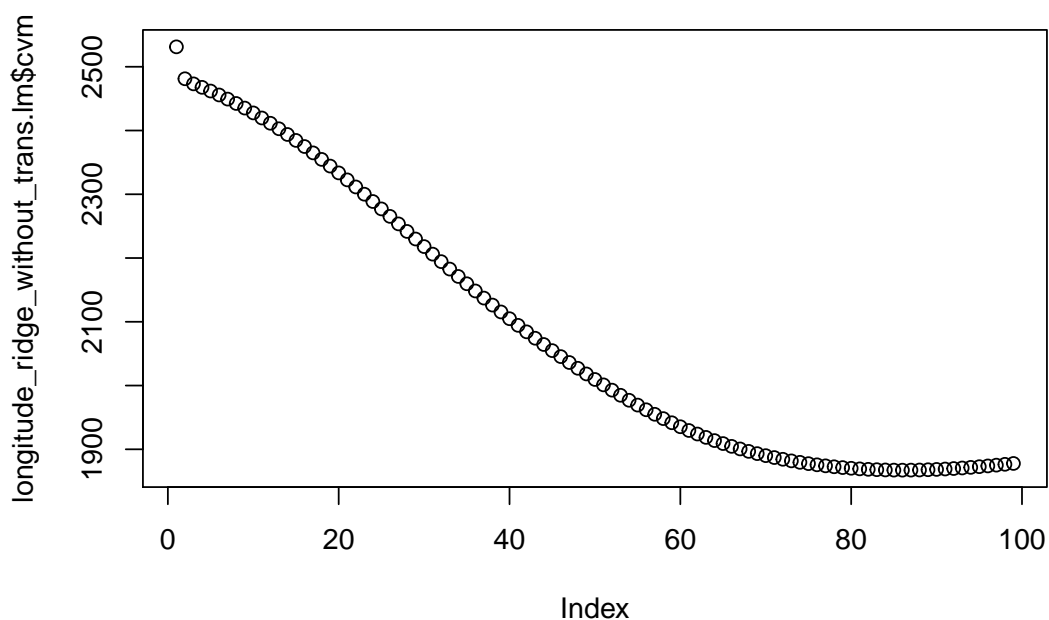
```



```

plot(longitude_ridge_without_trans.lm$cvm)

```



```
longitude_ridge_without_trans.pred <- predict(longitude_ridge_without_trans.lm,
      s = longitude_ridge_without_trans.lm$lambda.min,
      newx = data)
longitude_ridge_without_trans.pred <- (longitude_ridge_without_trans.pred^
      lo_lambda_new-1)/lo_lambda_new
longitude_ridge_without_trans.R2 = var(longitude_ridge_without_trans.pred)/var(longitude_new_trans)
longitude_ridge_without_trans.R2
```

```
##          1
## 1 0.2936222
```

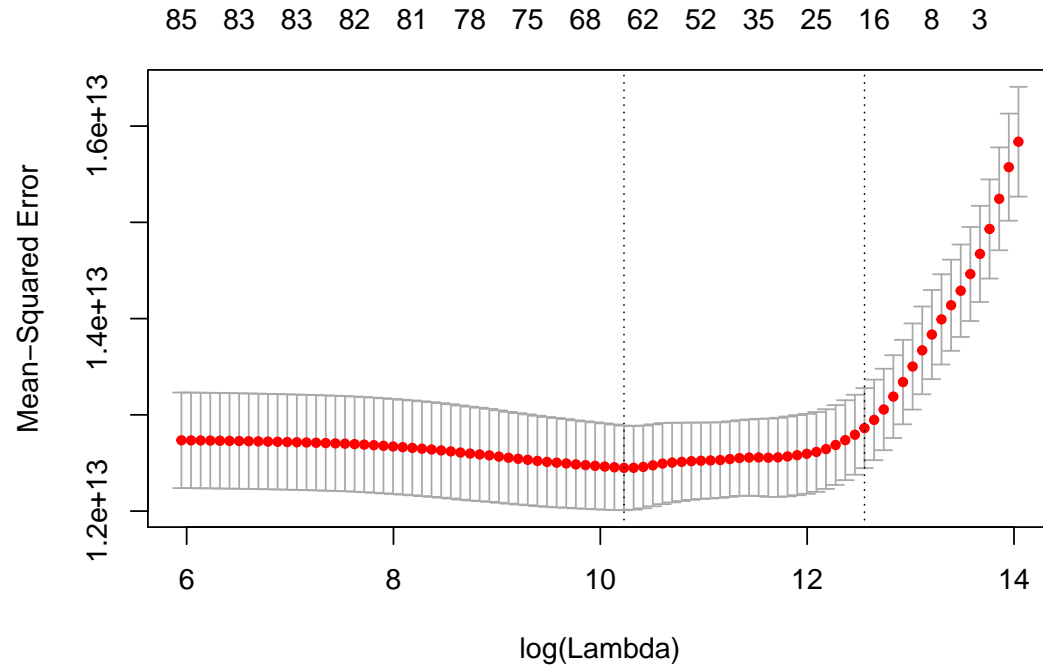
```
longitude_ridge_without_trans.mse = mean((longitude_ridge_without_trans.pred -
      longitude_new_trans)^2)
longitude_ridge_without_trans.mse
```

```
## [1] 3532.671
```

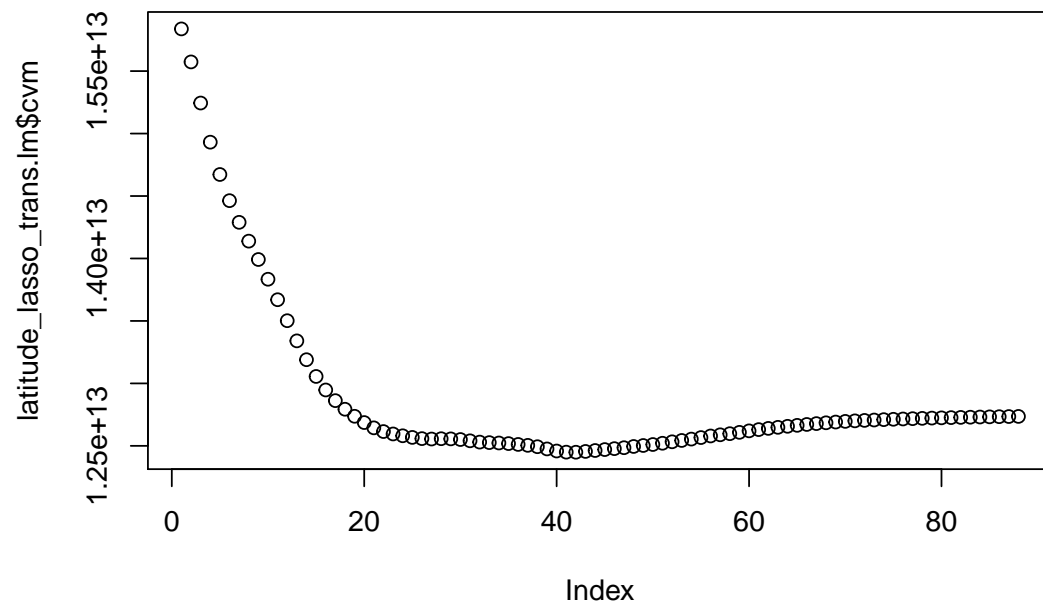
## 1.3b lasso

### latitude

```
layout(1)
latitude_lasso_trans.lm = cv.glmnet(x=data,y=latitude_new_trans,alpha=1,
      nfold = 10,family = "gaussian")
plot(latitude_lasso_trans.lm)
```



```
plot(latitude_lasso_trans.lm$cvm)
```



```
latitude_lasso_trans.pred <- predict(latitude_lasso_trans.lm, s =  
  latitude_lasso_trans.lm$lambda.min, newx = data)
```



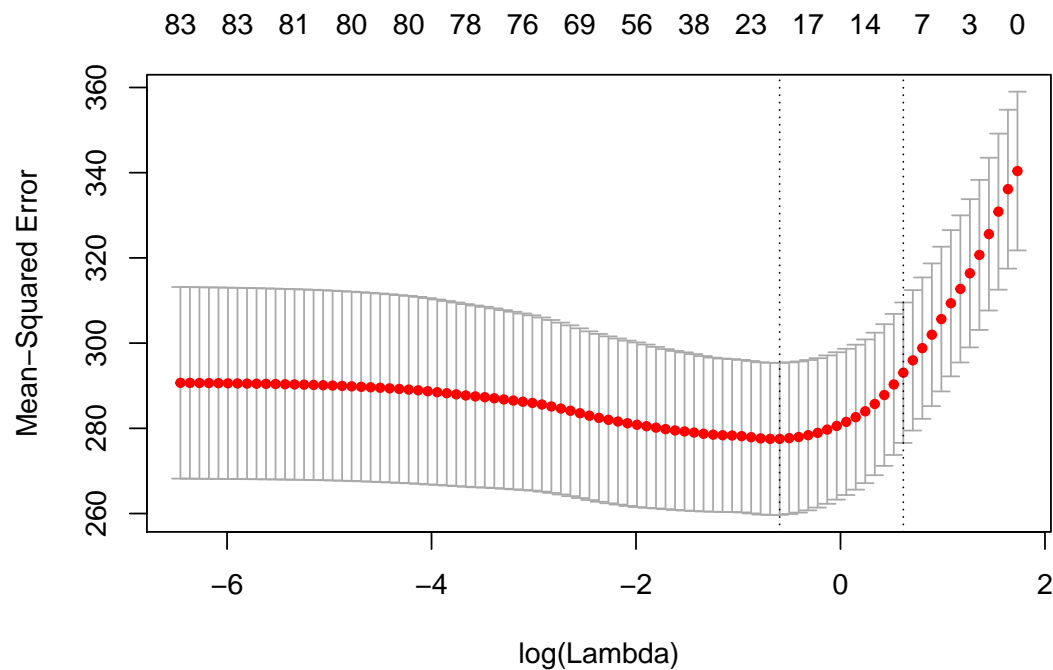
```
#latitude_lasso_trans.pred.original = (latitude_lasso_trans.pred * la_lambda_new + 1)^(1/la_lambda_new)
latitude_lasso_trans.R2 = var(latitude_lasso_trans.pred)/var(latitude_new_trans)
latitude_lasso_trans.R2
```

```
##          1
## 1 0.2640511
```

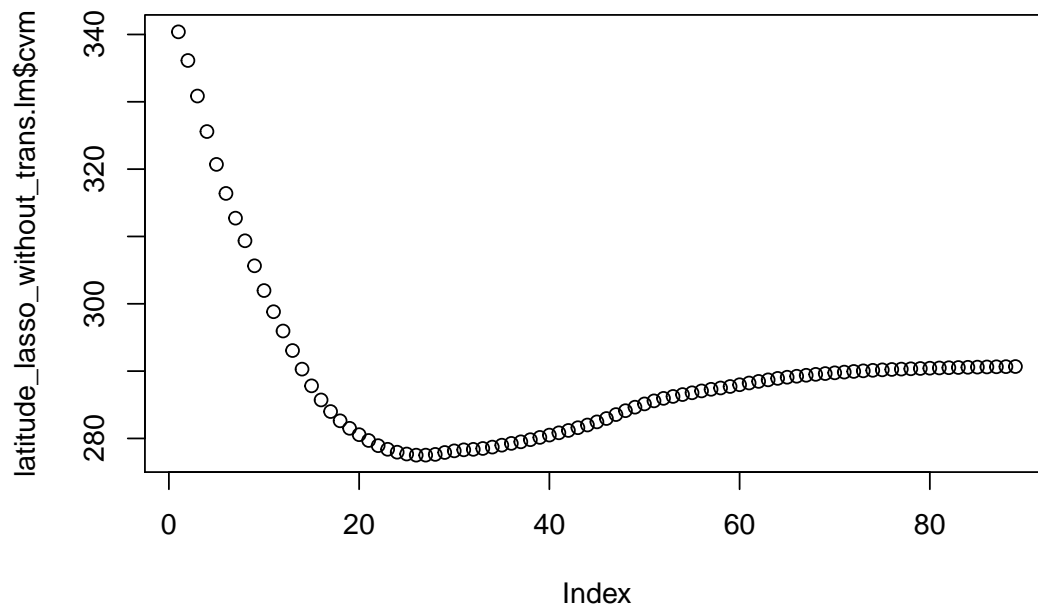
```
latitude_lasso_trans.mse = mean((latitude_lasso_trans.pred - latitude_new_trans)^2)
latitude_lasso_trans.mse
```

```
## [1] 1.098345e+13
```

```
latitude_lasso_without_trans.lm = cv.glmnet(x=data,y=latitude_new,alpha=1,
                                             nfold = 10,family = "gaussian")
plot(latitude_lasso_without_trans.lm )
```



```
plot(latitude_lasso_without_trans.lm$cvm)
```

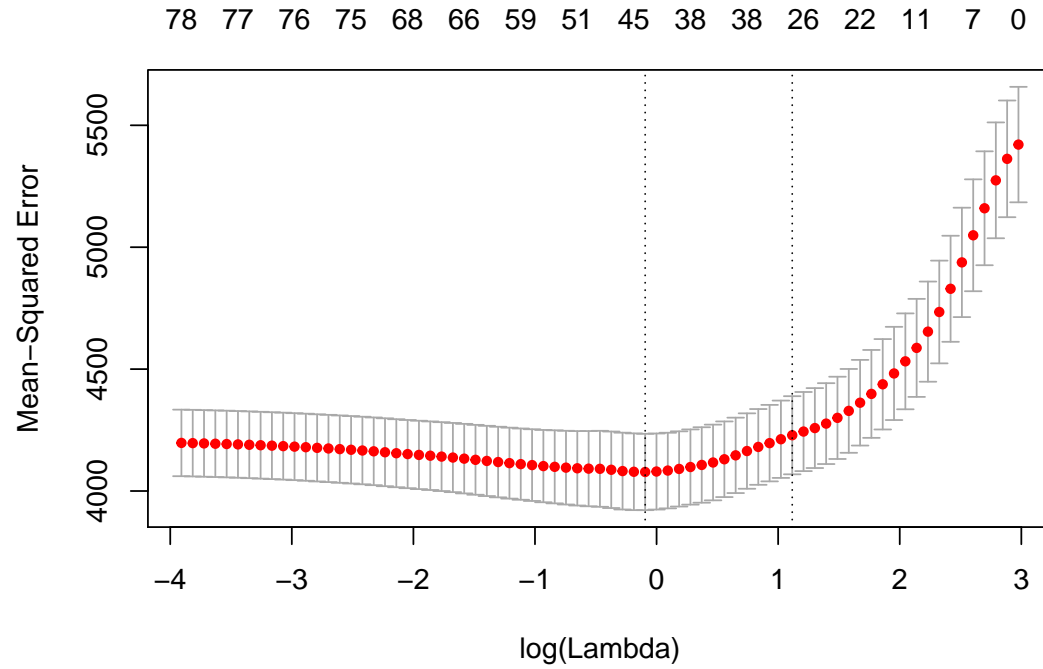


```
latitude_lasso_without_trans.pred <- predict(latitude_lasso_without_trans.lm,
      s = latitude_lasso_without_trans.lm$lambda.min,
      newx = data)
#latitude_lasso.pred.original = (latitude_lasso.pred * la_lambda_new + 1)^(1/la_lambda_new)
latitude_lasso_without_trans.pred <- (latitude_lasso_without_trans.pred^
      la_lambda_new-1)/la_lambda_new
latitude_lasso_without_trans.R2 = var(latitude_lasso_without_trans.pred)/var(latitude_new_trans)
latitude_lasso_without_trans.R2

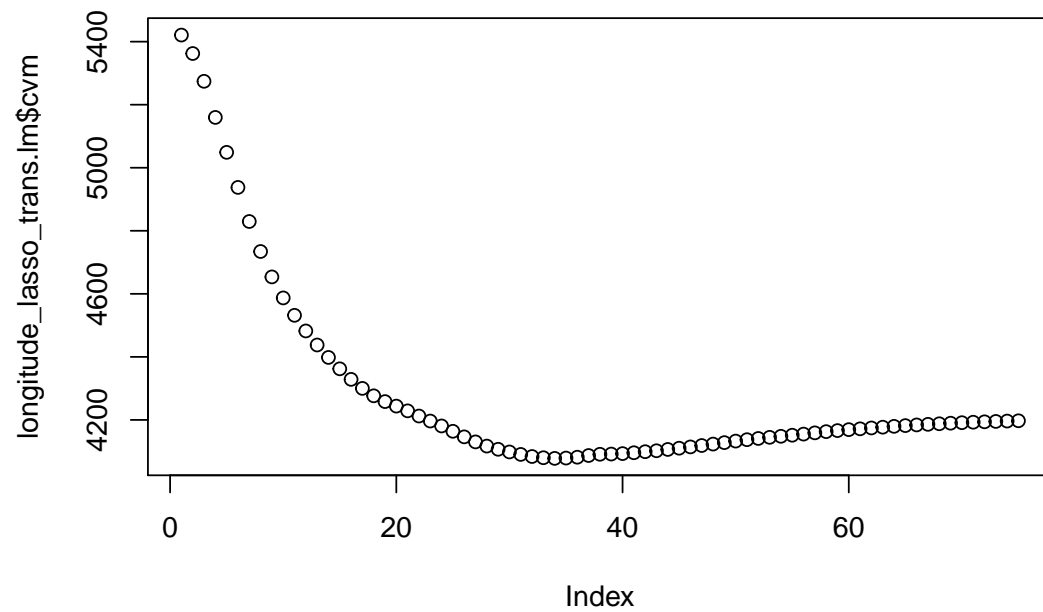
##          1
## 1 0.2016096

latitude_lasso_without_trans.mse = mean((latitude_lasso_without_trans.pred
      - latitude_new_trans)^2)
latitude_lasso_without_trans.mse

## [1] 1.240553e+13
#longitude
longitude_lasso_trans.lm = cv.glmnet(x=data,y=longitude_new_trans,alpha=1,
      nfold = 10,family = "gaussian")
plot(longitude_lasso_trans.lm)
```



```
plot(longitude_lasso_trans.lm$cvm)
```



```
longitude_lasso_trans.pred <- predict(longitude_lasso_trans.lm, s =  
  longitude_lasso_trans.lm$lambda.min, newx = data)
```

```

longitude_lasso_trans.R2 = var(longitude_lasso_trans.pred)/var(longitude_new_trans)
longitude_lasso_trans.R2

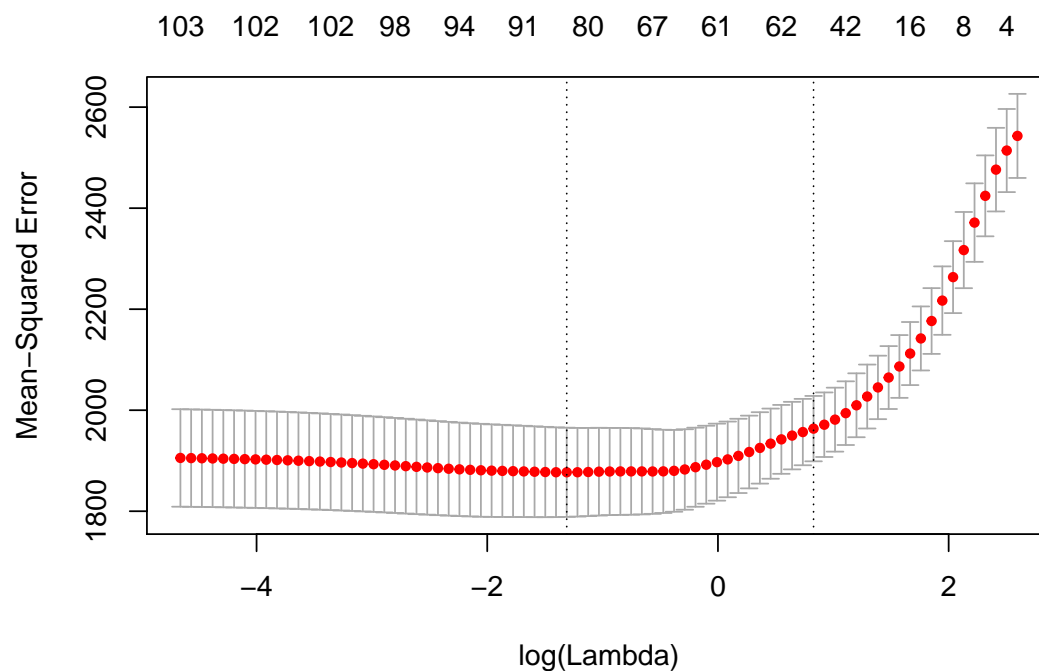
##           1
## 1 0.2708683

longitude_lasso_trans.mse = mean((longitude_lasso_trans.pred - longitude_new_trans)^2)
longitude_lasso_trans.mse

## [1] 3618.02

longitude_lasso_without_trans.lm = cv.glmnet(x=data,y=longitude_new,alpha=1,
                                              nfold = 10,family = "gaussian")
plot(longitude_lasso_without_trans.lm )

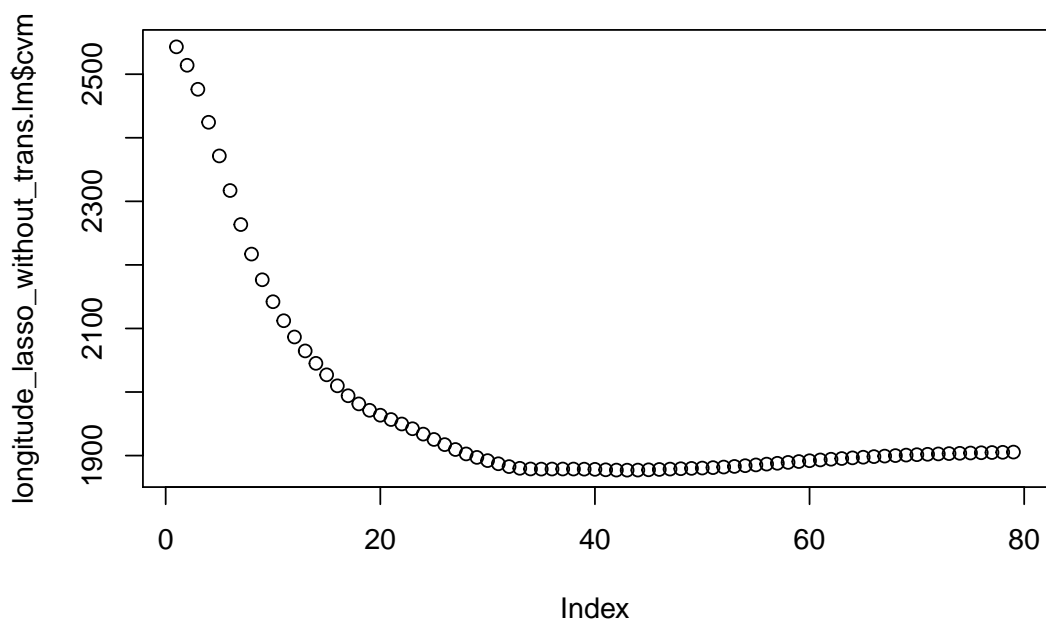
```



```

plot(longitude_lasso_without_trans.lm$cvm)

```



```
longitude_lasso_without_trans.pred <- predict(longitude_lasso_without_trans.lm,
      s = longitude_lasso_without_trans.lm$lambda.min,
      newx = data)
longitude_lasso_without_trans.pred <- (longitude_lasso_without_trans.pred^
      lo_lambda_new-1)/lo_lambda_new
longitude_lasso_without_trans.R2 = var(longitude_lasso_without_trans.pred)/var(longitude_new_trans)
longitude_lasso_without_trans.R2

##          1
## 1 0.3147185

longitude_lasso_without_trans.mse = mean((longitude_lasso_without_trans.pred -
      longitude_new_trans)^2)
longitude_lasso_without_trans.mse

## [1] 3493.799

model = c("latitude_new_trans.lm","latitude_new_without_trans.lm",
      "latitude_ridge_trans.lm","latitude_ridge_without_trans.lm",
      "latitude_lasso_trans.lm","latitude_lasso_without_trans.lm",
      "longitude_new_trans.lm","longitude_new_without_trans.lm",
      "longitude_ridge_trans.lm","longitude_ridge_without_trans.lm",
      "longitude_lasso_trans.lm","longitude_lasso_without_trans.lm"
      )
R2 = c(latitude_new_trans.r2,latitude_new_without_trans.r2,
      latitude_ridge_trans.R2,latitude_ridge_without_trans.R2,
      latitude_lasso_trans.R2,latitude_lasso_without_trans.R2,
      longitude_new_trans.r2,longitude_new_without_trans.r2,
      longitude_ridge_trans.R2,longitude_ridge_without_trans.R2,
      longitude_lasso_trans.R2,longitude_lasso_without_trans.R2
```

```
)
mse = c(latitude_new_trans.mse,latitude_new_without_trans.mse,
        latitude_ridge_trans.mse,latitude_ridge_without_trans.mse,
        latitude_lasso_trans.mse,latitude_lasso_without_trans.mse,
        longitude_new_trans.mse,longitude_new_without_trans.mse,
        longitude_ridge_trans.mse,longitude_ridge_without_trans.mse,
        longitude_lasso_trans.mse,longitude_lasso_without_trans.mse
)
cbind(model,R2,mse)
```

```
##      model                                R2
## [1,] "latitude_new_trans.lm"              "0.278205162234764"
## [2,] "latitude_new_without_trans.lm"      "0.24116849301382"
## [3,] "latitude_ridge_trans.lm"           "0.2409244464397298"
## [4,] "latitude_ridge_without_trans.lm"    "0.246259572258212"
## [5,] "latitude_lasso_trans.lm"           "0.264051079585654"
## [6,] "latitude_lasso_without_trans.lm"    "0.201609622712887"
## [7,] "longitude_new_trans.lm"            "0.318705154418536"
## [8,] "longitude_new_without_trans.lm"     "0.318176624480052"
## [9,] "longitude_ridge_trans.lm"          "0.293035773421009"
## [10,] "longitude_ridge_without_trans.lm"  "0.293622207449689"
## [11,] "longitude_lasso_trans.lm"         "0.270868290616546"
## [12,] "longitude_lasso_without_trans.lm"  "0.314718521104191"
##      mse
## [1,] "10659299446201.1"
## [2,] "11231193445874.9"
## [3,] "11095000365922.8"
## [4,] "11730270427932.2"
## [5,] "10983446421506"
## [6,] "12405533949603"
## [7,] "3441.22950864091"
## [8,] "3442.1343335657"
## [9,] "3533.25648606586"
## [10,] "3532.6713372049"
## [11,] "3618.0203618532"
## [12,] "3493.79880603119"
```

## 2

```
read.table("default_of_credit_card_clients.txt",header=T)->data
data[1,]
```

```
##  ID LIMIT_BAL SEX EDUCATION MARRIAGE AGE PAY_0 PAY_2 PAY_3 PAY_4 PAY_5
## 1   1    20000  2         2         1  24     2     2    -1    -1    -2
##  PAY_6 BILL_AMT1 BILL_AMT2 BILL_AMT3 BILL_AMT4 BILL_AMT5 BILL_AMT6
## 1    -2     3913     3102     689         0         0         0
##  PAY_AMT1 PAY_AMT2 PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6 default
## 1         0     689         0         0         0         0         1
```

```
dim(data)
```

```
## [1] 30000    25
```

```

x = as.matrix(data[,2:24])
y = data[25]
y <- as.factor(as.matrix(y))
alpha = c(0,0.25,0.5,0.75,1)
acc = NULL
for (i in 1:length(alpha)){
  lm = cv.glmnet(x, y, family = "binomial", alpha = alpha[i], type.measure = "class",
                nfold = 10)
  #plot(lm,main=paste("alpha = ",alpha[i]))
  lm.pred <- predict(lm, x, type = "class", s = "lambda.min")
  lm.acc = sum(lm.pred == y)/length(y)
  acc[i] = lm.acc
}
acc

```

```
## [1] 0.8069667 0.8099667 0.8105667 0.8104000 0.8103000
```