Lab3

GroupB:Kyla Hayworth, Lyn Peterson, Yilin Li

2019/9/24

```r
d <- read.csv("http://andrewpbray.github.io/data/crime-train.csv")

group_B_fit <- function(training_data) {
  # Select out variables to be fitted in the model
  training_data = d[1:600,c(8,49,54,127)]

  # Filter out the observations with missing data
  filtered_data = training_data[rowSums(!is.na(training_data)) == 4,]
  filtered_data = mutate(filtered_data,NumIllegsr = sqrt(NumIlleg))

  # run lm() to fit your model.
  m1 = lm(ViolentCrimesPerPop ~ racePctWhite + PctKids2Par +NumIllegsr, data = filtered_data)

  # on the last line, simply put m1, your final model.
  # this will return it as output.
  m1
}

group_B_MSE <- function(model, data) {
  # when you run group_A_fit, save the output and put
  # it here as the "model" argument.

  data = mutate(data, NumIllegsr = sqrt(NumIlleg))
  # use your model to find the predicted values for
  # all of the observations in the data set.
  p = predict(model, data)
  true_values = data$ViolentCrimesPerPop
  # use those fitted values and the true values to
  # compute the MSE, which this function should return
  # as output.
  MSE = mean((p - true_values)^2)

  MSE
}
```