Project v

Isaiah Thompson Ocansey

4/15/2022

Communities and Crime Data Set ***

```
#loading the data
df0 <- read.table(file = "crime.csv",sep=",", header = T, na.strings = c("NA", "", " "),
                  stringsAsFactors = T)
dim(df0)
```

```
## [1] 1994  128
```

```
head(df0)
```

```
##   state county community       communityname fold population householdsize
## 1     8     NA        NA        Lakewoodcity    1       0.19          0.33
## 2    53     NA        NA         Tukwilacity    1       0.00          0.16
## 3    24     NA        NA        Aberdeentown    1       0.00          0.42
## 4    34      5     81440 Willingborotownship    1       0.04          0.77
## 5    42     95      6096   Bethlehemtownship    1       0.01          0.55
## 6     6     NA        NA   SouthPasadenacity    1       0.02          0.28
##   racepctblack racePctWhite racePctAsian racePctHisp agePct12t21 agePct12t29
## 1         0.02         0.90         0.12        0.17        0.34        0.47
## 2         0.12         0.74         0.45        0.07        0.26        0.59
## 3         0.49         0.56         0.17        0.04        0.39        0.47
## 4         1.00         0.08         0.12        0.10        0.51        0.50
## 5         0.02         0.95         0.09        0.05        0.38        0.38
## 6         0.06         0.54         1.00        0.25        0.31        0.48
##   agePct16t24 agePct65up numbUrban pctUrban medIncome pctWWage pctWFarmSelf
## 1        0.29       0.32      0.20      1.0      0.37     0.72         0.34
## 2        0.35       0.27      0.02      1.0      0.31     0.72         0.11
## 3        0.28       0.32      0.00      0.0      0.30     0.58         0.19
## 4        0.34       0.21      0.06      1.0      0.58     0.89         0.21
## 5        0.23       0.36      0.02      0.9      0.50     0.72         0.16
## 6        0.27       0.37      0.04      1.0      0.52     0.68         0.20
##   pctWInvInc pctWSocSec pctWPubAsst pctWRetire medFamInc perCapInc whitePerCap
## 1       0.60       0.29        0.15       0.43      0.39      0.40        0.39
## 2       0.45       0.25        0.29       0.39      0.29      0.37        0.38
## 3       0.39       0.38        0.40       0.84      0.28      0.27        0.29
## 4       0.43       0.36        0.20       0.82      0.51      0.36        0.40
## 5       0.68       0.44        0.11       0.71      0.46      0.43        0.41
## 6       0.61       0.28        0.15       0.25      0.62      0.72        0.76
```

```
##   blackPerCap indianPerCap AsianPerCap OtherPerCap HispPerCap NumUnderPov
## 1        0.32         0.27        0.27        0.36       0.41        0.08
## 2        0.33         0.16        0.30        0.22       0.35        0.01
## 3        0.27         0.07        0.29        0.28       0.39        0.01
## 4        0.39         0.16        0.25        0.36       0.44        0.01
## 5        0.28         0.00        0.74        0.51       0.48        0.00
## 6        0.77         0.28        0.52        0.48       0.60        0.01
##   PctPopUnderPov PctLess9thGrade PctNotHSGrad PctBSorMore PctUnemployed
## 1           0.19            0.10         0.18        0.48          0.27
## 2           0.24            0.14         0.24        0.30          0.27
## 3           0.27            0.27         0.43        0.19          0.36
## 4           0.10            0.09         0.25        0.31          0.33
## 5           0.06            0.25         0.30        0.33          0.12
## 6           0.12            0.13         0.12        0.80          0.10
##   PctEmploy PctEmplManu PctEmplProfServ PctOccupManu PctOccupMgmtProf
## 1      0.68        0.23            0.41         0.25             0.52
## 2      0.73        0.57            0.15         0.42             0.36
## 3      0.58        0.32            0.29         0.49             0.32
## 4      0.71        0.36            0.45         0.37             0.39
## 5      0.65        0.67            0.38         0.42             0.46
## 6      0.65        0.19            0.77         0.06             0.91
##   MalePctDivorce MalePctNevMarr FemalePctDiv TotalPctDiv PersPerFam PctFam2Par
## 1           0.68           0.40         0.75        0.75       0.35       0.55
## 2           1.00           0.63         0.91        1.00       0.29       0.43
## 3           0.63           0.41         0.71        0.70       0.45       0.42
## 4           0.34           0.45         0.49        0.44       0.75       0.65
## 5           0.22           0.27         0.20        0.21       0.51       0.91
## 6           0.49           0.57         0.61        0.58       0.44       0.62
##   PctKids2Par PctYoungKids2Par PctTeen2Par PctWorkMomYoungKids PctWorkMom
## 1        0.59             0.61        0.56                0.74       0.76
## 2        0.47             0.60        0.39                0.46       0.53
## 3        0.44             0.43        0.43                0.71       0.67
## 4        0.54             0.83        0.65                0.85       0.86
## 5        0.91             0.89        0.85                0.40       0.60
## 6        0.69             0.87        0.53                0.30       0.43
##   NumIlleg PctIlleg NumImmig PctImmigRecent PctImmigRec5 PctImmigRec8
## 1     0.04     0.14     0.03           0.24         0.27         0.37
## 2     0.00     0.24     0.01           0.52         0.62         0.64
## 3     0.01     0.46     0.00           0.07         0.06         0.15
## 4     0.03     0.33     0.02           0.11         0.20         0.30
## 5     0.00     0.06     0.00           0.03         0.07         0.20
## 6     0.00     0.11     0.04           0.30         0.35         0.43
##   PctImmigRec10 PctRecentImmig PctRecImmig5 PctRecImmig8 PctRecImmig10
## 1          0.39           0.07         0.07         0.08          0.08
## 2          0.63           0.25         0.27         0.25          0.23
## 3          0.19           0.02         0.02         0.04          0.05
## 4          0.31           0.05         0.08         0.11          0.11
## 5          0.27           0.01         0.02         0.04          0.05
## 6          0.47           0.50         0.50         0.56          0.57
##   PctSpeakEnglOnly PctNotSpeakEnglWell PctLargHouseFam PctLargHouseOccup
## 1             0.89                0.06            0.14              0.13
## 2             0.84                0.10            0.16              0.10
## 3             0.88                0.04            0.20              0.20
## 4             0.81                0.08            0.56              0.62
```

```
## 5             0.88                0.05            0.16              0.19
## 6             0.45                0.28            0.25              0.19
##   PersPerOccupHous PersPerOwnOccHous PersPerRentOccHous PctPersOwnOccup
## 1             0.33              0.39               0.28            0.55
## 2             0.17              0.29               0.17            0.26
## 3             0.46              0.52               0.43            0.42
## 4             0.85              0.77               1.00            0.94
## 5             0.59              0.60               0.37            0.89
## 6             0.29              0.53               0.18            0.39
##   PctPersDenseHous PctHousLess3BR MedNumBR HousVacant PctHousOccup
## 1             0.09           0.51      0.5       0.21         0.71
## 2             0.20           0.82      0.0       0.02         0.79
## 3             0.15           0.51      0.5       0.01         0.86
## 4             0.12           0.01      0.5       0.01         0.97
## 5             0.02           0.19      0.5       0.01         0.89
## 6             0.26           0.73      0.0       0.02         0.84
##   PctHousOwnOcc PctVacantBoarded PctVacMore6Mos MedYrHousBuilt PctHousNoPhone
## 1          0.52             0.05           0.26           0.65           0.14
## 2          0.24             0.02           0.25           0.65           0.16
## 3          0.41             0.29           0.30           0.52           0.47
## 4          0.96             0.60           0.47           0.52           0.11
## 5          0.87             0.04           0.55           0.73           0.05
## 6          0.30             0.16           0.28           0.25           0.02
##   PctWOFullPlumb OwnOccLowQuart OwnOccMedVal OwnOccHiQuart RentLowQ RentMedian
## 1           0.06           0.22         0.19          0.18     0.36       0.35
## 2           0.00           0.21         0.20          0.21     0.42       0.38
## 3           0.45           0.18         0.17          0.16     0.27       0.29
## 4           0.11           0.24         0.21          0.19     0.75       0.70
## 5           0.14           0.31         0.31          0.30     0.40       0.36
## 6           0.05           0.94         1.00          1.00     0.67       0.63
##   RentHighQ MedRent MedRentPctHousInc MedOwnCostPctInc MedOwnCostPctIncNoMtg
## 1      0.38    0.34              0.38             0.46                  0.25
## 2      0.40    0.37              0.29             0.32                  0.18
## 3      0.27    0.31              0.48             0.39                  0.28
## 4      0.77    0.89              0.63             0.51                  0.47
## 5      0.38    0.38              0.22             0.51                  0.21
## 6      0.68    0.62              0.47             0.59                  0.11
##   NumInShelters NumStreet PctForeignBorn PctBornSameState PctSameHouse85
## 1          0.04         0           0.12             0.42           0.50
## 2          0.00         0           0.21             0.50           0.34
## 3          0.00         0           0.14             0.49           0.54
## 4          0.00         0           0.19             0.30           0.73
## 5          0.00         0           0.11             0.72           0.64
## 6          0.00         0           0.70             0.42           0.49
##   PctSameCity85 PctSameState85 LemasSwornFT LemasSwFTPerPop LemasSwFTFieldOps
## 1          0.51           0.64         0.03            0.13              0.96
## 2          0.60           0.52           NA              NA                NA
## 3          0.67           0.56           NA              NA                NA
## 4          0.64           0.65           NA              NA                NA
## 5          0.61           0.53           NA              NA                NA
## 6          0.73           0.64           NA              NA                NA
##   LemasSwFTFieldPerPop LemasTotalReq LemasTotReqPerPop PolicReqPerOffic
## 1                 0.17          0.06              0.18             0.44
## 2                   NA            NA                NA               NA
```

```
## 3                   NA            NA                 NA                NA
## 4                   NA            NA                 NA                NA
## 5                   NA            NA                 NA                NA
## 6                   NA            NA                 NA                NA
##   PolicPerPop RacialMatchCommPol PctPolicWhite PctPolicBlack PctPolicHisp
## 1        0.13               0.94          0.93          0.03         0.07
## 2          NA                 NA            NA            NA           NA
## 3          NA                 NA            NA            NA           NA
## 4          NA                 NA            NA            NA           NA
## 5          NA                 NA            NA            NA           NA
## 6          NA                 NA            NA            NA           NA
##   PctPolicAsian PctPolicMinor OfficAssgnDrugUnits NumKindsDrugsSeiz
## 1           0.1          0.07                0.02              0.57
## 2            NA            NA                  NA                NA
## 3            NA            NA                  NA                NA
## 4            NA            NA                  NA                NA
## 5            NA            NA                  NA                NA
## 6            NA            NA                  NA                NA
##   PolicAveOTWorked LandArea PopDens PctUsePubTrans PolicCars PolicOperBudg
## 1             0.29     0.12    0.26           0.20      0.06          0.04
## 2               NA     0.02    0.12           0.45        NA            NA
## 3               NA     0.01    0.21           0.02        NA            NA
## 4               NA     0.02    0.39           0.28        NA            NA
## 5               NA     0.04    0.09           0.02        NA            NA
## 6               NA     0.01    0.58           0.10        NA            NA
##   LemasPctPolicOnPatr LemasGangUnitDeploy LemasPctOfficDrugUn PolicBudgPerPop
## 1                 0.9                 0.5                0.32            0.14
## 2                  NA                  NA                0.00              NA
## 3                  NA                  NA                0.00              NA
## 4                  NA                  NA                0.00              NA
## 5                  NA                  NA                0.00              NA
## 6                  NA                  NA                0.00              NA
##   ViolentCrimesPerPop
## 1                0.20
## 2                0.67
## 3                0.43
## 4                0.12
## 5                0.03
## 6                0.14
```

The data above has 1994 observations with 128 variables.

DATA PPREPARATION ***

1(a) Remove the first five columns from the data since these are not predictive. . . . .

```r
# Removing first five columns
df <- df0[,-c(1:5)]
head(df)
```

```
##   population householdsize racepctblack racePctWhite racePctAsian racePctHisp
```

```
## 1       0.19           0.33          0.02          0.90          0.12          0.17
## 2       0.00           0.16          0.12          0.74          0.45          0.07
## 3       0.00           0.42          0.49          0.56          0.17          0.04
## 4       0.04           0.77          1.00          0.08          0.12          0.10
## 5       0.01           0.55          0.02          0.95          0.09          0.05
## 6       0.02           0.28          0.06          0.54          1.00          0.25
##   agePct12t21 agePct12t29 agePct16t24 agePct65up numbUrban pctUrban medIncome
## 1        0.34        0.47        0.29       0.32      0.20      1.0      0.37
## 2        0.26        0.59        0.35       0.27      0.02      1.0      0.31
## 3        0.39        0.47        0.28       0.32      0.00      0.0      0.30
## 4        0.51        0.50        0.34       0.21      0.06      1.0      0.58
## 5        0.38        0.38        0.23       0.36      0.02      0.9      0.50
## 6        0.31        0.48        0.27       0.37      0.04      1.0      0.52
##   pctWWage pctWFarmSelf pctWInvInc pctWSocSec pctWPubAsst pctWRetire medFamInc
## 1     0.72         0.34       0.60       0.29        0.15       0.43      0.39
## 2     0.72         0.11       0.45       0.25        0.29       0.39      0.29
## 3     0.58         0.19       0.39       0.38        0.40       0.84      0.28
## 4     0.89         0.21       0.43       0.36        0.20       0.82      0.51
## 5     0.72         0.16       0.68       0.44        0.11       0.71      0.46
## 6     0.68         0.20       0.61       0.28        0.15       0.25      0.62
##   perCapInc whitePerCap blackPerCap indianPerCap AsianPerCap OtherPerCap
## 1      0.40        0.39        0.32         0.27        0.27        0.36
## 2      0.37        0.38        0.33         0.16        0.30        0.22
## 3      0.27        0.29        0.27         0.07        0.29        0.28
## 4      0.36        0.40        0.39         0.16        0.25        0.36
## 5      0.43        0.41        0.28         0.00        0.74        0.51
## 6      0.72        0.76        0.77         0.28        0.52        0.48
##   HispPerCap NumUnderPov PctPopUnderPov PctLess9thGrade PctNotHSGrad
## 1       0.41        0.08           0.19            0.10         0.18
## 2       0.35        0.01           0.24            0.14         0.24
## 3       0.39        0.01           0.27            0.27         0.43
## 4       0.44        0.01           0.10            0.09         0.25
## 5       0.48        0.00           0.06            0.25         0.30
## 6       0.60        0.01           0.12            0.13         0.12
##   PctBSorMore PctUnemployed PctEmploy PctEmplManu PctEmplProfServ PctOccupManu
## 1        0.48          0.27      0.68        0.23            0.41         0.25
## 2        0.30          0.27      0.73        0.57            0.15         0.42
## 3        0.19          0.36      0.58        0.32            0.29         0.49
## 4        0.31          0.33      0.71        0.36            0.45         0.37
## 5        0.33          0.12      0.65        0.67            0.38         0.42
## 6        0.80          0.10      0.65        0.19            0.77         0.06
##   PctOccupMgmtProf MalePctDivorce MalePctNevMarr FemalePctDiv TotalPctDiv
## 1             0.52           0.68           0.40         0.75        0.75
## 2             0.36           1.00           0.63         0.91        1.00
## 3             0.32           0.63           0.41         0.71        0.70
## 4             0.39           0.34           0.45         0.49        0.44
## 5             0.46           0.22           0.27         0.20        0.21
## 6             0.91           0.49           0.57         0.61        0.58
##   PersPerFam PctFam2Par PctKids2Par PctYoungKids2Par PctTeen2Par
## 1       0.35       0.55        0.59             0.61        0.56
## 2       0.29       0.43        0.47             0.60        0.39
## 3       0.45       0.42        0.44             0.43        0.43
## 4       0.75       0.65        0.54             0.83        0.65
## 5       0.51       0.91        0.91             0.89        0.85
```

```
## 6       0.44       0.62        0.69             0.87        0.53
##   PctWorkMomYoungKids PctWorkMom NumIlleg PctIlleg NumImmig PctImmigRecent
## 1                0.74       0.76     0.04     0.14     0.03           0.24
## 2                0.46       0.53     0.00     0.24     0.01           0.52
## 3                0.71       0.67     0.01     0.46     0.00           0.07
## 4                0.85       0.86     0.03     0.33     0.02           0.11
## 5                0.40       0.60     0.00     0.06     0.00           0.03
## 6                0.30       0.43     0.00     0.11     0.04           0.30
##   PctImmigRec5 PctImmigRec8 PctImmigRec10 PctRecentImmig PctRecImmig5
## 1         0.27         0.37          0.39           0.07         0.07
## 2         0.62         0.64          0.63           0.25         0.27
## 3         0.06         0.15          0.19           0.02         0.02
## 4         0.20         0.30          0.31           0.05         0.08
## 5         0.07         0.20          0.27           0.01         0.02
## 6         0.35         0.43          0.47           0.50         0.50
##   PctRecImmig8 PctRecImmig10 PctSpeakEnglOnly PctNotSpeakEnglWell
## 1         0.08          0.08             0.89                0.06
## 2         0.25          0.23             0.84                0.10
## 3         0.04          0.05             0.88                0.04
## 4         0.11          0.11             0.81                0.08
## 5         0.04          0.05             0.88                0.05
## 6         0.56          0.57             0.45                0.28
##   PctLargHouseFam PctLargHouseOccup PersPerOccupHous PersPerOwnOccHous
## 1            0.14              0.13             0.33              0.39
## 2            0.16              0.10             0.17              0.29
## 3            0.20              0.20             0.46              0.52
## 4            0.56              0.62             0.85              0.77
## 5            0.16              0.19             0.59              0.60
## 6            0.25              0.19             0.29              0.53
##   PersPerRentOccHous PctPersOwnOccup PctPersDenseHous PctHousLess3BR MedNumBR
## 1               0.28            0.55             0.09           0.51      0.5
## 2               0.17            0.26             0.20           0.82      0.0
## 3               0.43            0.42             0.15           0.51      0.5
## 4               1.00            0.94             0.12           0.01      0.5
## 5               0.37            0.89             0.02           0.19      0.5
## 6               0.18            0.39             0.26           0.73      0.0
##   HousVacant PctHousOccup PctHousOwnOcc PctVacantBoarded PctVacMore6Mos
## 1       0.21         0.71          0.52             0.05           0.26
## 2       0.02         0.79          0.24             0.02           0.25
## 3       0.01         0.86          0.41             0.29           0.30
## 4       0.01         0.97          0.96             0.60           0.47
## 5       0.01         0.89          0.87             0.04           0.55
## 6       0.02         0.84          0.30             0.16           0.28
##   MedYrHousBuilt PctHousNoPhone PctWOFullPlumb OwnOccLowQuart OwnOccMedVal
## 1           0.65           0.14           0.06           0.22         0.19
## 2           0.65           0.16           0.00           0.21         0.20
## 3           0.52           0.47           0.45           0.18         0.17
## 4           0.52           0.11           0.11           0.24         0.21
## 5           0.73           0.05           0.14           0.31         0.31
## 6           0.25           0.02           0.05           0.94         1.00
##   OwnOccHiQuart RentLowQ RentMedian RentHighQ MedRent MedRentPctHousInc
## 1          0.18     0.36       0.35      0.38    0.34              0.38
## 2          0.21     0.42       0.38      0.40    0.37              0.29
## 3          0.16     0.27       0.29      0.27    0.31              0.48
```

```
## 4         0.19     0.75       0.70      0.77    0.89              0.63
## 5         0.30     0.40       0.36      0.38    0.38              0.22
## 6         1.00     0.67       0.63      0.68    0.62              0.47
##   MedOwnCostPctInc MedOwnCostPctIncNoMtg NumInShelters NumStreet PctForeignBorn
## 1             0.46                  0.25          0.04         0           0.12
## 2             0.32                  0.18          0.00         0           0.21
## 3             0.39                  0.28          0.00         0           0.14
## 4             0.51                  0.47          0.00         0           0.19
## 5             0.51                  0.21          0.00         0           0.11
## 6             0.59                  0.11          0.00         0           0.70
##   PctBornSameState PctSameHouse85 PctSameCity85 PctSameState85 LemasSwornFT
## 1             0.42           0.50          0.51           0.64         0.03
## 2             0.50           0.34          0.60           0.52           NA
## 3             0.49           0.54          0.67           0.56           NA
## 4             0.30           0.73          0.64           0.65           NA
## 5             0.72           0.64          0.61           0.53           NA
## 6             0.42           0.49          0.73           0.64           NA
##   LemasSwFTPerPop LemasSwFTFieldOps LemasSwFTFieldPerPop LemasTotalReq
## 1            0.13              0.96                 0.17          0.06
## 2              NA                NA                   NA            NA
## 3              NA                NA                   NA            NA
## 4              NA                NA                   NA            NA
## 5              NA                NA                   NA            NA
## 6              NA                NA                   NA            NA
##   LemasTotReqPerPop PolicReqPerOffic PolicPerPop RacialMatchCommPol
## 1              0.18             0.44        0.13               0.94
## 2                NA               NA          NA                 NA
## 3                NA               NA          NA                 NA
## 4                NA               NA          NA                 NA
## 5                NA               NA          NA                 NA
## 6                NA               NA          NA                 NA
##   PctPolicWhite PctPolicBlack PctPolicHisp PctPolicAsian PctPolicMinor
## 1          0.93          0.03         0.07           0.1          0.07
## 2            NA            NA           NA            NA            NA
## 3            NA            NA           NA            NA            NA
## 4            NA            NA           NA            NA            NA
## 5            NA            NA           NA            NA            NA
## 6            NA            NA           NA            NA            NA
##   OfficAssgnDrugUnits NumKindsDrugsSeiz PolicAveOTWorked LandArea PopDens
## 1                0.02              0.57             0.29     0.12    0.26
## 2                  NA                NA               NA     0.02    0.12
## 3                  NA                NA               NA     0.01    0.21
## 4                  NA                NA               NA     0.02    0.39
## 5                  NA                NA               NA     0.04    0.09
## 6                  NA                NA               NA     0.01    0.58
##   PctUsePubTrans PolicCars PolicOperBudg LemasPctPolicOnPatr
## 1           0.20      0.06          0.04                 0.9
## 2           0.45        NA            NA                  NA
## 3           0.02        NA            NA                  NA
## 4           0.28        NA            NA                  NA
## 5           0.02        NA            NA                  NA
## 6           0.10        NA            NA                  NA
##   LemasGangUnitDeploy LemasPctOfficDrugUn PolicBudgPerPop ViolentCrimesPerPop
## 1                 0.5                0.32            0.14                0.20
```

```
## 2                    NA               0.00               NA               0.67
## 3                    NA               0.00               NA               0.43
## 4                    NA               0.00               NA               0.12
## 5                    NA               0.00               NA               0.03
## 6                    NA               0.00               NA               0.14
```

```
dim(df)
```

```
## [1] 1994  123
```

The data now has 1994 observations and 123 variables.

1(b) Take a look at the missing percentage of each remaining variable. Remove those with heavy missing, say, over 60%. . . . .

```
#Removing remaining variables with heavy percentages (>60%)
missingrate <- data.frame()
nr <- NROW(df)
nc <- NCOL(df)
Var_name <- variable.names(df)
for (i in 1:nc) {
  na <- sum(is.na(df[,i]))
  na_rate <- (na/nr)*100
  result <- list(Number_Missing = na, Missing_Rate = na_rate,
                 Variable = Var_name[i])
  missingrate <- rbind(missingrate, result, stringsAsFactors = F)
}
tail(missingrate)
```

```
##     Number_Missing Missing_Rate            Variable
## 118           1675     84.00201       PolicOperBudg
## 119           1675     84.00201 LemasPctPolicOnPatr
## 120           1675     84.00201 LemasGangUnitDeploy
## 121              0      0.00000 LemasPctOfficDrugUn
## 122           1675     84.00201      PolicBudgPerPop
## 123              0      0.00000 ViolentCrimesPerPop
```

```
# removing variables with  missing percentages > 60%
removemissing <- missingrate$Missing_Rate > 60.0
removemissing1 <- missingrate$Missing_Rate == 0.0
No.missing <- names(df[, removemissing1])
sum(!removemissing)
```

```
## [1] 101
```

```
df <- df[, !removemissing]
dim(df)
```

```
## [1] 1994  101
```

From the above, we have 101 remaining variables with missing percentage less than 60%.

1(c) Impute or replace the remaining missing values appropriately. . . . .

```
#Missing  value imputation
set.seed(125)
# number of variables left
print("Number of Variables")
```

```
## [1] "Number of Variables"
```

```
length(names(df))
```

```
## [1] 101
```

```
# variables with no missing values
print("List of Variables with no missing values")
```

```
## [1] "List of Variables with no missing values"
```

```
head(No.missing)
```

```
## [1] "population"    "householdsize" "racepctblack"  "racePctWhite"
## [5] "racePctAsian"  "racePctHisp"
```

```
print("Total number of Variables with no missing values")
```

```
## [1] "Total number of Variables with no missing values"
```

```
length(No.missing)
```

```
## [1] 100
```

```
suppressPackageStartupMessages(library(mice))
df_imputed <- mice(df, printFlag = F)
```

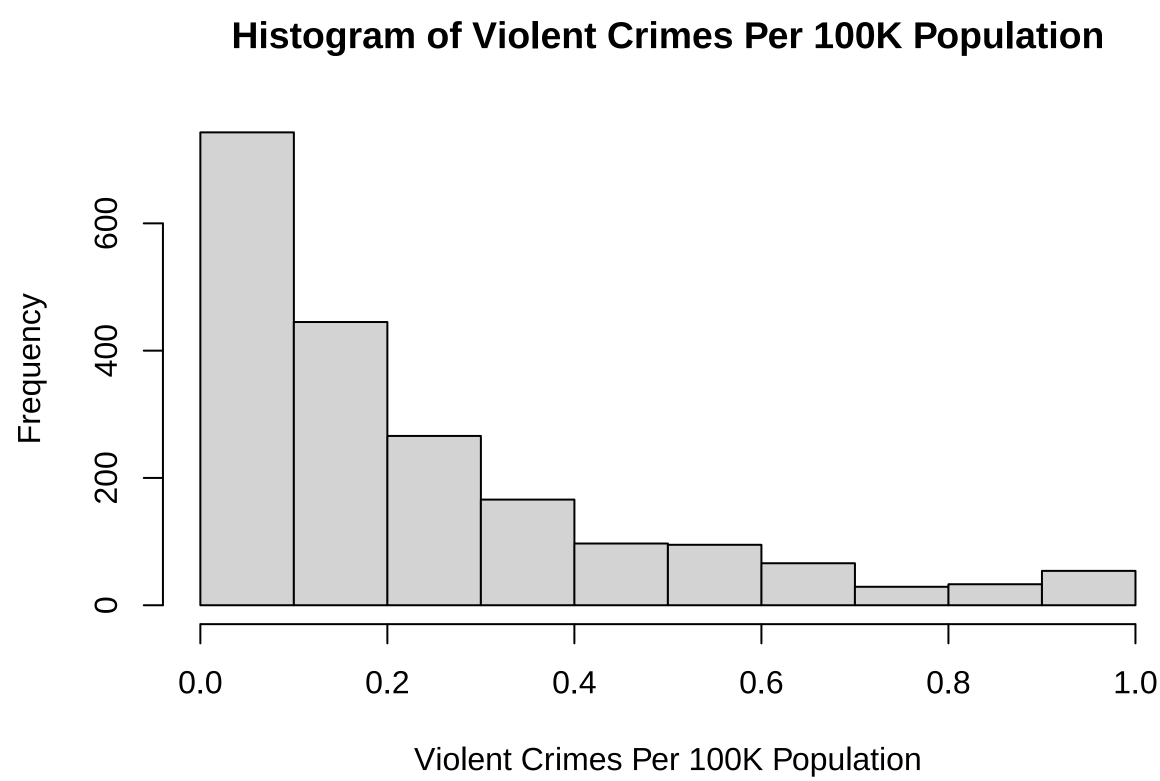
```
## Warning: Number of logged events: 25
```

```
data <- complete(df_imputed, 1)
data <- as.data.frame(data)
rm(df_imputed)
```

After imputing all the missing values. It is observed that out of the 101 variables remaining, 100 of them have no missing values.

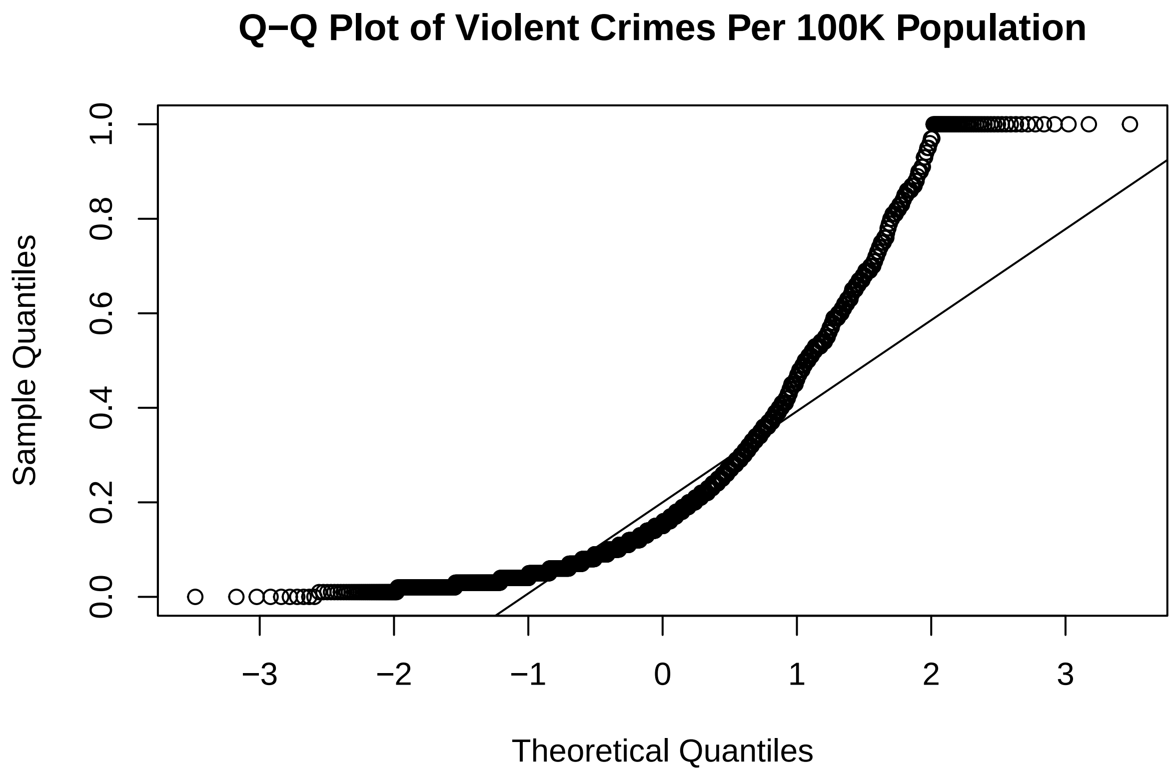
1(d) Conduct some EDA, which could be involved. In particular, check the distribution of the target variable ViolentCrimesPerPop. . . . .

```r
# histogram of Violent Crimes Per 100K Population
hist(data$ViolentCrimesPerPop, xlab = "Violent Crimes Per 100K Population",
     main = "Histogram of Violent Crimes Per 100K Population")
```



We can observe from the above histogram that ViolentCrimesPerPop is positively skewed which is a violation of the normality assumption.

```r
# Q-Q plot of Violent Crimes Per 100K Population
qqnorm(data$ViolentCrimesPerPop, main = "Q-Q Plot of Violent Crimes Per 100K Population")
qqline(data$ViolentCrimesPerPop)
```



We can observe from the above that there is a clear deviation of the plot from the line in the graph on the Q-Q plot above which confirms the violation of the normality assumption by the histogram.

```
# THE SHAPIRO-WILKS NORMALITY TEST: A LARGE P-VALUE WOULD JUSTIFY NORMALITY
shapiro.test(data$ViolentCrimesPerPop)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$ViolentCrimesPerPop
## W = 0.82162, p-value < 2.2e-16
```

We can also see from the Shairo-Wilk normality test that the p-value is very small which also confirms violation of the normality assumption.

(2)Partitioning Data: Randomly partition your data into two sets: the training set D1 and the test set D2 with a ratio of 2:1. In order for your results to be reproducible, report the random seed that you use in the partitioning.

. . . .

```
#Training-Test Split.
set.seed(500)
sampleData <- sample(nrow(data), (2.0/3.0)*nrow(data), replace = FALSE)
# training set
```

```r
D1 <- data[sampleData, ]
# test set
D2 <- data[-sampleData, ]
yobs <- D2$y
dim(D1)
```

```
## [1] 1329  101
```

```r
dim(D2)
```

```
## [1] 665 101
```

We can observed that the training data has 1329 observations and the test data set has 665 observations with 101 variables each.

(3)Linear Regression: First consider the ordinary linear regression model with variable selection or sparse estimation. Select any two methods out of the following options to determine two final linear models.

....

```r
# full model
set.seed(500)
formula0 <- ViolentCrimesPerPop ~ . -1
y <- D1[, all.vars(formula0)[1]]
X <- as.data.frame(model.matrix(as.formula(formula0),D1))
y.t <- D2[, all.vars(formula0)[1]]
X.t <- as.data.frame(model.matrix(as.formula(formula0),D2))
DAT <- data.frame(cbind(ViolentCrimesPerPop=y.t, X.t))
# model fit
fit.full <- lm(ViolentCrimesPerPop ~ ., data = D1)

summary(fit.full)
```

```
##
## Call:
## lm(formula = ViolentCrimesPerPop ~ ., data = D1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46685 -0.07221 -0.01217  0.05471  0.73648
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.262868   0.253066   1.039 0.299133
## population             0.587509   0.495758   1.185 0.236218
## householdsize         -0.049603   0.105328  -0.471 0.637767
## racepctblack           0.238816   0.061184   3.903 0.000100 ***
## racePctWhite           0.029017   0.070230   0.413 0.679550
## racePctAsian           0.003730   0.042081   0.089 0.929375
## racePctHisp            0.081799   0.063894   1.280 0.200705
## agePct12t21            0.061489   0.132450   0.464 0.642553
## agePct12t29           -0.317932   0.192514  -1.651 0.098897 .
## agePct16t24           -0.004085   0.204430  -0.020 0.984060
```

```
## agePct65up            0.001145   0.130642   0.009 0.993007
## numbUrban            -0.661970   0.483045  -1.370 0.170810
## pctUrban              0.068075   0.019619   3.470 0.000539 ***
## medIncome            -0.008697   0.210805  -0.041 0.967097
## pctWWage             -0.292768   0.110061  -2.660 0.007915 **
## pctWFarmSelf          0.017021   0.024876   0.684 0.493949
## pctWInvInc           -0.207479   0.081031  -2.560 0.010571 *
## pctWSocSec            0.075962   0.133327   0.570 0.568958
## pctWPubAsst           0.084765   0.057749   1.468 0.142412
## pctWRetire           -0.104394   0.045479  -2.295 0.021877 *
## medFamInc             0.053503   0.192646   0.278 0.781269
## perCapInc             0.078388   0.228531   0.343 0.731651
## whitePerCap          -0.234713   0.182783  -1.284 0.199345
## blackPerCap          -0.022290   0.031134  -0.716 0.474158
## indianPerCap         -0.055624   0.024058  -2.312 0.020937 *
## AsianPerCap           0.016124   0.023359   0.690 0.490168
## OtherPerCap           0.049333   0.023678   2.084 0.037408 *
## HispPerCap            0.023633   0.031914   0.741 0.459118
## NumUnderPov          -0.076968   0.184109  -0.418 0.675979
## PctPopUnderPov       -0.105738   0.078514  -1.347 0.178312
## PctLess9thGrade      -0.146718   0.085076  -1.725 0.084858 .
## PctNotHSGrad          0.139937   0.120103   1.165 0.244189
## PctBSorMore           0.142414   0.095746   1.487 0.137163
## PctUnemployed        -0.011535   0.051198  -0.225 0.821774
## PctEmploy             0.236323   0.098899   2.390 0.017019 *
## PctEmplManu          -0.053910   0.040474  -1.332 0.183116
## PctEmplProfServ      -0.010902   0.048893  -0.223 0.823596
## PctOccupManu          0.010800   0.069047   0.156 0.875736
## PctOccupMgmtProf      0.006677   0.106471   0.063 0.950009
## MalePctDivorce        0.577857   0.320553   1.803 0.071682 .
## MalePctNevMarr        0.226542   0.084493   2.681 0.007434 **
## FemalePctDiv          0.218865   0.419085   0.522 0.601594
## TotalPctDiv          -0.683573   0.690187  -0.990 0.322166
## PersPerFam           -0.017436   0.209902  -0.083 0.933813
## PctFam2Par            0.090018   0.198236   0.454 0.649840
## PctKids2Par          -0.298518   0.189971  -1.571 0.116350
## PctYoungKids2Par      0.043751   0.062124   0.704 0.481415
## PctTeen2Par          -0.006702   0.052822  -0.127 0.899053
## PctWorkMomYoungKids   0.116240   0.058016   2.004 0.045335 *
## PctWorkMom           -0.181688   0.066062  -2.750 0.006042 **
## NumIlleg             -0.076876   0.141569  -0.543 0.587209
## PctIlleg              0.142772   0.057962   2.463 0.013907 *
## NumImmig             -0.163155   0.107276  -1.521 0.128545
## PctImmigRecent        0.104415   0.050273   2.077 0.038012 *
## PctImmigRec5         -0.104160   0.082765  -1.259 0.208445
## PctImmigRec8         -0.054657   0.101549  -0.538 0.590513
## PctImmigRec10         0.082127   0.078979   1.040 0.298613
## PctRecentImmig       -0.122317   0.159089  -0.769 0.442126
## PctRecImmig5          0.030196   0.286462   0.105 0.916068
## PctRecImmig8          0.376541   0.339197   1.110 0.267176
## PctRecImmig10        -0.313283   0.271026  -1.156 0.247940
## PctSpeakEnglOnly      0.083217   0.082256   1.012 0.311887
## PctNotSpeakEnglWell  -0.186900   0.083438  -2.240 0.025270 *
## PctLargHouseFam       0.023245   0.278703   0.083 0.933543
```

```
## PctLargHouseOccup     -0.230272   0.294293  -0.782 0.434095
## PersPerOccupHous       0.584374   0.309599   1.888 0.059326 .
## PersPerOwnOccHous      0.008591   0.202348   0.042 0.966142
## PersPerRentOccHous    -0.233156   0.098552  -2.366 0.018145 *
## PctPersOwnOccup       -0.727426   0.433373  -1.679 0.093500 .
## PctPersDenseHous       0.233631   0.092893   2.515 0.012029 *
## PctHousLess3BR         0.138525   0.073984   1.872 0.061393 .
## MedNumBR               0.016690   0.024483   0.682 0.495560
## HousVacant             0.221109   0.093052   2.376 0.017645 *
## PctHousOccup          -0.030257   0.039002  -0.776 0.438032
## PctHousOwnOcc          0.629775   0.454897   1.384 0.166477
## PctVacantBoarded       0.054712   0.026399   2.072 0.038429 *
## PctVacMore6Mos        -0.086419   0.030846  -2.802 0.005165 **
## MedYrHousBuilt        -0.044172   0.036362  -1.215 0.224680
## PctHousNoPhone        -0.005570   0.044092  -0.126 0.899493
## PctWOFullPlumb        -0.017863   0.025510  -0.700 0.483926
## OwnOccLowQuart        -0.547895   0.254437  -2.153 0.031485 *
## OwnOccMedVal           0.597884   0.384729   1.554 0.120433
## OwnOccHiQuart         -0.098653   0.208733  -0.473 0.636562
## RentLowQ              -0.233688   0.082628  -2.828 0.004757 **
## RentMedian            -0.150252   0.201733  -0.745 0.456529
## RentHighQ             -0.078658   0.104033  -0.756 0.449744
## MedRent                0.489481   0.166620   2.938 0.003368 **
## MedRentPctHousInc      0.025239   0.040220   0.628 0.530422
## MedOwnCostPctInc      -0.094980   0.042259  -2.248 0.024782 *
## MedOwnCostPctIncNoMtg -0.074525   0.030713  -2.427 0.015388 *
## NumInShelters          0.157441   0.077646   2.028 0.042809 *
## NumStreet              0.201086   0.060011   3.351 0.000830 ***
## PctForeignBorn         0.243327   0.116595   2.087 0.037099 *
## PctBornSameState      -0.011756   0.053325  -0.220 0.825547
## PctSameHouse85        -0.014809   0.068773  -0.215 0.829541
## PctSameCity85          0.011679   0.045894   0.254 0.799171
## PctSameState85         0.029637   0.053362   0.555 0.578726
## LandArea              -0.008561   0.060020  -0.143 0.886601
## PopDens               -0.026318   0.037515  -0.702 0.483098
## PctUsePubTrans        -0.058930   0.028509  -2.067 0.038938 *
## LemasPctOfficDrugUn    0.032200   0.019119   1.684 0.092398 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1321 on 1228 degrees of freedom
## Multiple R-squared:  0.7123, Adjusted R-squared:  0.6889
## F-statistic: 30.41 on 100 and 1228 DF,  p-value: < 2.2e-16
```

From the above, the full model is significant since we have a p-value less than 0.05, F-statistic 30.4 and adjusted R-squared 68.9%

```
BIC(fit.full)
```

```
## [1] -980.9702
```

```
AIC(fit.full, k=2)
```

```
## [1] -1510.573
```

The values of BIC and AIC are -981 and -1511 respectively

3(a) Report the fitting results and identify variables that are important . . . .

Method one:LASSO

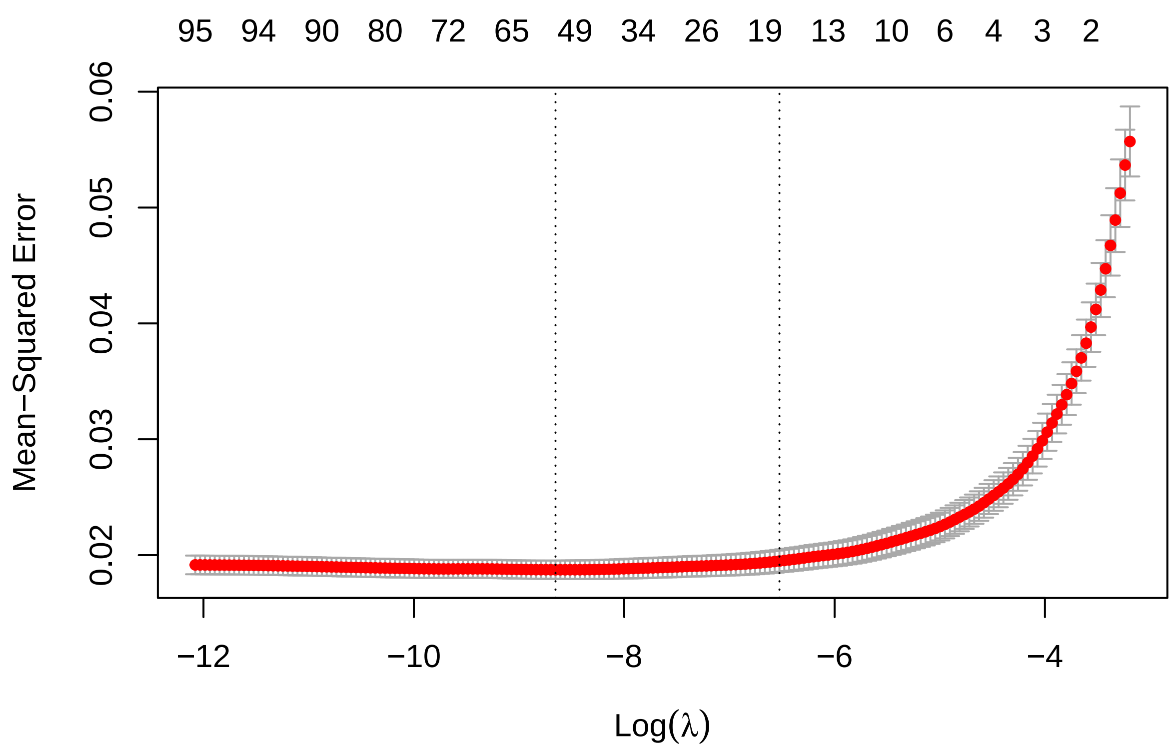
. . . .

```
set.seed(500)
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-3
```

```
cv.LASSO <- cv.glmnet(x=as.matrix(X), y=y, family="gaussian", alpha=1, nlambda=200,
                      standardize=F)
plot(cv.LASSO)
```



After performing LASSO, two models are found to be the best; one with 54 variables and the other with 18 variables however, we will choose the model with 18 variables.

```
beta.hat.lasso <- coef(cv.LASSO, s="lambda.1se")
cutoff <- 0
terms2 <- names(X)[abs(as.vector(beta.hat.lasso[-1])) > cutoff]
formula.LASSO <- as.formula(paste(c("ViolentCrimesPerPop ~ ", terms2),
                              collapse=" + "))
fit.LASSO <- lm(formula.LASSO, data = D1)
summary(fit.LASSO)
```

```
##
## Call:
## lm(formula = formula.LASSO, data = D1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45029 -0.07505 -0.01504  0.05207  0.79077
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.09106    0.08775   1.038 0.299629
## racepctblack          0.26259    0.04855   5.409 7.55e-08 ***
## racePctWhite          0.05201    0.05491   0.947 0.343729
## pctUrban              0.04807    0.01019   4.718 2.63e-06 ***
## pctWWage             -0.05464    0.03407  -1.604 0.109039
## pctWPubAsst           0.06320    0.04020   1.572 0.116144
## MalePctDivorce        0.19663    0.03390   5.800 8.30e-09 ***
## PctKids2Par          -0.11360    0.06863  -1.655 0.098109 .
## PctWorkMom           -0.05098    0.02716  -1.877 0.060704 .
## PctIlleg              0.21020    0.04743   4.432 1.01e-05 ***
## PctRecImmig10        -0.01637    0.06459  -0.253 0.799923
## PctPersDenseHous      0.13417    0.04326   3.101 0.001968 **
## HousVacant            0.07472    0.03827   1.953 0.051079 .
## PctHousOccup         -0.06079    0.02591  -2.346 0.019104 *
## PctVacantBoarded      0.04968    0.02253   2.205 0.027638 *
## MedOwnCostPctIncNoMtg -0.07543    0.02278  -3.312 0.000952 ***
## NumStreet             0.19137    0.04782   4.002 6.65e-05 ***
## PctForeignBorn        0.10880    0.06613   1.645 0.100126
## LemasPctOfficDrugUn   0.02248    0.01857   1.210 0.226309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1361 on 1310 degrees of freedom
## Multiple R-squared:  0.674,  Adjusted R-squared:  0.6695
## F-statistic: 150.5 on 18 and 1310 DF,  p-value: < 2.2e-16
```

After applying LASSO to the training data set (D1), the variables;racepctblack, pctUrban, MalePctDivorce, PctIlleg, PctPersDenseHous, PctHousOccup, PctVacantBoarded, MedOwnCostPctIncNoMtg and NumStreet are significant with Adjusted R-Squared of 67%

```
library(DAAG)
```

```
## Loading required package: lattice
```

```r
CV <- CVlm(data=DAT, form.lm=fit.LASSO, m=10, seed=500, plotit=F, printit=T)
```

```
## Analysis of Variance Table
##
## Response: ViolentCrimesPerPop
##                        Df Sum Sq Mean Sq F value  Pr(>F)
## racepctblack            1  12.19   12.19  675.82 < 2e-16 ***
## racePctWhite            1   3.75    3.75  207.99 < 2e-16 ***
## pctUrban                1   0.00    0.00    0.08 0.78055
## pctWWage                1   1.77    1.77   98.22 < 2e-16 ***
## pctWPubAsst             1   0.51    0.51   28.07 1.6e-07 ***
## MalePctDivorce          1   1.24    1.24   68.92 6.0e-16 ***
## PctKids2Par             1   0.89    0.89   49.07 6.2e-12 ***
## PctWorkMom              1   0.40    0.40   22.26 2.9e-06 ***
## PctIlleg                1   0.22    0.22   12.02 0.00056 ***
## PctRecImmig10           1   0.01    0.01    0.28 0.59616
## PctPersDenseHous        1   0.18    0.18    9.81 0.00181 **
## HousVacant              1   0.69    0.69   38.13 1.2e-09 ***
## PctHousOccup            1   0.05    0.05    2.88 0.09034 .
## PctVacantBoarded        1   0.00    0.00    0.00 0.98504
## MedOwnCostPctIncNoMtg   1   0.03    0.03    1.68 0.19519
## NumStreet               1   0.07    0.07    3.78 0.05215 .
## PctForeignBorn          1   0.04    0.04    2.26 0.13349
## LemasPctOfficDrugUn     1   0.00    0.00    0.11 0.73744
## Residuals             646  11.65    0.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 66
##                         28     51    119   139    141   158    179     199
## Predicted            0.112  0.296  0.164 0.258  0.092 0.575  0.204  0.3221
## cvpred               0.112  0.315  0.176 0.252  0.101 0.560  0.211  0.3581
## ViolentCrimesPerPop  0.100  0.200  0.060 0.380  0.080 0.730  0.100  0.2600
## CV residual         -0.012 -0.115 -0.116 0.128 -0.021 0.170 -0.111 -0.0981
##                         288    307     319    328     353    359     365
## Predicted           -0.0455  0.383  0.1142 0.0927 0.07300  0.169  0.1442
## cvpred              -0.0507  0.381  0.1177 0.0991 0.07661  0.182  0.1496
## ViolentCrimesPerPop  0.0100  0.230  0.0400 0.1700 0.08000  0.030  0.0600
## CV residual          0.0607 -0.151 -0.0777 0.0709 0.00339 -0.152 -0.0896
##                         384     415     435   458    459    570     645    651
## Predicted            0.1585  0.1828  0.1157 0.695  0.203 0.2384 -0.0272 0.1127
## cvpred               0.1688  0.1882  0.1172 0.628  0.205 0.2592 -0.0350 0.1164
## ViolentCrimesPerPop  0.1500  0.1000  0.0400 1.000  0.070 0.3300  0.0500 0.1500
## CV residual         -0.0188 -0.0882 -0.0772 0.372 -0.135 0.0708  0.0850 0.0336
##                        667    709     739    768     769     813    891
## Predicted           0.1499 0.3283  0.2858  0.336  0.1377  0.2137 0.1078
## cvpred              0.1413 0.3261  0.2962  0.366  0.1411  0.2268 0.1027
## ViolentCrimesPerPop 0.2200 0.3600  0.2800  0.230  0.1000  0.1300 0.2000
## CV residual         0.0787 0.0339 -0.0162 -0.136 -0.0411 -0.0968 0.0973
##                          905    940     989   999   1065    1101   1104   1185
## Predicted           0.089993  0.1095 0.00306 0.619  0.380 0.00621  0.360 0.0365
```

```
## cvpred              0.089352  0.1157 0.00220 0.575  0.393 0.01065  0.381 0.0344
## ViolentCrimesPerPop 0.090000  0.0600 0.02000 1.000  0.260 0.03000  0.190 0.0800
## CV residual         0.000648 -0.0557 0.01780 0.425 -0.133 0.01935 -0.191 0.0456
##                       1224   1302   1339   1391   1401    1405    1438  1453
## Predicted            0.307 0.3068 0.2789 0.0871 0.0112  0.3200  0.1707 0.192
## cvpred               0.311 0.3067 0.2897 0.0909 0.0101  0.3055  0.1622 0.188
## ViolentCrimesPerPop  0.190 0.4000 0.3000 0.1800 0.0400  0.2700  0.1000 0.340
## CV residual         -0.121 0.0933 0.0103 0.0891 0.0299 -0.0355 -0.0622 0.152
##                        1499   1549  1568  1571    1584    1594   1639    1691
## Predicted            0.1046 0.0123 0.347 0.240  0.0825  0.0371  0.372  0.4203
## cvpred               0.1084 0.0122 0.340 0.233  0.0843  0.0384  0.417  0.4085
## ViolentCrimesPerPop  0.0600 0.0800 0.500 0.530  0.0700  0.0200  0.170  0.3900
## CV residual         -0.0484 0.0678 0.160 0.297 -0.0143 -0.0184 -0.247 -0.0185
##                       1721   1775  1776    1803    1804   1811    1862    1872
## Predicted            0.340  0.250 0.470  0.1078  0.2836 0.5953  0.2162 0.00999
## cvpred               0.337  0.283 0.466  0.1044  0.2958 0.5599  0.2294 0.01382
## ViolentCrimesPerPop  0.110  0.150 0.830  0.0700  0.2700 0.6200  0.1700 0.11000
## CV residual         -0.227 -0.133 0.364 -0.0344 -0.0258 0.0601 -0.0594 0.09618
##                        1929    1936    1974  1988
## Predicted            0.2592  0.0571  0.0501 0.028
## cvpred               0.2683  0.0571  0.0504 0.022
## ViolentCrimesPerPop  0.1700  0.0400  0.0400 0.040
## CV residual         -0.0983 -0.0171 -0.0104 0.018
##
## Sum of squares = 1.11    Mean square = 0.02    n = 66
##
## fold 2
## Observations in test set: 67
##                          4      6      23    57    67     89     90     94
## Predicted            0.331  0.242  0.2356 0.194 0.387 0.1944  0.178 0.0829
## cvpred               0.357  0.247  0.2293 0.201 0.389 0.1792  0.173 0.0815
## ViolentCrimesPerPop  0.120  0.140  0.2100 0.220 0.610 0.2700  0.150 0.1200
## CV residual         -0.237 -0.107 -0.0193 0.019 0.221 0.0908 -0.023 0.0385
##                        134    180     194    211    230     245     265   270
## Predicted           0.2035 0.1671 0.06416  0.387  0.570  0.2180 0.04993 0.174
## cvpred              0.2032 0.1689 0.06326  0.399  0.589  0.2208 0.04353 0.162
## ViolentCrimesPerPop 0.2700 0.2500 0.07000  0.290  0.420  0.1800 0.05000 0.690
## CV residual         0.0668 0.0811 0.00674 -0.109 -0.169 -0.0408 0.00647 0.528
##                        272     285   324    325   385    420    465    528
## Predicted           0.2111  0.2457  0.41 0.2502 0.508  0.216 0.3129  0.184
## cvpred              0.2124  0.2478  0.43 0.2398 0.523  0.215 0.3105  0.177
## ViolentCrimesPerPop 0.2400  0.1900  0.27 0.3000 0.540  0.100 0.3300  0.070
## CV residual         0.0276 -0.0578 -0.16 0.0602 0.017 -0.115 0.0195 -0.107
##                        603    618    693    764    886      911   925     943
## Predicted           0.1825 0.1824 0.5203  0.311  0.331  0.04478 0.257 0.14555
## cvpred              0.1728 0.1682 0.5137  0.302  0.326  0.04828 0.244 0.14816
## ViolentCrimesPerPop 0.2300 0.1800 0.6100  0.190  0.190  0.04000 0.390 0.15000
## CV residual         0.0572 0.0118 0.0963 -0.112 -0.136 -0.00828 0.146 0.00184
##                        955    983    991    1028  1040    1081   1085   1106
## Predicted           0.1987  0.229  0.148 0.00406 0.382  0.1245 0.0136  0.224
## cvpred              0.2149  0.223  0.144 0.00436 0.388  0.1252 0.0132  0.216
## ViolentCrimesPerPop 0.2400  0.210  0.120 0.01000 0.570  0.0300 0.0700  0.090
## CV residual         0.0251 -0.013 -0.024 0.00564 0.182 -0.0952 0.0568 -0.126
##                       1131      1175   1290   1362   1384    1387   1429
```

```
## Predicted            0.1852 -0.000145 0.1627  0.281  0.268  0.2336 0.4325
## cvpred               0.1864 -0.000059 0.1697  0.281  0.267  0.2378 0.4347
## ViolentCrimesPerPop  0.0900  0.000000 0.1900  0.120  0.160  0.1500 0.5100
## CV residual         -0.0964  0.000059 0.0203 -0.161 -0.107 -0.0878 0.0753
##                       1430      1431  1432    1479   1531    1542   1543  1583
## Predicted            0.0166 0.062975 0.332  0.0750 0.0592  0.342  0.309 0.468
## cvpred               0.0219 0.059726 0.333  0.0801 0.0567  0.351  0.321 0.421
## ViolentCrimesPerPop  0.0200 0.060000 0.510  0.0600 0.1300  0.210  0.220 0.660
## CV residual         -0.0019 0.000274 0.177 -0.0201 0.0733 -0.141 -0.101 0.239
##                       1600    1661    1686   1719  1748   1759  1814    1892
## Predicted           0.07839  0.0831  0.1518  0.460 0.327 0.0329 0.320  0.2643
## cvpred              0.07403  0.0860  0.1532  0.467 0.320 0.0314 0.313  0.2644
## ViolentCrimesPerPop 0.08000  0.0600  0.0900  0.210 0.500 0.2200 0.490  0.2300
## CV residual         0.00597 -0.0260 -0.0632 -0.257 0.180 0.1886 0.177 -0.0344
##                         1931   1942   1970  1971
## Predicted           -0.001785  0.257 0.6778 0.262
## cvpred              -0.000546  0.266 0.6597 0.263
## ViolentCrimesPerPop  0.020000  0.140 0.7500 0.530
## CV residual          0.020546 -0.126 0.0903 0.267
##
## Sum of squares = 1.1    Mean square = 0.02    n = 67
##
## fold 3
## Observations in test set: 67
##                         40    52      56     85   150    192    210     233
## Predicted            0.1712 0.372  0.2929  0.153 0.387 0.1267 0.1214 0.19045
## cvpred               0.1683 0.366  0.2927  0.158 0.384 0.1246 0.1207 0.19086
## ViolentCrimesPerPop  0.0700 0.680  0.2800  0.030 1.000 0.2100 0.1500 0.20000
## CV residual         -0.0983 0.314 -0.0127 -0.128 0.616 0.0854 0.0293 0.00914
##                        248    277     370     418   434   455    491   503
## Predicted            0.2967  0.525  0.1008  0.1743  0.56 0.478 0.0063 0.049
## cvpred               0.3102  0.531  0.1022  0.1696  0.57 0.470 0.0047 0.049
## ViolentCrimesPerPop  0.2200  0.340  0.0700  0.1300  0.18 0.640 0.0100 0.120
## CV residual         -0.0902 -0.191 -0.0322 -0.0396 -0.39 0.170 0.0053 0.071
##                        514    542   559       653     666    683   724     731
## Predicted            0.2287  0.208 0.201 -0.000977  0.2851  0.880 0.486  0.1414
## cvpred               0.2259  0.212 0.203 -0.000175  0.2955  0.947 0.470  0.1416
## ViolentCrimesPerPop  0.2200  0.130 0.360  0.040000  0.2500  0.530 0.820  0.0600
## CV residual         -0.0059 -0.082 0.157  0.040175 -0.0455 -0.417 0.350 -0.0816
##                        756    760    781    784    796    798     814     819
## Predicted            0.2318 0.0687  0.449 0.5550 0.2130  0.197  0.1619 -0.0299
## cvpred               0.2351 0.0707  0.450 0.5418 0.2134  0.196  0.1692 -0.0319
## ViolentCrimesPerPop  0.1600 0.1900  0.220 0.6100 0.2500  0.080  0.1000  0.2200
## CV residual         -0.0751 0.1193 -0.230 0.0682 0.0366 -0.116 -0.0692  0.2519
##                       839    873     887      901    942     944    1073
## Predicted            0.123 0.0318  0.3504  0.09136  0.277  0.0589  0.1642
## cvpred               0.122 0.0314  0.3312  0.09488  0.277  0.0604  0.1614
## ViolentCrimesPerPop  0.020 0.1300  0.2600  0.09000  0.140  0.0400  0.1000
## CV residual         -0.102 0.0986 -0.0712 -0.00488 -0.137 -0.0204 -0.0614
##                        1090    1127  1135    1145    1153    1229   1281
## Predicted            0.02479  0.2566 0.919  0.2658  0.1421 0.02152  0.179
## cvpred               0.02899  0.2476 0.915  0.2677  0.1451 0.02094  0.179
## ViolentCrimesPerPop  0.02000  0.1700 1.000  0.2100  0.1200 0.03000  0.050
## CV residual         -0.00899 -0.0776 0.085 -0.0577 -0.0251 0.00906 -0.129
```

```
##                       1304     1329    1343    1364   1389   1418  1503    1574
## Predicted            0.397  0.07265  0.0691  0.0319  0.282 0.0239 0.380  0.3518
## cvpred               0.382  0.07509  0.0661  0.0358  0.282 0.0196 0.376  0.3568
## ViolentCrimesPerPop  0.310  0.07000  0.0200  0.0200  0.120 0.2100 0.760  0.3100
## CV residual         -0.072 -0.00509 -0.0461 -0.0158 -0.162 0.1904 0.384 -0.0468
##                       1655    1681     1783   1788   1790    1826      1869
## Predicted            0.197  0.1714  0.04719  0.353 0.4848  0.2166  0.032712
## cvpred               0.203  0.1765  0.04511  0.358 0.4633  0.2157  0.030257
## ViolentCrimesPerPop  0.080  0.1400  0.04000  0.250 0.5300  0.1800  0.030000
## CV residual         -0.123 -0.0365 -0.00511 -0.108 0.0667 -0.0357 -0.000257
##                       1874    1891   1904   1908    1927    1960
## Predicted            0.138  0.1022  0.320  0.603  0.1023  0.1277
## cvpred               0.141  0.0974  0.335  0.610  0.1027  0.1254
## ViolentCrimesPerPop  0.040  0.0500  0.220  0.410  0.0900  0.1000
## CV residual         -0.101 -0.0474 -0.115 -0.200 -0.0127 -0.0254
##
## Sum of squares = 1.65    Mean square = 0.02    n = 67
##
## fold 4
## Observations in test set: 67
##                          70    101   152    169    177    181    191    196
## Predicted            0.1340 0.3659 0.469 0.4177 0.0460  0.335  0.281 0.0179
## cvpred               0.1413 0.3584 0.465 0.4221 0.0413  0.339  0.289 0.0174
## ViolentCrimesPerPop  0.0500 0.3900 0.860 0.4600 0.0700  0.120  0.100 0.0300
## CV residual         -0.0913 0.0316 0.395 0.0379 0.0287 -0.219 -0.189 0.0126
##                         206     232    250    254    269     279    315    374
## Predicted            0.2143  0.0839  0.138  0.219 0.0369  0.1363 0.1250 0.0224
## cvpred               0.2009  0.0839  0.140  0.227 0.0451  0.1413 0.1304 0.0255
## ViolentCrimesPerPop  0.1600  0.0600  0.030  0.090 0.0500  0.0500 0.1800 0.0400
## CV residual         -0.0409 -0.0239 -0.110 -0.137 0.0049 -0.0913 0.0496 0.0145
##                         466      485    486     518     521     578    621
## Predicted            0.1800  0.02224  0.501  0.1454  0.1616 -0.0154  0.528
## cvpred               0.1857  0.02679  0.521  0.1505  0.1687 -0.0128  0.530
## ViolentCrimesPerPop  0.0900  0.02000  0.250  0.0900  0.1300  0.0400  0.330
## CV residual         -0.0957 -0.00679 -0.271 -0.0605 -0.0387  0.0528 -0.200
##                       623    656     659     671     685     700    751
## Predicted           0.0728 0.0744  0.1805  0.3201  0.2857  0.5602  0.417
## cvpred              0.0820 0.0743  0.1794  0.3172  0.2821  0.5709  0.437
## ViolentCrimesPerPop 0.1000 0.0900  0.1600  0.2700  0.2600  0.5400  0.120
## CV residual         0.0180 0.0157 -0.0194 -0.0472 -0.0221 -0.0309 -0.317
##                         879     895     898    914     937    973     994
## Predicted            0.1177  0.1567  0.2576  0.280  0.1641 0.1547  0.0735
## cvpred               0.1118  0.1628  0.2574  0.291  0.1709 0.1597  0.0797
## ViolentCrimesPerPop  0.0800  0.1100  0.2200  0.160  0.1600 0.2200  0.0500
## CV residual         -0.0318 -0.0528 -0.0374 -0.131 -0.0109 0.0603 -0.0297
##                       1033   1074    1079   1094  1097    1133    1206    1211
## Predicted            0.1054 0.4299  0.1876 0.6818 0.162  0.0803  0.0678  0.1031
## cvpred               0.1191 0.4462  0.1849 0.6899 0.162  0.0862  0.0703  0.1046
## ViolentCrimesPerPop  0.0400 0.5100  0.1700 0.7100 0.440  0.0500  0.0500  0.0900
## CV residual         -0.0791 0.0638 -0.0149 0.0201 0.278 -0.0362 -0.0203 -0.0146
##                       1218   1220   1252   1282   1316    1399   1497     1562
## Predicted            0.390  0.100 0.1542 0.0719 0.1210  0.2125 0.0871 0.076422
## cvpred               0.389  0.098 0.1512 0.0711 0.1256  0.2198 0.0942 0.079319
## ViolentCrimesPerPop  0.250  0.060 0.2200 0.0900 0.1500  0.1400 0.1100 0.080000
```

```
## CV residual         -0.139 -0.038 0.0688 0.0189 0.0244 -0.0798 0.0158 0.000681
##                      1564    1599   1633  1634   1784    1786  1812    1820
## Predicted           0.729  0.2758  0.398 0.644 0.4344 0.00971 0.266  0.2416
## cvpred              0.718  0.2776  0.416 0.649 0.4397 0.00809 0.263  0.2528
## ViolentCrimesPerPop 1.000  0.2500  0.300 0.800 0.5300 0.06000 0.530  0.2400
## CV residual         0.282 -0.0276 -0.116 0.151 0.0903 0.05191 0.267 -0.0128
##                       1836   1870   1880    1882   1933     1987
## Predicted            0.6817 0.1308  0.674  0.0708 0.0307  0.01846
## cvpred               0.6808 0.1314  0.683  0.0737 0.0359  0.02123
## ViolentCrimesPerPop  0.6500 0.1900  0.200  0.0600 0.0400  0.02000
## CV residual         -0.0308 0.0586 -0.483 -0.0137 0.0041 -0.00123
##
## Sum of squares = 1.12    Mean square = 0.02    n = 67
##
## fold 5
## Observations in test set: 67
##                        15    16    27      32    97    131     187    236
## Predicted            0.252 0.147 0.513  0.2617 0.261  0.573  0.1398  0.288
## cvpred               0.248 0.147 0.503  0.2528 0.253  0.579  0.1494  0.297
## ViolentCrimesPerPop  0.210 0.300 0.840  0.1900 0.450  0.230  0.1200  0.110
## CV residual         -0.038 0.153 0.337 -0.0628 0.197 -0.349 -0.0294 -0.187
##                        301     331    337     344      377     380   421
## Predicted            0.1400  0.1083  0.154  0.0547  0.02576  0.2895 0.532
## cvpred               0.1466  0.1068  0.173  0.0550  0.02923  0.2855 0.515
## ViolentCrimesPerPop  0.1100  0.0700  0.070  0.0400  0.02000  0.2300 1.000
## CV residual         -0.0366 -0.0368 -0.103 -0.0150 -0.00923 -0.0555 0.485
##                        452     513     520     541     580     593     602
## Predicted            0.0513  0.2192  0.0125  0.1950  0.2123  0.0322  0.4306
## cvpred               0.0626  0.2226  0.0219  0.1955  0.2118  0.0340  0.4293
## ViolentCrimesPerPop  0.0500  0.1700  0.0000  0.1500  0.1400  0.0100  0.3900
## CV residual         -0.0126 -0.0526 -0.0219 -0.0455 -0.0718 -0.0240 -0.0393
##                        620     622    682    703    733     759     766
## Predicted           0.14194  0.0832 0.3369  0.226 0.6435 0.33363  0.0608
## cvpred              0.12414  0.0909 0.3312  0.241 0.6251 0.32356  0.0701
## ViolentCrimesPerPop 0.13000  0.0600 0.3700  0.070 0.6600 0.33000  0.0500
## CV residual         0.00586 -0.0309 0.0388 -0.171 0.0349 0.00644 -0.0201
##                        805     883    918    930     931     935  1060    1098
## Predicted            0.3461  0.0909  0.704 0.2134  0.0308  0.0558 0.295  0.1279
## cvpred               0.3518  0.0825  0.691 0.1974  0.0327  0.0576 0.298  0.1166
## ViolentCrimesPerPop  0.2800  0.0600  0.530 0.2600  0.0300  0.0300 0.310  0.0900
## CV residual         -0.0718 -0.0225 -0.161 0.0626 -0.0027 -0.0276 0.012 -0.0266
##                      1117   1132  1183  1205    1265   1315     1346    1404
## Predicted           0.274  0.176 0.530 0.236  0.0770 0.0483  0.11989  0.1684
## cvpred              0.268  0.160 0.524 0.220  0.0794 0.0461  0.11846  0.1691
## ViolentCrimesPerPop 0.470  0.060 0.690 0.400  0.0400 0.1500  0.11000  0.1000
## CV residual         0.202 -0.100 0.166 0.180 -0.0394 0.1039 -0.00846 -0.0691
##                       1410   1489    1500   1528  1555   1588   1609   1656
## Predicted            0.1308  0.203  0.1493 0.2315 0.335  0.222  0.324 0.3598
## cvpred               0.1291  0.202  0.1554 0.2267 0.321  0.206  0.337 0.3691
## ViolentCrimesPerPop  0.1100  0.090  0.1200 0.2800 0.780  0.100  0.130 0.4200
## CV residual         -0.0191 -0.112 -0.0354 0.0533 0.459 -0.106 -0.207 0.0509
##                       1673  1685  1687    1692    1709   1730    1737    1771
## Predicted           0.03104 0.703 0.302  0.3388  0.1046 0.2520  0.1174  0.4336
## cvpred              0.03645 0.663 0.296  0.3332  0.1069 0.2553  0.1054  0.4359
```

```
## ViolentCrimesPerPop 0.04000 1.000 0.650  0.2900  0.0200 0.3100  0.0800  0.3600
## CV residual         0.00355 0.337 0.354 -0.0432 -0.0869 0.0547 -0.0254 -0.0759
##                        1817   1835      1865    1881    1905    1967
## Predicted            0.0570 0.0590 -0.002261  0.2635  0.1123  0.3154
## cvpred               0.0574 0.0661  0.019227  0.2617  0.1109  0.3165
## ViolentCrimesPerPop  0.0100 0.1500  0.020000  0.2200  0.0800  0.3000
## CV residual         -0.0474 0.0839  0.000773 -0.0417 -0.0309 -0.0165
##
## Sum of squares = 1.36    Mean square = 0.02    n = 67
##
## fold 6
## Observations in test set: 67
##                       175    184      189     220   252     355    375   405
## Predicted           0.419 0.0161 0.052165  0.1132 0.460  0.0573 0.4396 0.306
## cvpred              0.404 0.0117 0.059843  0.1177 0.452  0.0561 0.4469 0.294
## ViolentCrimesPerPop 0.730 0.0200 0.060000  0.0600 0.630  0.0300 0.5100 0.680
## CV residual         0.326 0.0083 0.000157 -0.0577 0.178 -0.0261 0.0631 0.386
##                       500    515   583    605     639     673    688    710
## Predicted           0.596 0.1736 0.179 0.5159  0.2349  0.0717 0.5929 0.0604
## cvpred              0.566 0.1716 0.165 0.4964  0.2289  0.0741 0.5795 0.0640
## ViolentCrimesPerPop 1.000 0.2100 0.270 0.5900  0.1500  0.0400 0.6200 0.1000
## CV residual         0.434 0.0384 0.105 0.0936 -0.0789 -0.0341 0.0405 0.0360
##                       742    748    775     817    831     885     900   903
## Predicted           0.120 0.0830 0.1605  0.1568  0.154  0.1799  0.3289 0.160
## cvpred              0.125 0.0749 0.1631  0.1636  0.163  0.1842  0.3181 0.164
## ViolentCrimesPerPop 0.230 0.1700 0.2400  0.1200  0.080  0.0900  0.2600 0.550
## CV residual         0.105 0.0951 0.0769 -0.0436 -0.083 -0.0942 -0.0581 0.386
##                         915    936     963     980    1000    1001    1029
## Predicted            0.0676 0.0355  0.0243  0.1540  0.0552  0.1008  0.2559
## cvpred               0.0720 0.0388  0.0272  0.1529  0.0585  0.1016  0.2556
## ViolentCrimesPerPop  0.0400 0.0700  0.0100  0.0700  0.0400  0.0600  0.2400
## CV residual         -0.0320 0.0312 -0.0172 -0.0829 -0.0185 -0.0416 -0.0156
##                        1031    1068    1083   1084    1111   1119   1167
## Predicted            0.2863  0.3326 0.00880 0.1728  0.1001 0.1176 0.2814
## cvpred               0.2904  0.3372 0.00949 0.1616  0.1004 0.1222 0.2834
## ViolentCrimesPerPop  0.2000  0.3100 0.05000 0.2400  0.0300 0.1700 0.3100
## CV residual         -0.0904 -0.0272 0.04051 0.0784 -0.0704 0.0478 0.0266
##                        1178   1186    1187   1238  1296    1309    1334    1336
## Predicted            0.0906 0.2460 0.00574 0.4403 0.600  0.1817 0.34195 0.04876
## cvpred               0.0952 0.2558 0.00300 0.4373 0.573  0.1853 0.33156 0.04608
## ViolentCrimesPerPop  0.0600 0.3100 0.08000 0.4600 1.000  0.1100 0.34000 0.05000
## CV residual         -0.0352 0.0542 0.07700 0.0227 0.427 -0.0753 0.00844 0.00392
##                        1341   1347  1358   1363   1371   1434   1441    1522
## Predicted            0.1638  0.161 0.158 0.2271  0.478 0.0239  0.279  0.0740
## cvpred               0.1636  0.157 0.141 0.1963  0.481 0.0245  0.290  0.0761
## ViolentCrimesPerPop  0.0900  0.030 0.270 0.2600  0.350 0.0500  0.130  0.0300
## CV residual         -0.0736 -0.127 0.129 0.0637 -0.131 0.0255 -0.160 -0.0461
##                       1626   1699  1802    1827     1839    1920    1946
## Predicted            0.576  0.854 0.546  0.0557  0.57112  0.3327  0.1652
## cvpred               0.590  0.873 0.521  0.0585  0.57646  0.3367  0.1661
## ViolentCrimesPerPop  0.430  0.590 0.720  0.0200  0.57000  0.2700  0.1500
## CV residual         -0.160 -0.283 0.199 -0.0385 -0.00646 -0.0667 -0.0161
##                        1964   1965  1966   1981     1982   1989
## Predicted            0.5347 0.5700 0.544 0.1949  0.08723  0.595
```

```
## cvpred               0.5169 0.5799 0.550 0.1872  0.07433  0.593
## ViolentCrimesPerPop  0.4500 0.6000 0.690 0.2800  0.07000  0.190
## CV residual         -0.0669 0.0201 0.140 0.0928 -0.00433 -0.403
##
## Sum of squares = 1.38    Mean square = 0.02    n = 67
##
## fold 7
## Observations in test set: 66
##                         8      20       47      50    64      71    100     114
## Predicted           0.412  0.1009  0.20852  0.1639 0.361  0.0586  0.203  0.2656
## cvpred              0.388  0.1021  0.20494  0.1635 0.346  0.0688  0.215  0.2706
## ViolentCrimesPerPop 0.550  0.0300  0.20000  0.1200 0.570  0.0400  0.090  0.2000
## CV residual         0.162 -0.0721 -0.00494 -0.0435 0.224 -0.0288 -0.125 -0.0706
##                       129    171   216     257    308   332   363    456
## Predicted           0.416  0.299 0.265  0.2556 0.3006 0.164 0.784 0.0720
## cvpred              0.387  0.332 0.252  0.2549 0.2993 0.164 0.764 0.0702
## ViolentCrimesPerPop 0.620  0.230 0.490  0.1900 0.3200 0.350 1.000 0.1000
## CV residual         0.233 -0.102 0.238 -0.0649 0.0207 0.186 0.236 0.0298
##                         469    522     534   558    612    657     672   674
## Predicted            0.1856 0.1206 0.19802  0.35  0.296 0.2016 0.06510 0.397
## cvpred               0.1821 0.1183 0.19202  0.37  0.295 0.1934 0.06462 0.388
## ViolentCrimesPerPop  0.1500 0.1700 0.20000  0.36  0.130 0.2800 0.07000 0.610
## CV residual         -0.0321 0.0517 0.00798 -0.01 -0.165 0.0866 0.00538 0.222
##                       694    697     701     747   752    763   804   811
## Predicted            0.793 0.0268  0.4702  0.0576  0.23 0.1290 0.330 0.646
## cvpred               0.829 0.0275  0.4622  0.0558  0.25 0.1302 0.318 0.637
## ViolentCrimesPerPop  0.350 0.0400  0.4500  0.0200  0.10 0.1600 0.500 1.000
## CV residual         -0.479 0.0125 -0.0122 -0.0358 -0.15 0.0298 0.182 0.363
##                         826    893   954     977    995    1044   1067   1103
## Predicted            0.0595 0.0159 0.565  0.7812  0.141  0.0901 0.0519 0.0882
## cvpred               0.0660 0.0162 0.561  0.7929  0.135  0.0901 0.0548 0.0935
## ViolentCrimesPerPop  0.0200 0.0300 0.730  0.7700  0.030  0.0800 0.0700 0.3200
## CV residual         -0.0460 0.0138 0.169 -0.0229 -0.105 -0.0101 0.0152 0.2265
##                      1129  1160     1197    1223    1242    1245    1259
## Predicted            0.310 0.805  0.65250  0.2247  0.0989  0.5954  0.2687
## cvpred               0.325 0.767  0.65734  0.2243  0.0974  0.6196  0.2647
## ViolentCrimesPerPop  0.200 1.000  0.65000  0.1600  0.0600  0.5200  0.2100
## CV residual         -0.125 0.233 -0.00734 -0.0643 -0.0374 -0.0996 -0.0547
##                        1266      1269      1273   1289   1427    1490    1501
## Predicted            0.3368  0.02404  0.12686  0.164 0.0960 0.00981  0.0778
## cvpred               0.3556  0.02738  0.12434  0.164 0.0997 0.00747  0.0857
## ViolentCrimesPerPop  0.3400  0.02000  0.12000  0.060 0.1100 0.01000  0.0400
## CV residual         -0.0156 -0.00738 -0.00434 -0.104 0.0103 0.00253 -0.0457
##                        1579   1581    1659   1662    1710   1761   1809  1822
## Predicted            0.03900  0.587  0.0859 0.5480 0.22527 0.0638  0.466 0.216
## cvpred               0.03774  0.587  0.0826 0.5525 0.21891 0.0735  0.471 0.214
## ViolentCrimesPerPop  0.03000  0.450  0.0500 0.5900 0.22000 0.1200  0.260 0.380
## CV residual         -0.00774 -0.137 -0.0326 0.0375 0.00109 0.0465 -0.211 0.166
##                      1842   1849   1852    1924
## Predicted            0.473 0.0982  0.320  0.1622
## cvpred               0.482 0.1034  0.348  0.1556
## ViolentCrimesPerPop  0.160 0.1400  0.220  0.1100
## CV residual         -0.322 0.0366 -0.128 -0.0456
##
```

```
## Sum of squares = 1.24    Mean square = 0.02    n = 66
##
## fold 8
## Observations in test set: 66
##                         10     55    127    128   149     160   215      264
## Predicted           0.1081  0.153 0.1032  0.265 0.473  0.0838 0.112  0.01511
## cvpred              0.1085  0.175 0.1038  0.289 0.472  0.0846 0.107  0.01229
## ViolentCrimesPerPop 0.1500  0.100 0.1700  0.140 0.630  0.0600 0.270  0.01000
## CV residual         0.0415 -0.075 0.0662 -0.149 0.158 -0.0246 0.163 -0.00229
##                         276     298      300    336    362     443    477
## Predicted            0.3813  0.0374  0.34814 0.0218  0.289  0.0981  0.328
## cvpred               0.3956  0.0382  0.36264 0.0191  0.292  0.1025  0.344
## ViolentCrimesPerPop  0.3600  0.0200  0.36000 0.0300  0.220  0.0400  0.140
## CV residual         -0.0356 -0.0182 -0.00264 0.0109 -0.072 -0.0625 -0.204
##                           504    507    587    594     610    630    665    698
## Predicted            0.039798  0.354  0.658 0.0465 0.00405 0.0605  0.213  0.207
## cvpred               0.040782  0.375  0.664 0.0421 0.00372 0.0578  0.209  0.214
## ViolentCrimesPerPop  0.040000  0.130  0.560 0.0600 0.01000 0.1700  0.090  0.090
## CV residual         -0.000782 -0.245 -0.104 0.0179 0.00628 0.1122 -0.119 -0.124
##                         715    774    830   897    913    917    933     965
## Predicted           0.21352  0.178 0.2446 0.168 0.0541 0.0626  0.174  0.0277
## cvpred              0.22551  0.191 0.2549 0.165 0.0496 0.0570  0.181  0.0358
## ViolentCrimesPerPop 0.23000  0.000 0.3500 0.240 0.1300 0.0900  0.110  0.0100
## CV residual         0.00449 -0.191 0.0951 0.075 0.0804 0.0330 -0.071 -0.0258
##                        976    1003   1050   1061   1087   1099  1130    1136
## Predicted           0.0691  0.2582  0.205  0.435  0.185 0.1851 0.193  0.2189
## cvpred              0.0750  0.2648  0.220  0.441  0.196 0.1796 0.188  0.2374
## ViolentCrimesPerPop 0.0900  0.2000  0.020  0.270  0.090 0.1900 0.380  0.2100
## CV residual         0.0150 -0.0648 -0.200 -0.171 -0.106 0.0104 0.192 -0.0274
##                        1151  1163    1176   1200    1202   1226    1253   1360
## Predicted            0.1123 0.151 0.21598  0.144  0.0984 0.1314  0.1083  0.278
## cvpred               0.1107 0.143 0.21832  0.149  0.1070 0.1358  0.1102  0.275
## ViolentCrimesPerPop  0.0800 0.290 0.22000  0.060  0.0200 0.1800  0.0600  0.160
## CV residual         -0.0307 0.147 0.00168 -0.089 -0.0870 0.0442 -0.0502 -0.115
##                         1409   1451  1460   1471   1504   1524   1536    1569
## Predicted           0.002575  0.543 0.249  0.342 0.1602  0.430  0.180  0.1443
## cvpred              0.000193  0.543 0.246  0.368 0.1397  0.454  0.185  0.1334
## ViolentCrimesPerPop 0.030000  0.440 0.370  0.130 0.1500  0.290  0.050  0.0600
## CV residual         0.029807 -0.103 0.124 -0.238 0.0103 -0.164 -0.135 -0.0734
##                        1601   1604   1612    1627     1653  1664   1720   1731
## Predicted            0.0727 0.1923 0.0311  0.3847 -0.00947 0.264 0.2173 0.1256
## cvpred               0.0694 0.1872 0.0289  0.3931 -0.02431 0.249 0.2123 0.1172
## ViolentCrimesPerPop  0.0400 0.2000 0.0400  0.3000  0.06000 0.750 0.2300 0.1500
## CV residual         -0.0294 0.0128 0.0111 -0.0931  0.08431 0.501 0.0177 0.0328
##                        1773    1923  1959
## Predicted            0.1621  0.3627 0.222
## cvpred               0.1671  0.3685 0.215
## ViolentCrimesPerPop  0.0900  0.3200 0.340
## CV residual         -0.0771 -0.0485 0.125
##
## Sum of squares = 0.92    Mean square = 0.01    n = 66
##
## fold 9
## Observations in test set: 66
```

```
##                         61    113    116   229    235     238    242    327
## Predicted            0.0702 0.0259 0.0262 0.509 0.0406  0.0390  0.239  0.116
## cvpred               0.0648 0.0294 0.0143 0.511 0.0454  0.0446  0.234  0.114
## ViolentCrimesPerPop  0.0300 0.0600 0.0400 0.630 0.1200  0.0200  0.120  0.070
## CV residual         -0.0348 0.0306 0.0257 0.119 0.0746 -0.0246 -0.114 -0.044
##                       387     397   400     410    419    437  526    596
## Predicted           0.2823  0.4303 0.608  0.3172 0.1075  0.177 0.43  0.248
## cvpred              0.2926  0.4688 0.582  0.3123 0.0998  0.172 0.41  0.244
## ViolentCrimesPerPop 0.3900  0.3900 0.800  0.2500 0.2200  0.060 0.66  0.110
## CV residual         0.0974 -0.0788 0.218 -0.0623 0.1202 -0.112 0.25 -0.134
##                       637       648     663    670   676   689     696    711
## Predicted            0.409  0.000412  0.3820 0.7539 0.558 0.304  0.7559 0.0137
## cvpred               0.460 -0.003539  0.3727 0.7258 0.549 0.304  0.7515 0.0169
## ViolentCrimesPerPop  0.200  0.030000  0.3500 0.8100 0.950 0.500  0.7200 0.0300
## CV residual         -0.260  0.033539 -0.0227 0.0842 0.401 0.196 -0.0315 0.0131
##                      712    767      833     854     876    938      972
## Predicted           0.354 0.3222  0.02525 0.01007  0.1163 0.0123 -0.00797
## cvpred              0.358 0.3252  0.02908 0.00881  0.1138 0.0144 -0.00976
## ViolentCrimesPerPop 0.620 0.3400  0.02000 0.01000  0.0600 0.0600  0.09000
## CV residual         0.262 0.0148 -0.00908 0.00119 -0.0538 0.0456  0.09976
##                       986    1030   1078    1138    1174  1181   1203    1254
## Predicted           0.0142  0.4697 0.1369  0.1263 0.00860 0.499  0.239  0.4183
## cvpred              0.0155  0.4798 0.1376  0.1193 0.00635 0.491  0.240  0.4247
## ViolentCrimesPerPop 0.0500  0.4300 0.1500  0.0400 0.02000 0.690  0.140  0.3900
## CV residual         0.0345 -0.0498 0.0124 -0.0793 0.01365 0.199 -0.100 -0.0347
##                       1284    1331    1348   1394   1436   1454     1486   1502
## Predicted           0.3547  0.0233 0.05544  0.402 0.1868  0.289 -0.00802 0.2168
## cvpred              0.3438  0.0219 0.05621  0.398 0.1816  0.287 -0.00391 0.2245
## ViolentCrimesPerPop 0.3600  0.0200 0.06000  0.290 0.2100  0.130  0.03000 0.2500
## CV residual         0.0162 -0.0019 0.00379 -0.108 0.0284 -0.157  0.03391 0.0255
##                        1506    1616    1622    1649    1651   1671   1697
## Predicted           0.22077  0.1071  0.1234  0.0830  0.1598  0.556 0.0418
## cvpred              0.22359  0.1001  0.1347  0.0756  0.1622  0.548 0.0378
## ViolentCrimesPerPop 0.23000  0.0600  0.0900  0.0500  0.0900  0.480 0.1300
## CV residual         0.00641 -0.0401 -0.0447 -0.0256 -0.0722 -0.068 0.0922
##                        1728    1734   1743    1757     1770    1777   1823
## Predicted           0.41143  0.1586  0.325  0.2999  0.14105 0.01820  0.653
## cvpred              0.41538  0.1558  0.345  0.2893  0.14953 0.02581  0.673
## ViolentCrimesPerPop 0.42000  0.1100  0.150  0.2700  0.14000 0.03000  0.230
## CV residual         0.00462 -0.0458 -0.195 -0.0193 -0.00953 0.00419 -0.443
##                      1824   1828    1854   1954   1993
## Predicted           0.189 0.0631  0.0970  0.154 0.1581
## cvpred              0.186 0.0625  0.0976  0.155 0.1564
## ViolentCrimesPerPop 0.410 0.0900  0.0500  0.040 0.1900
## CV residual         0.224 0.0275 -0.0476 -0.115 0.0336
##
## Sum of squares = 1    Mean square = 0.02    n = 66
##
## fold 10
## Observations in test set: 66
##                          3       7    29      68     92     122       123
## Predicted           0.3521  0.0562 0.349  0.1206  0.194  0.0656  0.000838
## cvpred              0.3543  0.0586 0.336  0.1173  0.191  0.0624 -0.000167
## ViolentCrimesPerPop 0.4300  0.0300 0.490  0.0300  0.050  0.0500  0.030000
```

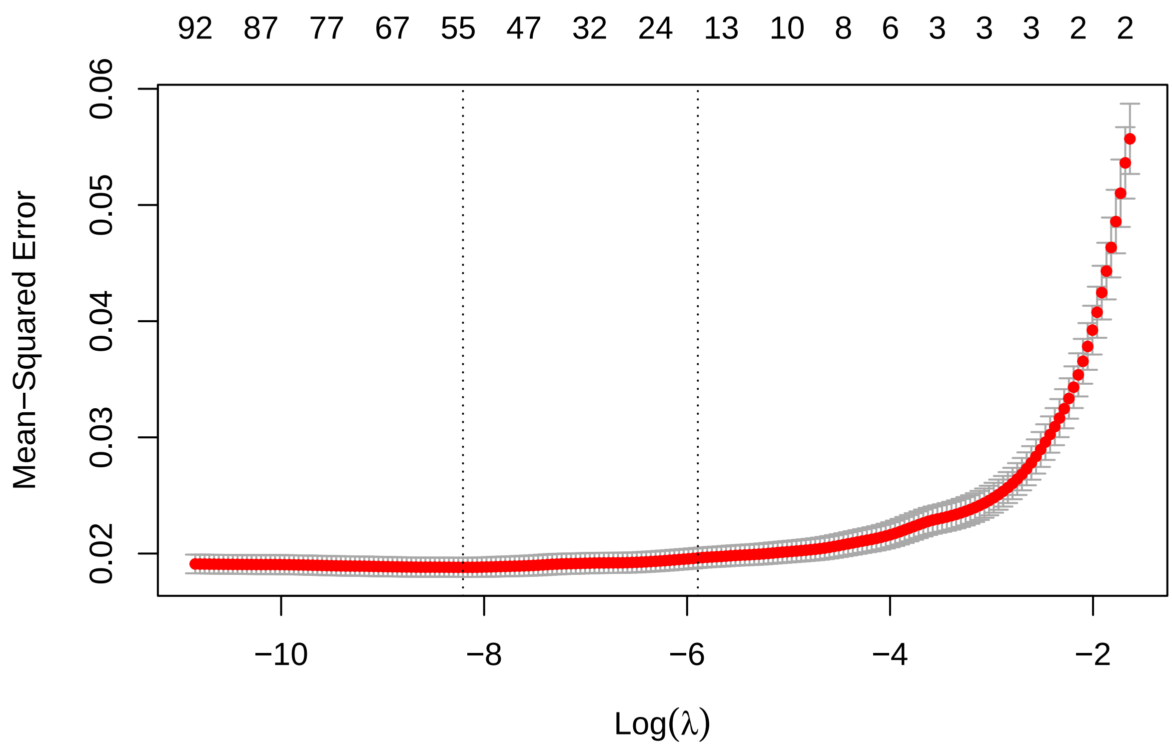
```
## CV residual         0.0757 -0.0286 0.154 -0.0873 -0.141 -0.0124  0.030167
##                       286   289    290    297      304    318    329     396
## Predicted             0.365 0.297 0.0634 0.6017  0.26010  0.393  0.326 0.09905
## cvpred                0.349 0.285 0.0672 0.6193  0.25316  0.433  0.322 0.09886
## ViolentCrimesPerPop   0.140 0.590 0.0900 0.6700  0.25000  0.310  0.170 0.10000
## CV residual          -0.209 0.305 0.0228 0.0507 -0.00316 -0.123 -0.152 0.00114
##                       402    414   422   429     480   502    505    544
## Predicted             0.667 0.9825 0.215 0.199  0.2854 0.171  0.344 0.0195
## cvpred                0.684 0.9604 0.213 0.188  0.2903 0.163  0.342 0.0205
## ViolentCrimesPerPop   0.400 1.0000 0.690 0.710  0.2400 0.240  0.210 0.0500
## CV residual          -0.284 0.0396 0.477 0.522 -0.0503 0.077 -0.132 0.0295
##                       552   568     569    584    753       790    794   829
## Predicted             0.158 0.202  0.4139  0.216  0.230  2.28e-03 0.1778 0.600
## cvpred                0.151 0.206  0.4176  0.219  0.224 -5.54e-05 0.1813 0.587
## ViolentCrimesPerPop   0.050 0.410  0.3300  0.130  0.110  4.00e-02 0.2100 1.000
## CV residual          -0.101 0.204 -0.0876 -0.089 -0.114  4.01e-02 0.0287 0.413
##                         860    862   1014   1019    1069   1125   1134   1139
## Predicted            0.26811  0.406 -0.021  0.270  0.0999  0.083 0.0301  0.243
## cvpred               0.27139  0.431 -0.023  0.283  0.0989  0.087 0.0355  0.256
## ViolentCrimesPerPop  0.28000  0.070  0.020  0.100  0.0700  0.030 0.0500  0.120
## CV residual          0.00861 -0.361  0.043 -0.183 -0.0289 -0.057 0.0145 -0.136
##                        1166    1184   1231  1233    1247   1270   1412     1521
## Predicted             0.3810  0.4479  0.173 0.285 0.26109  0.423 0.0578 -0.00756
## cvpred                0.4121  0.4751  0.177 0.273 0.26361  0.432 0.0582 -0.00712
## ViolentCrimesPerPop   0.3800  0.4100  0.000 0.370 0.27000  0.310 0.1500  0.03000
## CV residual          -0.0321 -0.0651 -0.177 0.097 0.00639 -0.122 0.0918  0.03712
##                        1527    1566   1580    1644   1670    1675   1676   1693
## Predicted            0.2212  0.0897 0.1193  0.1234  0.476  0.2044 0.0106  0.378
## cvpred               0.2161  0.0844 0.1116  0.1298  0.508  0.2042 0.0109  0.373
## ViolentCrimesPerPop  0.2600  0.0400 0.1400  0.0300  0.250  0.1100 0.1800  0.200
## CV residual          0.0439 -0.0444 0.0284 -0.0998 -0.258 -0.0942 0.1691 -0.173
##                       1722     1763    1794    1800  1832   1875    1895    1900
## Predicted            0.657  0.03374  0.0533  0.2825 0.486 0.1545  0.0890 0.00676
## cvpred               0.650  0.03291  0.0524  0.2732 0.462 0.1627  0.0902 0.00456
## ViolentCrimesPerPop  1.000  0.03000  0.0200  0.1900 0.810 0.2100  0.0300 0.08000
## CV residual          0.350 -0.00291 -0.0324 -0.0832 0.348 0.0473 -0.0602 0.07544
##                        1913   1926   1957
## Predicted            0.1227  0.441  0.319
## cvpred               0.1269  0.470  0.314
## ViolentCrimesPerPop  0.1700  0.290  0.300
## CV residual          0.0431 -0.180 -0.014
##
## Sum of squares = 1.8    Mean square = 0.03    n = 66
##
## Overall (Sum over all 66 folds)
##     ms
## 0.0191
```

```
MSE.LASSO<- mean((CV$ViolentCrimesPerPop-CV$cvpred)^2)
MSE.LASSO
```

```
## [1] 0.0191
```

Method two: ADAPTIVE LASSO

```
set.seed(500)
library(MESS)
library(glmnet)
wt <- adaptive.weights(x=X, y=D1$ViolentCrimesPerPop, weight.method="univariate")
cv.ALASSO <- cv.glmnet(x=as.matrix(X), y=y, family="gaussian", alpha=1,
nlambda=200,
 penalty.factor=as.numeric(wt$weights),
standardize=FALSE)
plot(cv.ALASSO)
```



From the Adaptive Lasso Plot, two models are found to be the best; one with 54 variable and the other with 15 variables.

```
beta.hat.alasso <- coef(cv.ALASSO, s="lambda.1se")
cutoff <- 0
terms3 <- names(X)[abs(as.vector(beta.hat.alasso[-1])) > cutoff]
formula.ALASSO <- as.formula(paste(c("y ~ ", terms3),
 collapse=" + "))
fit.ALASSO <- lm(formula.ALASSO, data =D1)
summary(fit.ALASSO)
```

```
##
## Call:
## lm(formula = formula.ALASSO, data = D1)
```

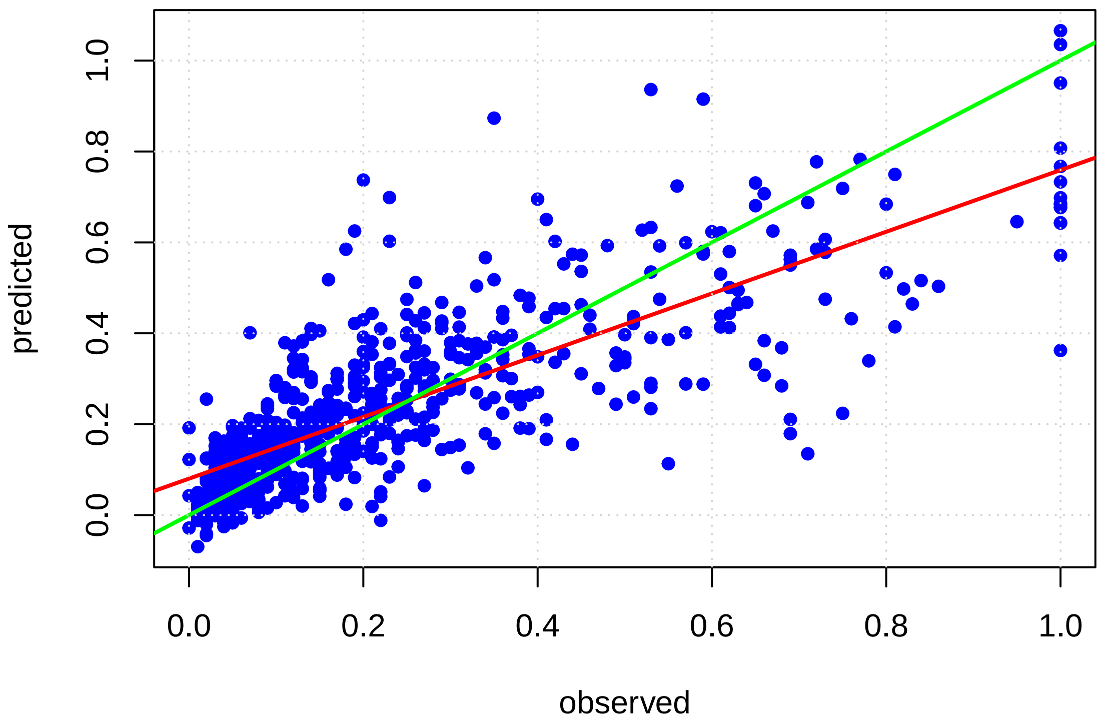
```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4404 -0.0746 -0.0153  0.0486  0.7691
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.000456   0.084384   -0.01   0.9957
## racepctblack         0.222900   0.044578    5.00  6.5e-07 ***
## racePctWhite         0.016779   0.051501    0.33   0.7446
## pctUrban             0.047661   0.009877    4.83  1.6e-06 ***
## pctWWage            -0.086610   0.031555   -2.74   0.0061 **
## pctWPubAsst          0.052168   0.036282    1.44   0.1507
## perCapInc            0.070746   0.030439    2.32   0.0203 *
## MalePctDivorce       0.215161   0.034751    6.19  8.0e-10 ***
## PctKids2Par         -0.116718   0.071027   -1.64   0.1006
## PctIlleg             0.189379   0.047076    4.02  6.1e-05 ***
## PctPersDenseHous     0.237119   0.039560    5.99  2.6e-09 ***
## HousVacant           0.105613   0.035340    2.99   0.0029 **
## PctVacantBoarded     0.044906   0.021759    2.06   0.0392 *
## MedRentPctHousInc    0.039889   0.026001    1.53   0.1252
## NumStreet            0.183633   0.046878    3.92  9.4e-05 ***
## LemasPctOfficDrugUn  0.021877   0.018429    1.19   0.2354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.137 on 1313 degrees of freedom
## Multiple R-squared:  0.67,   Adjusted R-squared:  0.667
## F-statistic:  178 on 15 and 1313 DF,  p-value: <2e-16
```

From the Adaptive lasso, the variables;racepctblack, pctUrban, pctWWage, perCapInc, MalePctDivorce, PctIlleg, PctPersDenseHous, HousVacant, PctVacantBoarded and NumStreet are significant with the adjusted R-Squared of 66.7%.

3(b) Apply the model to the test set D2 and report the the mean square error of prediction (MSEP) ....

```r
MSE <- function(yobs, yhat, plot.it=TRUE, title=""){
    if (plot.it) {
        par(mfrow=c(1,1), mar=rep(4,4))
        plot(yobs, yhat, xlab="observed", ylab="predicted", main=title, col="blue", pch=19, cex=0.8)
        abline(lm(yhat~yobs), col="red", lty=1, lwd=2)
        abline(a=0, b=1, col="green", lty=1, lwd=2)
        grid()
    }
    MSEP <-  mean((yobs-yhat)^2)
    return(MSEP)
}
```

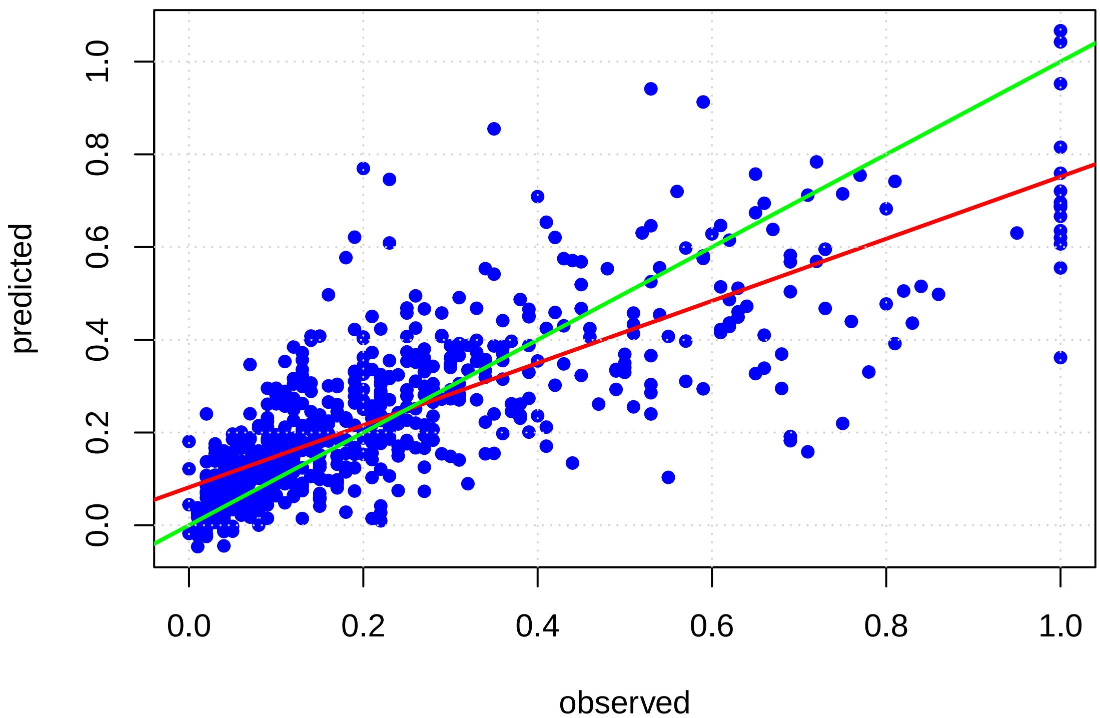
```r
#MSEP of LASSO
yhat.LASSO<-predict(fit.LASSO,newdata=D2)
MSEP.LASSO<-MSE(y.t,yhat.LASSO)
```




```
MSEP.LASSO
```

```
## [1] 0.0186
```

```
#MSEP Error of the Adaptive Lasso
yhat.alasso<-predict(fit.ALASSO,newdata = D2)
MSEP.alasso<-MSE(y.t,yhat.alasso)
```




```
MSEP.alasso
```

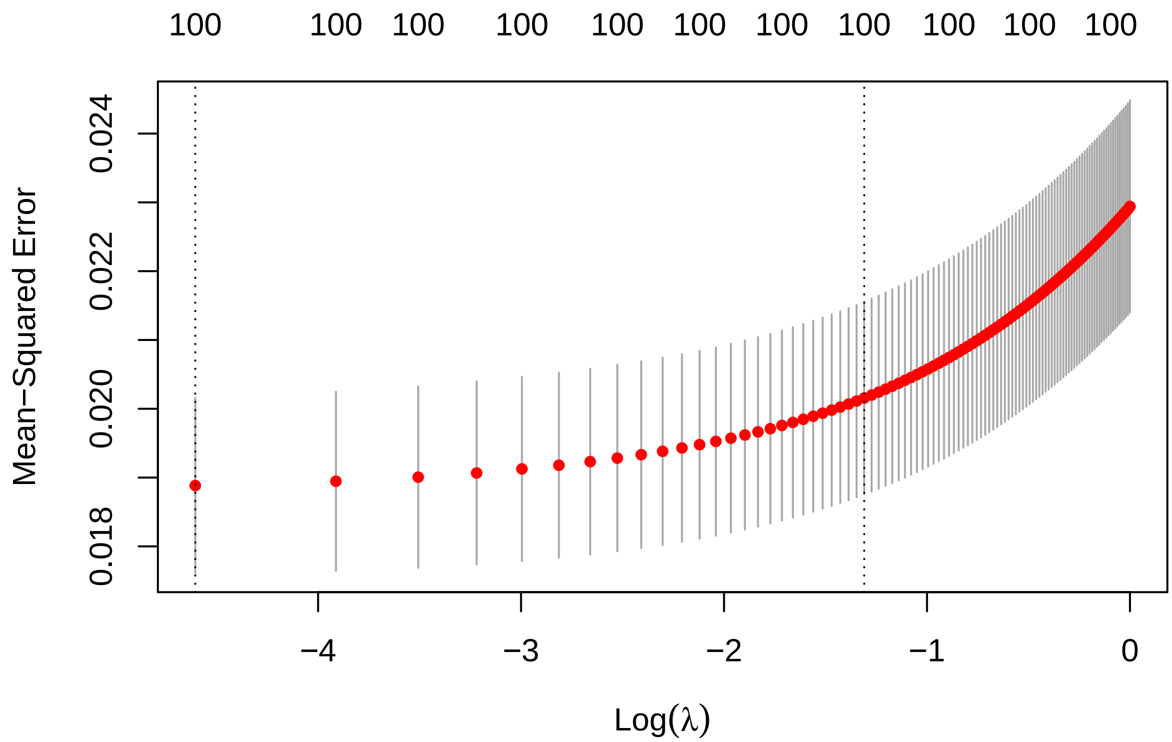
```
## [1] 0.0189
```

From the the MSEP of the two methods , it can be observed that the MSEP of LASSO is smaller than the MSEP of ALASSO

(4)Extended Linear Modeling: Referring to the sample R code R09.R from the class website, fit at least three other models of your own choice from the following models using the training set D1: • Ridge Regression (RR) . . . .

(i) Ridge Regression (RR) =========================

```
library(glmnet)
lambda <- seq(0.0, 1.0, 0.01)
cv.RR <- cv.glmnet(x=as.matrix(X), y=y, alpha = 0, lambda = lambda, nfolds=10)
plot(cv.RR)
```

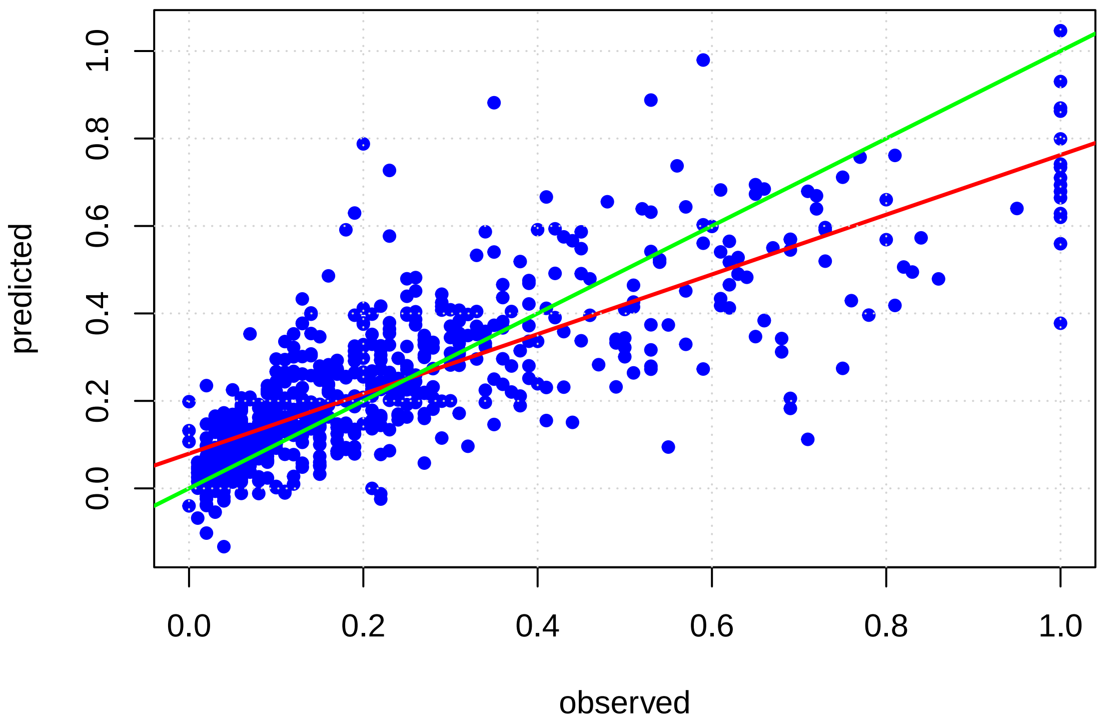



```
lmbd0 <- cv.RR$lambda.min
lmbd0
```

```
## [1] 0.01
```

The best(minimum lambda value) lambda using cross-validation is 0.01.

```
fit.RR <- cv.RR$glmnet.fit
yhat.RR <- predict(fit.RR, s=lmbd0, newx = as.matrix(X.t))
MSEP.RR <- MSE(y.t ,yhat.RR)
```




```
MSEP.RR
```

```
## [1] 0.0183
```

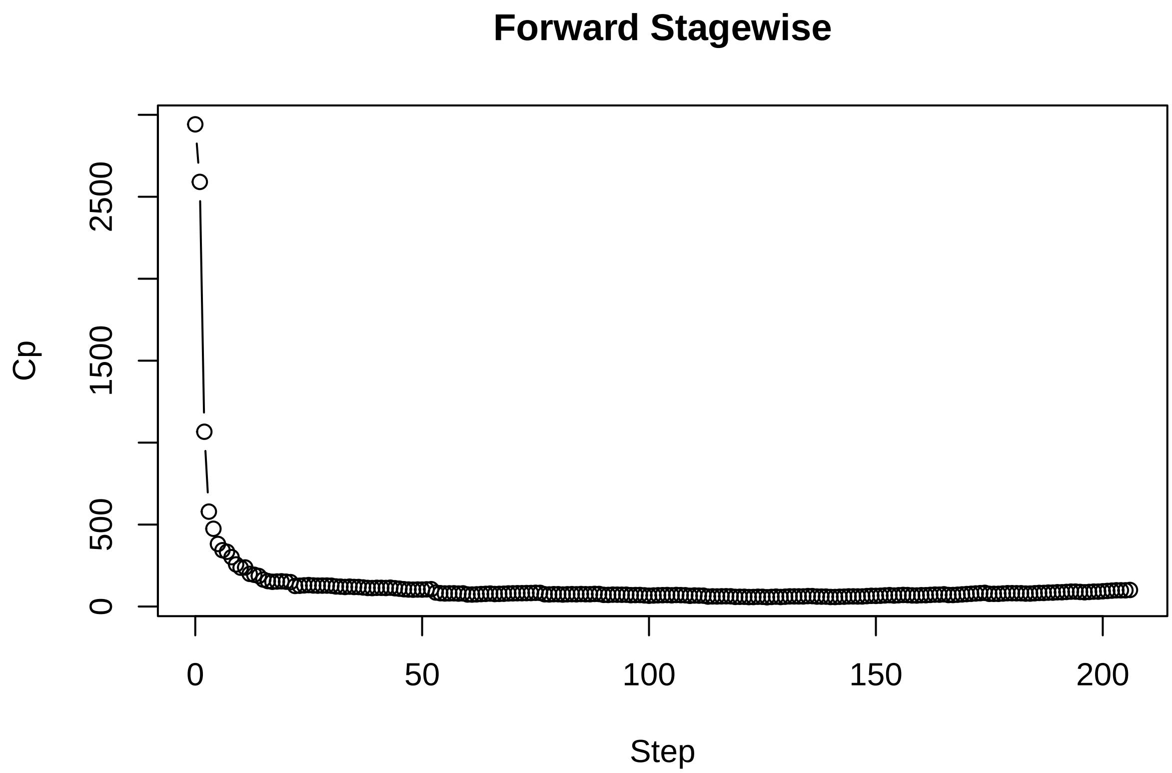
The MSEP of the ridge regression is 0.0183

(ii) Stagewise regression ===

```
library(lars)
```

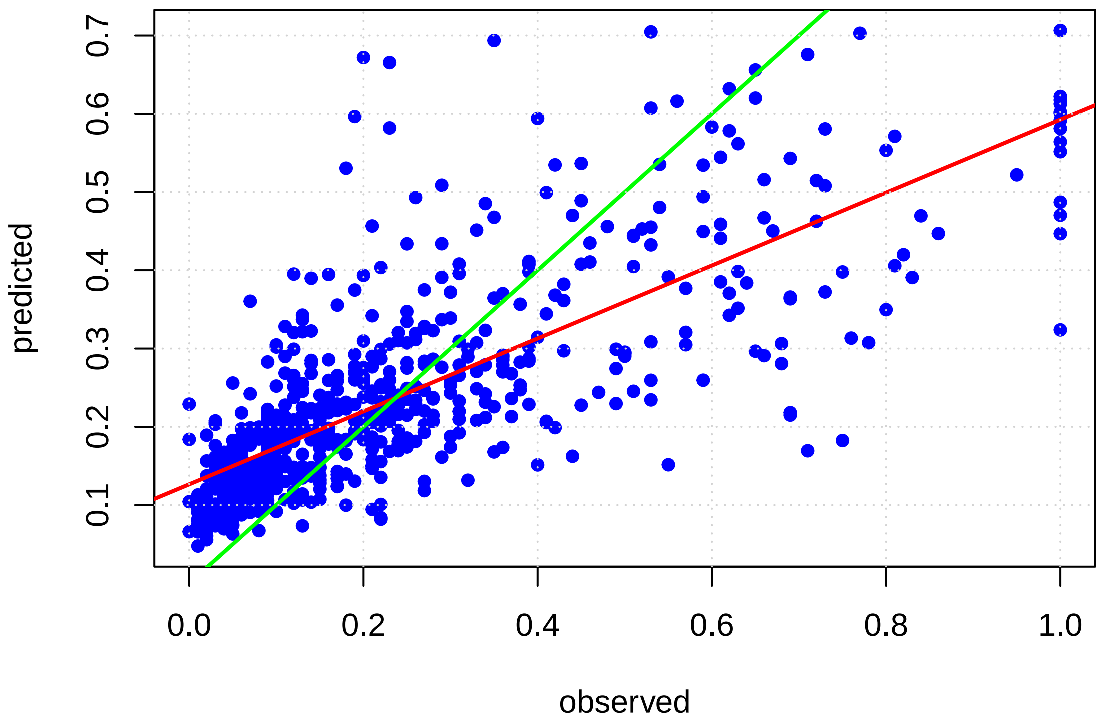
```
## Loaded lars 1.3
```

```
fit.stagewise <- lars(as.matrix(X),y,type="for", trace = F, normalize = F,
                      intercept = TRUE)
plot(fit.stagewise, xvar= "step", breaks = TRUE, plottype = "Cp")
```



From the plot above it is seen that after the 5th step the plot start to decrease slowly.Thus, the maximum number of step is 10

```
b.max.steps <- 5
yhat.stagewise <- predict(fit.stagewise,  s=b.max.steps, newx=X.t)$fit
MSEP.stagewise <- MSE(y.t,yhat.stagewise)
```




```
MSEP.stagewise
```

```
## [1] 0.022
```

From the above, the MSEP of the Stagewise regression is 0.022.

(iii) Least angle regression (LAR)

```
set.seed(500)
fit.lar <- lars(as.matrix(X),y=y,type="lar", trace = FALSE, normalize = TRUE, intercept = TRUE)
summary(fit.lar)
```

```
## LARS/LAR
## Call: lars(x = as.matrix(X), y = y, type = "lar", trace = FALSE, normalize = TRUE,
## Call:     intercept = TRUE)
##      Df  Rss     Cp
## 0     1 74.5 2941.9
## 1     2 73.6 2896.0
## 2     3 48.2 1439.4
## 3     4 33.4  595.7
## 4     5 32.6  552.8
## 5     6 31.5  487.9
```

```
## 6      7 30.4  430.0
## 7      8 29.0  347.5
## 8      9 27.8  284.3
## 9     10 27.7  278.6
## 10    11 26.5  210.0
## 11    12 26.4  211.5
## 12    13 26.4  209.4
## 13    14 25.9  184.5
## 14    15 25.6  166.1
## 15    16 25.4  157.8
## 16    17 25.3  157.6
## 17    18 25.3  158.3
## 18    19 25.0  142.7
## 19    20 25.0  142.1
## 20    21 24.9  142.1
## 21    22 24.7  131.4
## 22    23 24.6  130.2
## 23    24 24.6  131.5
## 24    25 24.5  126.7
## 25    26 24.4  124.6
## 26    27 24.3  119.1
## 27    28 24.2  116.4
## 28    29 24.2  114.9
## 29    30 24.1  114.1
## 30    31 24.1  112.8
## 31    32 24.0  111.3
## 32    33 23.8  103.7
## 33    34 23.8  104.4
## 34    35 23.7   98.2
## 35    36 23.7  100.0
## 36    37 23.7  101.6
## 37    38 23.6  101.6
## 38    39 23.5   98.1
## 39    40 23.3   86.2
## 40    41 23.2   83.4
## 41    42 23.1   82.1
## 42    43 23.1   84.0
## 43    44 23.1   83.5
## 44    45 23.1   85.3
## 45    46 23.1   87.1
## 46    47 23.1   86.7
## 47    48 23.0   88.4
## 48    49 23.0   85.2
## 49    50 22.9   86.6
## 50    51 22.9   87.8
## 51    52 22.9   86.7
## 52    53 22.9   88.0
## 53    54 22.8   89.1
## 54    55 22.8   91.0
## 55    56 22.7   82.4
## 56    57 22.7   84.2
## 57    58 22.5   79.8
## 58    59 22.5   81.5
## 59    60 22.5   83.1
```

```
## 60   61 22.5   84.0
## 61   62 22.5   84.4
## 62   63 22.4   79.7
## 63   64 22.4   80.8
## 64   65 22.3   79.0
## 65   66 22.3   80.7
## 66   67 22.3   82.0
## 67   68 22.2   80.2
## 68   69 22.1   78.6
## 69   70 22.1   80.4
## 70   71 22.1   81.8
## 71   72 22.1   83.2
## 72   73 22.1   82.6
## 73   74 22.0   82.8
## 74   75 21.9   79.0
## 75   76 21.9   76.5
## 76   77 21.9   78.3
## 77   78 21.8   79.7
## 78   79 21.8   80.2
## 79   80 21.7   75.9
## 80   81 21.7   75.5
## 81   82 21.6   75.6
## 82   83 21.6   76.9
## 83   84 21.6   78.6
## 84   85 21.6   79.7
## 85   86 21.6   81.5
## 86   87 21.6   83.4
## 87   88 21.6   84.5
## 88   89 21.6   86.4
## 89   90 21.5   86.4
## 90   91 21.5   87.9
## 91   92 21.5   89.5
## 92   93 21.5   91.0
## 93   94 21.5   91.5
## 94   95 21.5   91.0
## 95   96 21.4   92.3
## 96   97 21.4   93.3
## 97   98 21.4   95.3
## 98   99 21.4   97.0
## 99  100 21.4   99.0
## 100 101 21.4  101.0
```

```
BETA <- as.matrix(fit.lar$beta)
L1.norm <- function(x) sum(abs(x))
norm.L1 <- apply(BETA, 1, L1.norm)
norm.L1 <- norm.L1/max(norm.L1)
df <- as.vector(unlist(fit.lar$df))
Cp <- as.vector(fit.lar$Cp)
RSS <- as.vector(fit.lar$RSS)
lambda <- c(as.vector(fit.lar$lambda), 0)
cbind(df, norm.L1, lambda, RSS, Cp)
```

```
##      df  norm.L1   lambda  RSS     Cp
## 0     1 0.000000 6.396694 74.5 2941.9
```

```
## 1      2 0.000507 6.331101 73.6 2896.0
## 2      3 0.020710 4.025461 48.2 1439.4
## 3      4 0.040455 1.907473 33.4  595.7
## 4      5 0.042561 1.777178 32.6  552.8
## 5      6 0.045985 1.586226 31.5  487.9
## 6      7 0.049405 1.399828 30.4  430.0
## 7      8 0.058707 1.123272 29.0  347.5
## 8      9 0.069258 0.883831 27.8  284.3
## 9     10 0.070491 0.853040 27.7  278.6
## 10    11 0.083595 0.592749 26.5  210.0
## 11    12 0.083683 0.591229 26.4  211.5
## 12    13 0.084520 0.576839 26.4  209.4
## 13    14 0.090167 0.488883 25.9  184.5
## 14    15 0.095409 0.411835 25.6  166.1
## 15    16 0.097974 0.370903 25.4  157.8
## 16    17 0.100203 0.361604 25.3  157.6
## 17    18 0.101854 0.356434 25.3  158.3
## 18    19 0.124503 0.288676 25.0  142.7
## 19    20 0.127910 0.277843 25.0  142.1
## 20    21 0.130528 0.270374 24.9  142.1
## 21    22 0.140079 0.232418 24.7  131.4
## 22    23 0.142670 0.222798 24.6  130.2
## 23    24 0.143416 0.220724 24.6  131.5
## 24    25 0.150932 0.201928 24.5  126.7
## 25    26 0.157113 0.192781 24.4  124.6
## 26    27 0.169392 0.174658 24.3  119.1
## 27    28 0.177300 0.163431 24.2  116.4
## 28    29 0.183169 0.154769 24.2  114.9
## 29    30 0.188249 0.147849 24.1  114.1
## 30    31 0.194359 0.140004 24.1  112.8
## 31    32 0.201288 0.131074 24.0  111.3
## 32    33 0.217297 0.107727 23.8  103.7
## 33    34 0.219822 0.104370 23.8  104.4
## 34    35 0.235640 0.087184 23.7   98.2
## 35    36 0.235969 0.086816 23.7  100.0
## 36    37 0.236760 0.086064 23.7  101.6
## 37    38 0.240300 0.082818 23.6  101.6
## 38    39 0.250689 0.074869 23.5   98.1
## 39    40 0.274288 0.061020 23.3   86.2
## 40    41 0.281330 0.056745 23.2   83.4
## 41    42 0.286184 0.053991 23.1   82.1
## 42    43 0.286394 0.053895 23.1   84.0
## 43    44 0.290523 0.052005 23.1   83.5
## 44    45 0.290752 0.051905 23.1   85.3
## 45    46 0.291128 0.051750 23.1   87.1
## 46    47 0.294866 0.050313 23.1   86.7
## 47    48 0.295443 0.050092 23.0   88.4
## 48    49 0.304094 0.046770 23.0   85.2
## 49    50 0.305050 0.046407 22.9   86.6
## 50    51 0.306745 0.045821 22.9   87.7
## 51    52 0.312706 0.043912 22.9   86.7
## 52    53 0.313856 0.043494 22.9   88.0
## 53    54 0.315442 0.042970 22.8   89.1
## 54    55 0.315479 0.042958 22.8   91.0
```

```
## 55   56 0.336166 0.036909 22.7   82.4
## 56   57 0.336427 0.036837 22.7   84.2
## 57   58 0.351058 0.032775 22.5   79.8
## 58   59 0.351619 0.032641 22.5   81.6
## 59   60 0.352699 0.032385 22.5   83.1
## 60   61 0.355175 0.031798 22.5   84.0
## 61   62 0.358968 0.030971 22.5   84.4
## 62   63 0.378213 0.027367 22.4   79.7
## 63   64 0.381407 0.026830 22.4   80.8
## 64   65 0.396110 0.024720 22.3   79.0
## 65   66 0.397696 0.024533 22.3   80.7
## 66   67 0.400586 0.024212 22.3   82.0
## 67   68 0.417588 0.022412 22.2   80.2
## 68   69 0.433146 0.020786 22.1   78.6
## 69   70 0.434248 0.020680 22.1   80.4
## 70   71 0.437002 0.020411 22.1   81.7
## 71   72 0.439360 0.020182 22.1   83.2
## 72   73 0.451369 0.019036 22.1   82.6
## 73   74 0.459426 0.018342 22.0   82.8
## 74   75 0.488298 0.015902 21.9   79.0
## 75   76 0.513889 0.013759 21.9   76.5
## 76   77 0.515482 0.013639 21.9   78.3
## 77   78 0.518864 0.013387 21.8   79.7
## 78   79 0.528115 0.012719 21.8   80.2
## 79   80 0.577636 0.009366 21.7   75.9
## 80   81 0.600208 0.008244 21.7   75.5
## 81   82 0.619604 0.007315 21.6   75.6
## 82   83 0.628364 0.006908 21.6   76.9
## 83   84 0.631826 0.006746 21.6   78.6
## 84   85 0.643454 0.006255 21.6   79.6
## 85   86 0.645722 0.006167 21.6   81.5
## 86   87 0.646570 0.006138 21.6   83.4
## 87   88 0.660616 0.005681 21.6   84.5
## 88   89 0.663191 0.005616 21.6   86.4
## 89   90 0.697594 0.004819 21.5   86.4
## 90   91 0.708093 0.004576 21.5   87.9
## 91   92 0.716609 0.004434 21.5   89.5
## 92   93 0.726205 0.004276 21.5   91.0
## 93   94 0.763486 0.003670 21.5   91.5
## 94   95 0.840630 0.002422 21.5   91.0
## 95   96 0.868372 0.001982 21.4   92.3
## 96   97 0.939256 0.000867 21.4   93.3
## 97   98 0.940590 0.000846 21.4   95.3
## 98   99 0.982645 0.000232 21.4   97.0
## 99  100 0.990871 0.000118 21.4   99.0
## 100 101 1.000000 0.000000 21.4  101.0
```

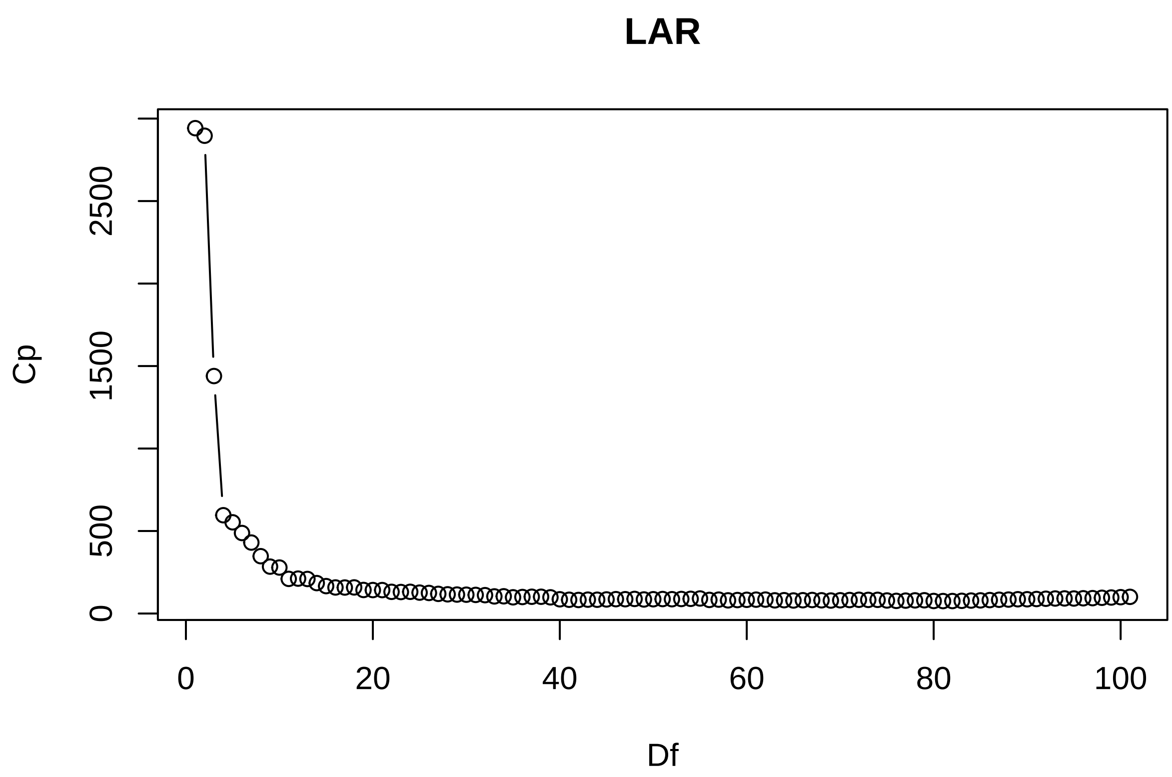
```r
b.lambda <- lambda[Cp==min(Cp)]
b.lambda
```

```
## [1] 0.00824
```

```
b.L1norm <- norm.L1[Cp==min(Cp)]
b.L1norm
```

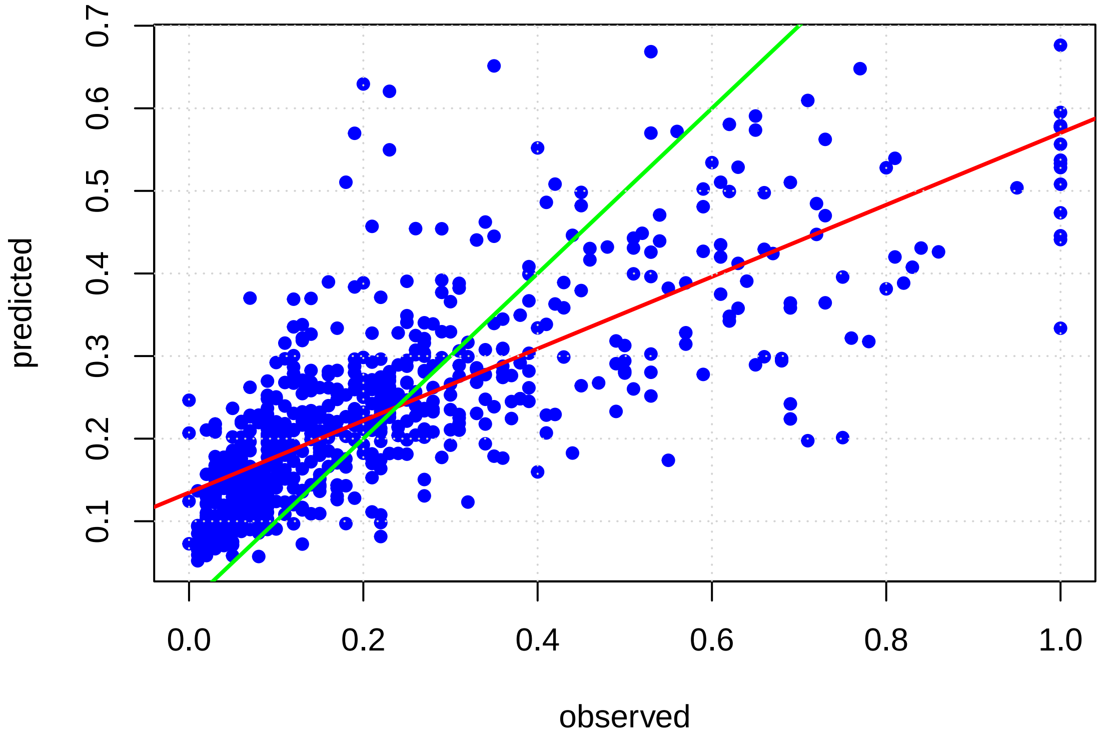
```
##  80
## 0.6
```

```
plot(fit.lar, xvar= "df", plottype="Cp", omit.zeros = F)
```



From the plot above, the graph starts to decrease after the 5th step which indicates that the best step is 5

```
b.max.steps <- 5
yhat.lar <- predict(fit.lar,  s=b.max.steps, newx=as.matrix(X.t))$fit
MSEP.lar <- MSE(y.t,yhat.lar)
```




```
MSEP.lar
```

```
## [1] 0.0225
```

The MSEP of the Least angle regression (LAR) is 0.0225.

Comparing Models using their MSE

```
cbind(MSEP.LASSO, MSEP.alasso,MSEP.stagewise, MSEP.RR, MSEP.lar)
```

```
##      MSEP.LASSO MSEP.alasso MSEP.stagewise MSEP.RR MSEP.lar
## [1,]     0.0186      0.0189          0.022  0.0183   0.0225
```

We compare the MSEP of LASSO, adaptive LASSO, Stagewise regression,ridge regression and Least angle regression. Irt can be observed from the above that the Ridge regression has the lowest MSEP error followed by LASSO, Adaptive Lasso, Stagewise regression and Least angle regression.