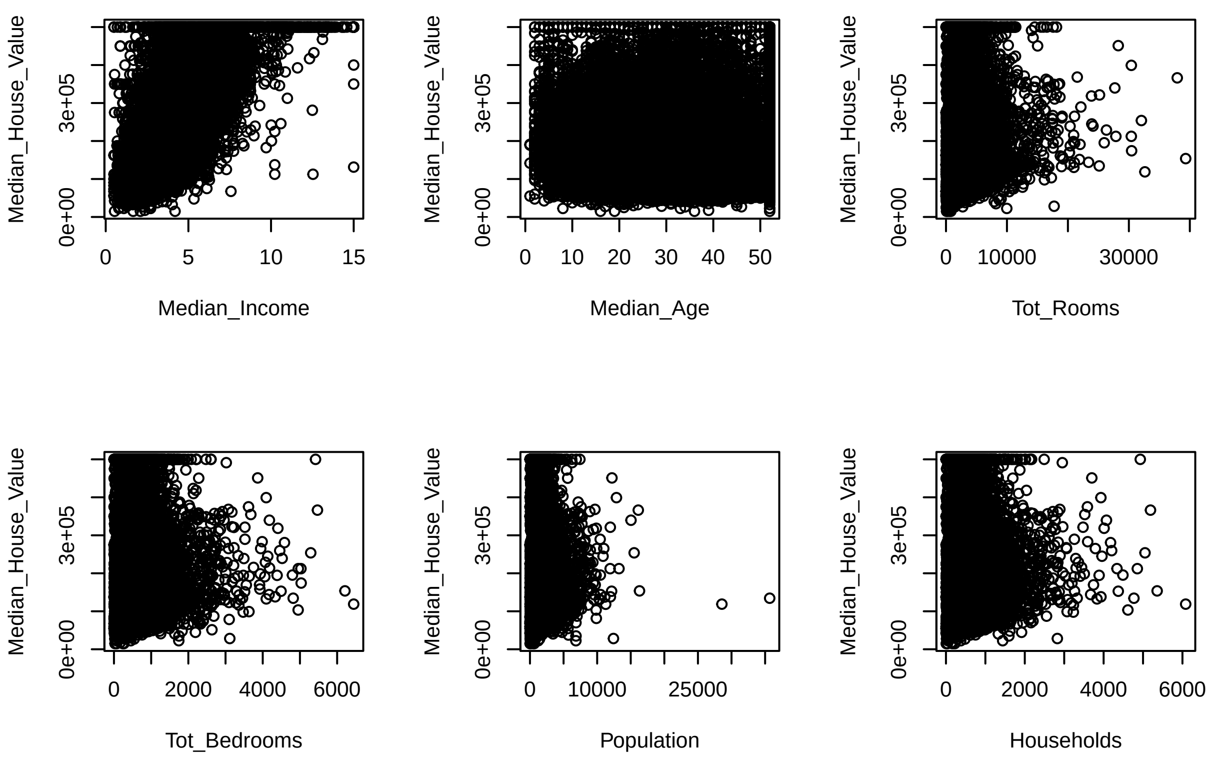
Project

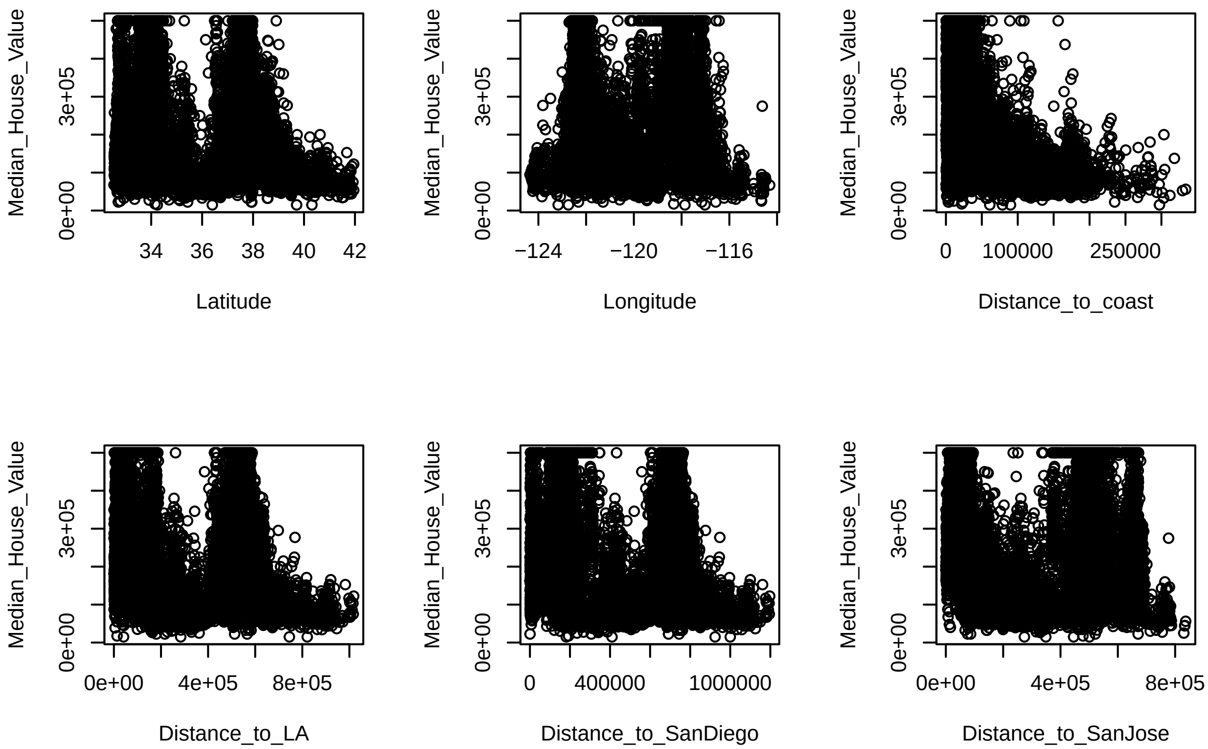
Manikanta Reddy Kallam

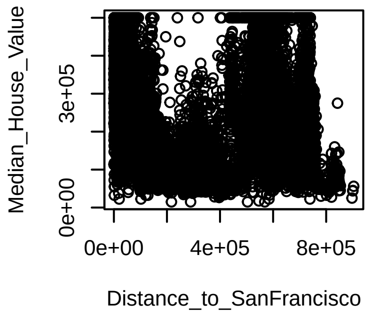
2023-05-05

Data Loading and Statistical analysis

```
data_df <- read.csv("./data/California_Houses.csv")
par(mfrow=c(2,3))
plot(Median_House_Value ~ .,data = data_df)
```




Median_House_Value
0e+00
3e+05
34
36
38
40
42
Latitude
−124
−120
−116
Longitude
0
100000
250000
Distance_to_coast
0e+00
4e+05
8e+05
Distance_to_LA
0
400000
1000000
Distance_to_SanDiego
Distance_to_SanJose



Statistical Ananlysis

```
# adding another column to show closest distance to major city
data_df <- cbind(data_df,c(sqldf("select min(Distance_to_LA, Distance_to_SanDiego, Distance_to_SanJose,

summary(data_df)
```

```
##  Median_House_Value Median_Income      Median_Age      Tot_Rooms
##  Min.   : 14999     Min.   : 0.4999   Min.   : 1.00   Min.   :    2
##  1st Qu.:119600     1st Qu.: 2.5634   1st Qu.:18.00   1st Qu.: 1448
##  Median :179700     Median : 3.5348   Median :29.00   Median : 2127
##  Mean   :206856     Mean   : 3.8707   Mean   :28.64   Mean   : 2636
##  3rd Qu.:264725     3rd Qu.: 4.7432   3rd Qu.:37.00   3rd Qu.: 3148
##  Max.   :500001     Max.   :15.0001   Max.   :52.00   Max.   :39320
##   Tot_Bedrooms      Population      Households        Latitude
##  Min.   :   1.0   Min.   :    3   Min.   :   1.0   Min.   :32.54
##  1st Qu.: 295.0   1st Qu.:  787   1st Qu.: 280.0   1st Qu.:33.93
##  Median : 435.0   Median : 1166   Median : 409.0   Median :34.26
##  Mean   : 537.9   Mean   : 1425   Mean   : 499.5   Mean   :35.63
##  3rd Qu.: 647.0   3rd Qu.: 1725   3rd Qu.: 605.0   3rd Qu.:37.71
##  Max.   :6445.0   Max.   :35682   Max.   :6082.0   Max.   :41.95
##    Longitude      Distance_to_coast  Distance_to_LA     Distance_to_SanDiego
##  Min.   :-124.3   Min.   :   120.7   Min.   :    420.6   Min.   :    484.9
##  1st Qu.:-121.8   1st Qu.:  9079.8   1st Qu.:  32111.3   1st Qu.: 159426.4
```

```
## Median :-118.5   Median : 20522.0   Median : 173667.5   Median : 214739.8
## Mean   :-119.6   Mean   : 40509.3   Mean   : 269422.0   Mean   : 398164.9
## 3rd Qu.:-118.0   3rd Qu.: 49830.4   3rd Qu.: 527156.2   3rd Qu.: 705795.4
## Max.   :-114.3   Max.   :333804.7   Max.   :1018260.1   Max.   :1196919.3
## Distance_to_SanJose Distance_to_SanFrancisco Closest_Distance
## Min.   :   569.4    Min.   :   456.1         Min.   :   420.6
## 1st Qu.:113119.9    1st Qu.:117395.5         1st Qu.: 17213.3
## Median :459758.9    Median :526546.7         Median : 36142.1
## Mean   :349187.6    Mean   :386688.4         Mean   : 69509.3
## 3rd Qu.:516946.5    3rd Qu.:584552.0         3rd Qu.: 93057.1
## Max.   :836762.7    Max.   :903627.7         Max.   :489626.4
```

```
# plot
```

MLR Model

```
#split train and test
set.seed(123)
train_flag <- sample.split(data_df,SplitRatio = 0.75)
train <- subset(data_df,train_flag==TRUE)
test <- subset(data_df,train_flag==FALSE)

#model fitting
lm1 <- lm(Median_House_Value ~ .,data = train)
summary(lm1)
```

```
##
## Call:
## lm(formula = Median_House_Value ~ ., data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -527607  -42184  -11337   28475  877363
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.828e+06  2.757e+05  -6.631 3.44e-11 ***
## Median_Income             3.809e+04  3.960e+02  96.182  < 2e-16 ***
## Median_Age                7.343e+02  5.325e+01  13.790  < 2e-16 ***
## Tot_Rooms                -3.136e+00  9.321e-01  -3.364  0.00077 ***
## Tot_Bedrooms              9.485e+01  8.054e+00  11.776  < 2e-16 ***
## Population               -4.021e+01  1.263e+00 -31.827  < 2e-16 ***
## Households                3.993e+01  8.667e+00   4.607 4.11e-06 ***
## Latitude                 -5.861e+04  3.417e+03 -17.151  < 2e-16 ***
## Longitude                -3.166e+04  2.002e+03 -15.814  < 2e-16 ***
## Distance_to_coast         3.700e-02  2.955e-02   1.252  0.21060
## Distance_to_LA           -1.565e-01  8.789e-03 -17.802  < 2e-16 ***
## Distance_to_SanDiego      4.371e-01  3.535e-02  12.363  < 2e-16 ***
## Distance_to_SanJose       1.140e-01  2.704e-02   4.215 2.51e-05 ***
## Distance_to_SanFrancisco  2.333e-02  3.110e-02   0.750  0.45323
## Closest_Distance         -2.606e-01  1.760e-02 -14.807  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68130 on 15121 degrees of freedom
## Multiple R-squared:  0.6505, Adjusted R-squared:  0.6502
## F-statistic:  2010 on 14 and 15121 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(1,2))


#Transformation based on BoxCox
boxCox(lm1)
summary(powerTransform(lm1))
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.1273        0.13       0.1062       0.1485
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df       pval
## LR test, lambda = (0) 141.3173  1 < 2.22e-16
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 5700.092  1 < 2.22e-16
```

```r
lm_transformed <- lm((Median_House_Value^0.15) ~ .,data = train)
summary(lm_transformed)
```

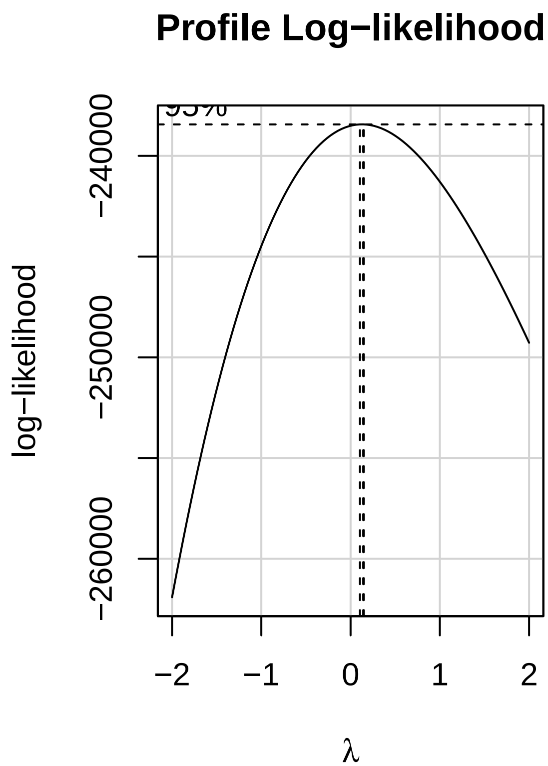
```
##
## Call:
## lm(formula = (Median_House_Value^0.15) ~ ., data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0516 -0.1807 -0.0230  0.1587  3.7170
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -7.739e+00  1.183e+00  -6.539 6.39e-11 ***
## Median_Income            1.511e-01  1.700e-03  88.912  < 2e-16 ***
## Median_Age               4.590e-04  2.286e-04   2.008 0.044627 *
## Tot_Rooms                2.395e-06  4.001e-06   0.599 0.549439
## Tot_Bedrooms             3.872e-04  3.457e-05  11.198  < 2e-16 ***
## Population              -1.733e-04  5.423e-06 -31.955  < 2e-16 ***
## Households               1.169e-04  3.720e-05   3.142 0.001678 **
## Latitude                -1.786e-01  1.467e-02 -12.174  < 2e-16 ***
## Longitude               -1.607e-01  8.593e-03 -18.703  < 2e-16 ***
## Distance_to_coast       -8.134e-07  1.269e-07  -6.412 1.47e-10 ***
## Distance_to_LA          -7.958e-07  3.773e-08 -21.093  < 2e-16 ***
## Distance_to_SanDiego     1.352e-06  1.517e-07   8.908  < 2e-16 ***
## Distance_to_SanJose      4.257e-07  1.161e-07   3.667 0.000246 ***
```

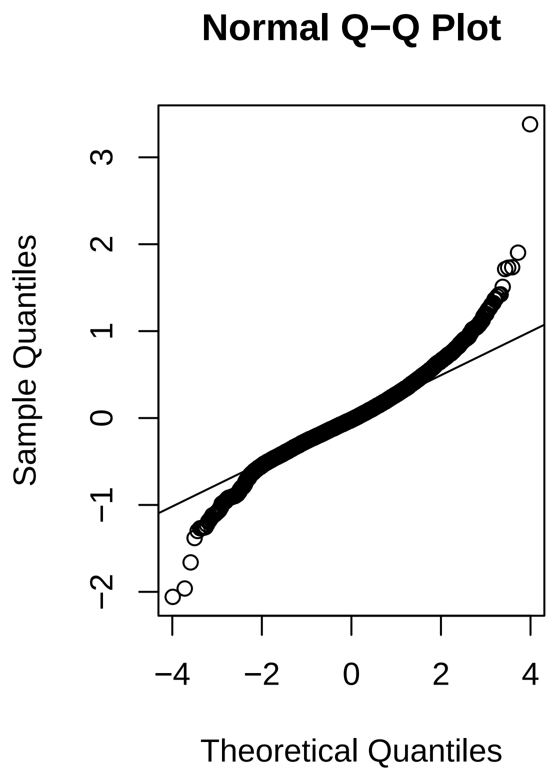
```
## Distance_to_SanFrancisco  2.259e-07  1.335e-07   1.692 0.090632 .
## Closest_Distance         -1.685e-06  7.555e-08 -22.298  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2925 on 15121 degrees of freedom
## Multiple R-squared:  0.6845, Adjusted R-squared:  0.6842
## F-statistic:  2343 on 14 and 15121 DF,  p-value: < 2.2e-16
```

```
# removing insignificant column
lm_preds <- lm((Median_House_Value^0.15) ~ Median_Income+Median_Age+Tot_Rooms+Tot_Bedrooms+Population+Ho
summary(lm_preds)
```

```
##
## Call:
## lm(formula = (Median_House_Value^0.15) ~ Median_Income + Median_Age +
##     Tot_Rooms + Tot_Bedrooms + Population + Households + Latitude +
##     Longitude + Distance_to_coast + Distance_to_LA + Distance_to_SanJose +
##     Distance_to_SanFrancisco, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0574 -0.1834 -0.0194  0.1684  3.3801
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -6.731e+00  7.377e-01  -9.125  < 2e-16 ***
## Median_Income             1.568e-01  1.708e-03  91.832  < 2e-16 ***
## Median_Age                1.538e-03  2.268e-04   6.781 1.23e-11 ***
## Tot_Rooms                -1.136e-05  4.015e-06  -2.830  0.00466 **
## Tot_Bedrooms              4.408e-04  3.487e-05  12.641  < 2e-16 ***
## Population               -1.630e-04  5.487e-06 -29.713  < 2e-16 ***
## Households                1.130e-04  3.761e-05   3.004  0.00267 **
## Latitude                 -7.775e-02  6.922e-03 -11.233  < 2e-16 ***
## Longitude                -1.283e-01  7.215e-03 -17.786  < 2e-16 ***
## Distance_to_coast        -2.599e-06  9.996e-08 -26.001  < 2e-16 ***
## Distance_to_LA           -7.295e-07  3.323e-08 -21.954  < 2e-16 ***
## Distance_to_SanJose       8.592e-07  1.160e-07   7.404 1.39e-13 ***
## Distance_to_SanFrancisco -9.687e-07  1.243e-07  -7.794 6.89e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2972 on 15123 degrees of freedom
## Multiple R-squared:  0.6741, Adjusted R-squared:  0.6739
## F-statistic:  2607 on 12 and 15123 DF,  p-value: < 2.2e-16
```

```
# Assumptions
qqnorm(resid(lm_preds))
qqline(resid(lm_transformed))
```






```r
#constat variance test
plot(predict(lm_preds),rstandard(lm_preds),xlab="Fitted values",ylab = "Standardized Residuals")
abline(h=0)
bptest(lm_transformed)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lm_transformed
## BP = 1038.8, df = 14, p-value < 2.2e-16
```

```r
#AIC for predictor selector
lm_AIC <- step(lm_preds)
```

```
## Start:  AIC=-36717.03
## (Median_House_Value^0.15) ~ Median_Income + Median_Age + Tot_Rooms +
##     Tot_Bedrooms + Population + Households + Latitude + Longitude +
##     Distance_to_coast + Distance_to_LA + Distance_to_SanJose +
##     Distance_to_SanFrancisco
##
##                            Df Sum of Sq    RSS    AIC
## <none>                                  1335.8 -36717
## - Tot_Rooms                 1      0.71 1336.5 -36711
## - Households                1      0.80 1336.6 -36710
## - Median_Age                1      4.06 1339.9 -36673
```

```
## - Distance_to_SanJose       1      4.84 1340.7 -36664
## - Distance_to_SanFrancisco  1      5.37 1341.2 -36658
## - Latitude                  1     11.14 1347.0 -36593
## - Tot_Bedrooms              1     14.11 1349.9 -36560
## - Longitude                 1     27.94 1363.8 -36406
## - Distance_to_LA            1     42.57 1378.4 -36244
## - Distance_to_coast         1     59.72 1395.5 -36057
## - Population                1     77.99 1413.8 -35860
## - Median_Income             1    744.90 2080.7 -30011
```

```
lm_AIC
```

```
##
## Call:
## lm(formula = (Median_House_Value^0.15) ~ Median_Income + Median_Age +
##     Tot_Rooms + Tot_Bedrooms + Population + Households + Latitude +
##     Longitude + Distance_to_coast + Distance_to_LA + Distance_to_SanJose +
##     Distance_to_SanFrancisco, data = train)
##
## Coefficients:
##              (Intercept)             Median_Income                Median_Age
##               -6.731e+00                 1.568e-01                 1.538e-03
##                Tot_Rooms              Tot_Bedrooms                Population
##               -1.136e-05                 4.408e-04                -1.630e-04
##               Households                  Latitude                 Longitude
##                1.130e-04                -7.775e-02                -1.283e-01
##        Distance_to_coast            Distance_to_LA       Distance_to_SanJose
##               -2.599e-06                -7.295e-07                 8.592e-07
## Distance_to_SanFrancisco
##               -9.687e-07
```

```
# VIF
vcov(lm_preds)
```

```
##                            (Intercept) Median_Income    Median_Age
## (Intercept)               5.441687e-01  2.549928e-05  8.830579e-06
## Median_Income             2.549928e-05  2.917084e-06  5.875999e-08
## Median_Age                8.830579e-06  5.875999e-08  5.145905e-08
## Tot_Rooms                -1.240739e-07 -4.179893e-09  2.584755e-11
## Tot_Bedrooms             -1.386142e-06  2.143393e-08  4.049584e-10
## Population               -4.720971e-08  1.997367e-09  5.915328e-11
## Households                2.201022e-06 -6.004393e-09 -2.013072e-10
## Latitude                  2.195590e-03  2.039295e-07  5.587716e-08
## Longitude                 5.141887e-03  4.229975e-07  1.155513e-07
## Distance_to_coast        -4.046844e-08  2.571913e-11  1.072244e-12
## Distance_to_LA            4.497629e-09  6.852151e-12  1.975471e-12
## Distance_to_SanJose      -1.953087e-09  4.633188e-12 -5.814380e-12
## Distance_to_SanFrancisco -1.754220e-08 -1.060802e-13  6.415649e-12
##                              Tot_Rooms  Tot_Bedrooms    Population
## (Intercept)              -1.240739e-07 -1.386142e-06 -4.720971e-08
## Median_Income            -4.179893e-09  2.143393e-08  1.997367e-09
## Median_Age                2.584755e-11  4.049584e-10  5.915328e-11
## Tot_Rooms                 1.612030e-11 -7.167732e-11 -6.803780e-12
```

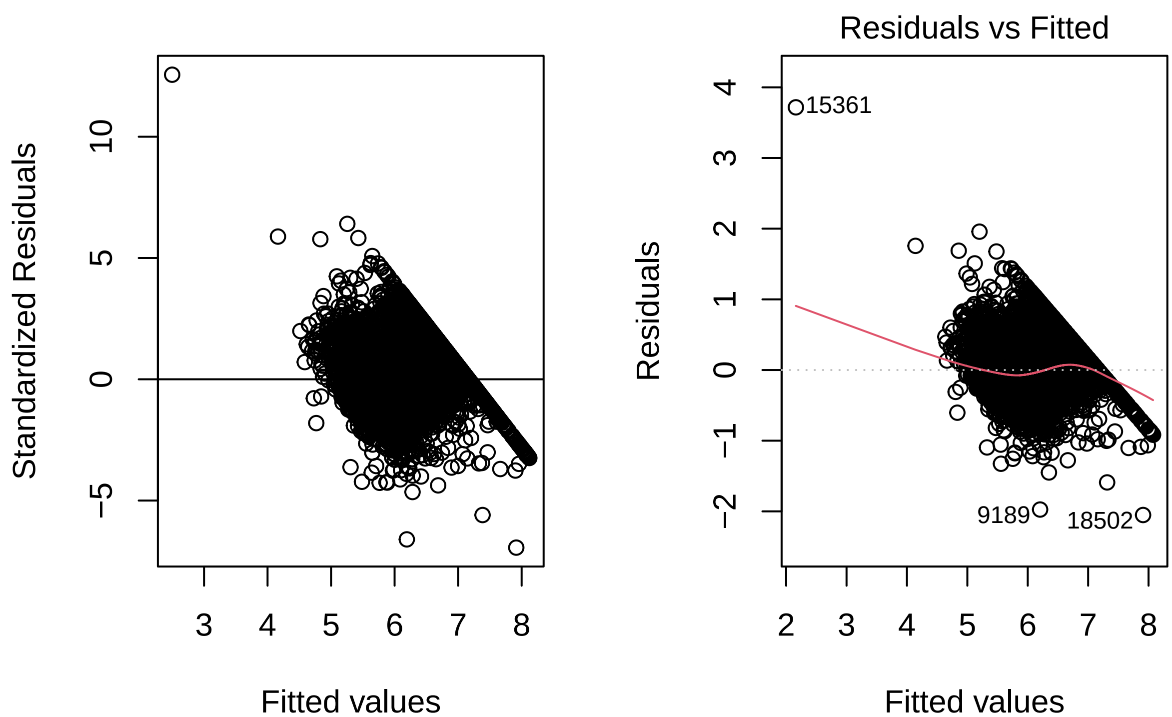
```
## Tot_Bedrooms            -7.167732e-11  1.215749e-09  5.383543e-11
## Population              -6.803780e-12  5.383543e-11  3.011129e-11
## Households               1.158240e-11 -1.079139e-09 -1.017233e-10
## Latitude                 1.829327e-10 -7.115630e-09 -1.716193e-09
## Longitude               -1.164506e-09 -1.266193e-08 -7.621359e-10
## Distance_to_coast       -3.773001e-14  5.519841e-14  2.019140e-14
## Distance_to_LA          -1.660286e-14  4.094714e-14  1.497840e-14
## Distance_to_SanJose     -8.923591e-15 -7.210179e-14  4.062744e-14
## Distance_to_SanFrancisco  7.435721e-15  1.043890e-13 -3.702362e-14
##                           Households      Latitude     Longitude
## (Intercept)              2.201022e-06  2.195590e-03  5.141887e-03
## Median_Income           -6.004393e-09  2.039295e-07  4.229975e-07
## Median_Age              -2.013072e-10  5.587716e-08  1.155513e-07
## Tot_Rooms                1.158240e-11  1.829327e-10 -1.164506e-09
## Tot_Bedrooms            -1.079139e-09 -7.115630e-09 -1.266193e-08
## Population              -1.017233e-10 -1.716193e-09 -7.621359e-10
## Households               1.414719e-09  1.203224e-08  2.187480e-08
## Latitude                 1.203224e-08  4.791452e-05  3.234083e-05
## Longitude                2.187480e-08  3.234083e-05  5.206094e-05
## Distance_to_coast        1.167568e-13 -5.414877e-10 -4.936999e-10
## Distance_to_LA           2.206893e-14 -9.361196e-11  1.331505e-11
## Distance_to_SanJose     -4.560048e-14 -2.812751e-10 -1.082628e-10
## Distance_to_SanFrancisco  2.077341e-14  2.763843e-10 -5.179780e-11
##                          Distance_to_coast Distance_to_LA Distance_to_SanJose
## (Intercept)                  -4.046844e-08   4.497629e-09       -1.953087e-09
## Median_Income                 2.571913e-11   6.852151e-12        4.633188e-12
## Median_Age                    1.072244e-12   1.975471e-12       -5.814380e-12
## Tot_Rooms                    -3.773001e-14  -1.660286e-14       -8.923591e-15
## Tot_Bedrooms                  5.519841e-14   4.094714e-14       -7.210179e-14
## Population                    2.019140e-14   1.497840e-14        4.062744e-14
## Households                    1.167568e-13   2.206893e-14       -4.560048e-14
## Latitude                     -5.414877e-10  -9.361196e-11       -2.812751e-10
## Longitude                    -4.936999e-10   1.331505e-11       -1.082628e-10
## Distance_to_coast             9.991377e-15   3.330578e-16        4.422326e-15
## Distance_to_LA                3.330578e-16   1.104050e-15       -1.173169e-15
## Distance_to_SanJose           4.422326e-15  -1.173169e-15        1.346611e-14
## Distance_to_SanFrancisco     -3.763319e-15   1.125143e-15       -1.385885e-14
##                          Distance_to_SanFrancisco
## (Intercept)                         -1.754220e-08
## Median_Income                       -1.060802e-13
## Median_Age                           6.415649e-12
## Tot_Rooms                            7.435721e-15
## Tot_Bedrooms                         1.043890e-13
## Population                          -3.702362e-14
## Households                           2.077341e-14
## Latitude                             2.763843e-10
## Longitude                           -5.179780e-11
## Distance_to_coast                   -3.763319e-15
## Distance_to_LA                       1.125143e-15
## Distance_to_SanJose                 -1.385885e-14
## Distance_to_SanFrancisco             1.544587e-14
```

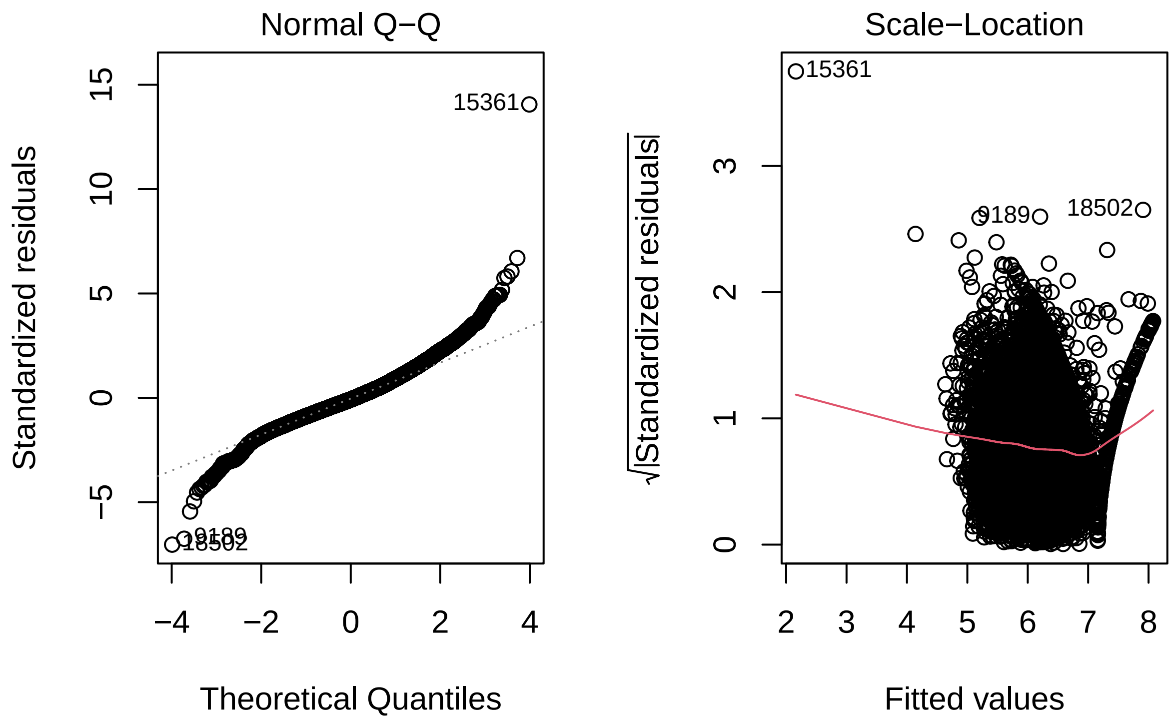
```r
mlr.vif <- vif(lm_preds)
mlr_vif <- data.frame(Columns=names(mlr.vif),vif_value=mlr.vif)
```

```
sqldf("select * from mlr_vif where vif_value<=10")
```

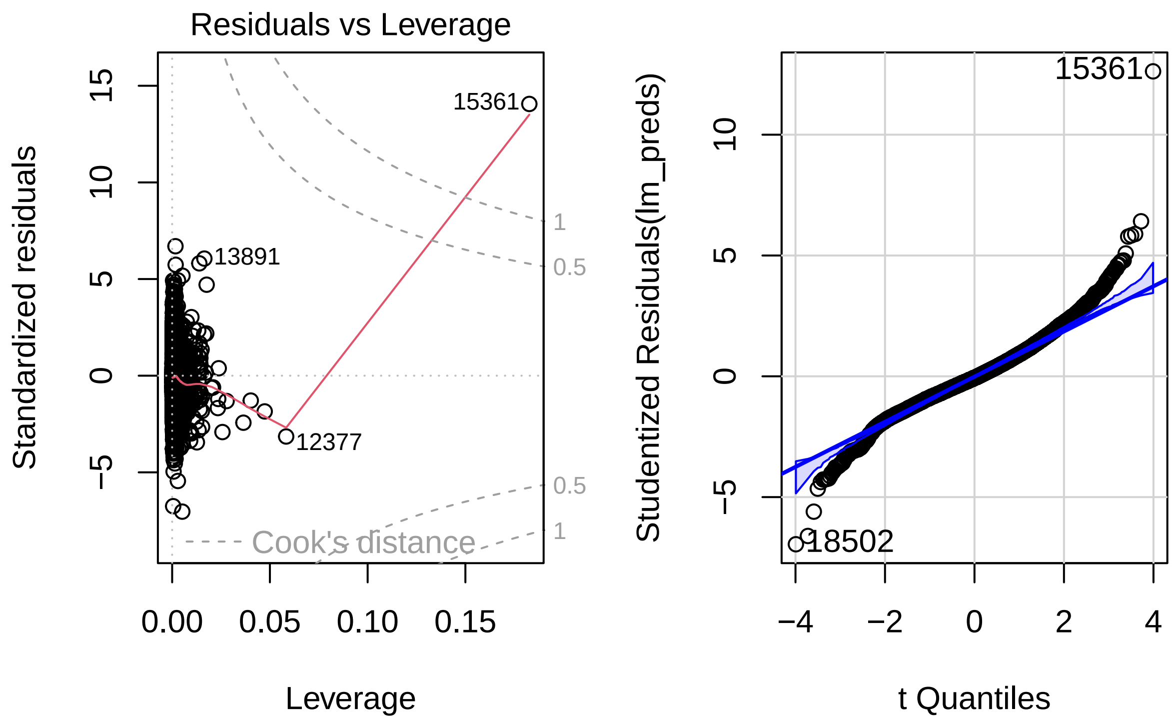
```
##             Columns vif_value
## 1     Median_Income  1.811116
## 2        Median_Age  1.395060
## 3        Population  6.413894
## 4 Distance_to_coast  4.117661
```

```
#Make predictions
lm_preds.probs <- predict(lm_preds,newdata = test[-1])

#confusion_matrix
lm_preds.cm <- table(prediction=lm_preds.probs,actual=test$Median_House_Value)
#addmargins(lm_preds.cm)
cm<-data.frame(lm_preds.cm)

#MISC/test
plot(lm_transformed)
```






```
qqPlot(lm_preds)
```




```
## 15361 18502
## 11265 13568
```

```
sqldf("select count(*) from data_df where Tot_Bedrooms <= 500")
```

```
##   count(*)
## 1    12303
```

```
# ROC CURVE

rc <- roc(Median_House_Value ~ predict(lm_preds),data = train)
```

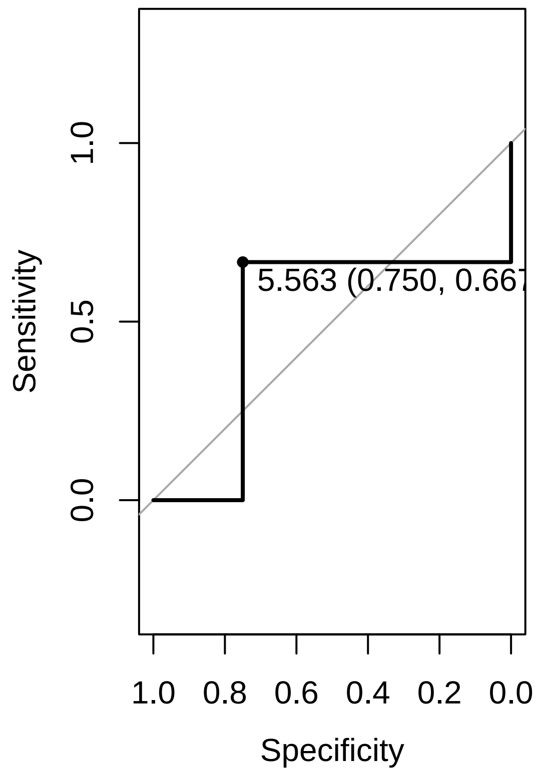
```
## Warning in roc.default(response, predictors[, 1], ...): 'response' has more
## than two levels. Consider setting 'levels' explicitly or using 'multiclass.roc'
## instead
```

```
## Setting levels: control = 14999, case = 22500
```

```
## Setting direction: controls < cases
```

```
plot(rc,print.thres="best")
auc(rc)
```

```
## Area under the curve: 0.5
```



KNN MODEL

```
knn_train <- sqldf("select Median_House_Value,Median_Income, Median_Age, Tot_Rooms, Distance_to_coast, C
knn_test <- sqldf("select Median_House_Value,Median_Income, Median_Age, Tot_Rooms, Distance_to_coast, Cl

train_scale <- scale(knn_train)
test_scale <- scale(knn_test)

c_knn <- knn(train = train_scale,test = test_scale,cl=knn_train$Median_House_Value,k=6)
```

Random Forest

```
set.seed(123)
rf_classifier <- randomForest(x=train[-1],y=train$Median_House_Value,ntree = 500)
rf_predict <- predict(rf_classifier,newdata = test[-1])

rf_cm <- data.frame(predicted_value=rf_predict,actual_value=test$Median_House_Value)
accuracy_matrix <- sqldf("select case
        when predicted_value >= (actual_value-(0.3*actual_value)) and predicted_value <= (actual_value+
```

```
        else 0
        end as flag
      from rf_cm")
accuracy <- sqldf("select count(*) as accuracy from accuracy_matrix where flag=1")/nrow(accuracy_matrix)
accuracy
```

```
##    accuracy
## 1 0.8846294
```

XGBoost

```
xgb_classifier <- xgboost(data=as.matrix(train[-1]),label = train$Median_House_Value,nrounds=100,object
```

```
## [1]  train-rmse:172166.398433
## [2]  train-rmse:128090.036932
## [3]  train-rmse:98276.474342
## [4]  train-rmse:79342.444810
## [5]  train-rmse:66925.631029
## [6]  train-rmse:59296.774811
## [7]  train-rmse:54825.709968
## [8]  train-rmse:51875.268928
## [9]  train-rmse:49779.604504
## [10] train-rmse:48342.423292
## [11] train-rmse:46835.283411
## [12] train-rmse:45799.033110
## [13] train-rmse:44948.446237
## [14] train-rmse:44302.768135
## [15] train-rmse:43718.078315
## [16] train-rmse:42983.712331
## [17] train-rmse:42310.206620
## [18] train-rmse:41995.381489
## [19] train-rmse:41537.250904
## [20] train-rmse:41060.074472
## [21] train-rmse:40256.356672
## [22] train-rmse:39928.783411
## [23] train-rmse:39646.494451
## [24] train-rmse:39203.211269
## [25] train-rmse:38884.162994
## [26] train-rmse:38610.352340
## [27] train-rmse:38335.534083
## [28] train-rmse:38062.639725
## [29] train-rmse:37709.041497
## [30] train-rmse:37309.734070
## [31] train-rmse:37214.739387
## [32] train-rmse:36957.231362
## [33] train-rmse:36686.654582
## [34] train-rmse:36248.359505
## [35] train-rmse:36033.441341
## [36] train-rmse:35776.047412
## [37] train-rmse:35486.457906
## [38] train-rmse:35310.891290
```

```
## [39] train-rmse:35213.957309
## [40] train-rmse:35018.337531
## [41] train-rmse:34747.714404
## [42] train-rmse:34556.494673
## [43] train-rmse:34431.884368
## [44] train-rmse:34334.086709
## [45] train-rmse:34065.511235
## [46] train-rmse:33827.139149
## [47] train-rmse:33576.640172
## [48] train-rmse:33259.422294
## [49] train-rmse:33059.292096
## [50] train-rmse:32909.418605
## [51] train-rmse:32625.090946
## [52] train-rmse:32439.070817
## [53] train-rmse:32319.960798
## [54] train-rmse:32152.959628
## [55] train-rmse:31873.814217
## [56] train-rmse:31715.095351
## [57] train-rmse:31527.229054
## [58] train-rmse:31373.878759
## [59] train-rmse:31200.399669
## [60] train-rmse:31103.951491
## [61] train-rmse:30948.825385
## [62] train-rmse:30811.974770
## [63] train-rmse:30671.467423
## [64] train-rmse:30510.585114
## [65] train-rmse:30456.995786
## [66] train-rmse:30291.185187
## [67] train-rmse:30077.804081
## [68] train-rmse:29868.307554
## [69] train-rmse:29743.216378
## [70] train-rmse:29614.536490
## [71] train-rmse:29456.062350
## [72] train-rmse:29218.502538
## [73] train-rmse:29099.026841
## [74] train-rmse:28907.625607
## [75] train-rmse:28755.061823
## [76] train-rmse:28662.736794
## [77] train-rmse:28474.116043
## [78] train-rmse:28280.489834
## [79] train-rmse:28070.792107
## [80] train-rmse:27935.310867
## [81] train-rmse:27908.077929
## [82] train-rmse:27785.383295
## [83] train-rmse:27645.026627
## [84] train-rmse:27420.013450
## [85] train-rmse:27340.702700
## [86] train-rmse:27263.101911
## [87] train-rmse:27213.418598
## [88] train-rmse:27077.234314
## [89] train-rmse:26955.989069
## [90] train-rmse:26881.838447
## [91] train-rmse:26789.499172
## [92] train-rmse:26684.695859
```

```
## [93] train-rmse:26672.967067
## [94] train-rmse:26595.825460
## [95] train-rmse:26515.423282
## [96] train-rmse:26402.732669
## [97] train-rmse:26362.332558
## [98] train-rmse:26227.543386
## [99] train-rmse:26052.961896
## [100]    train-rmse:25986.936601
```

```r
xgb_predictions <- predict(xgb_classifier, as.matrix(test[-1]))
```