week1_tutorial

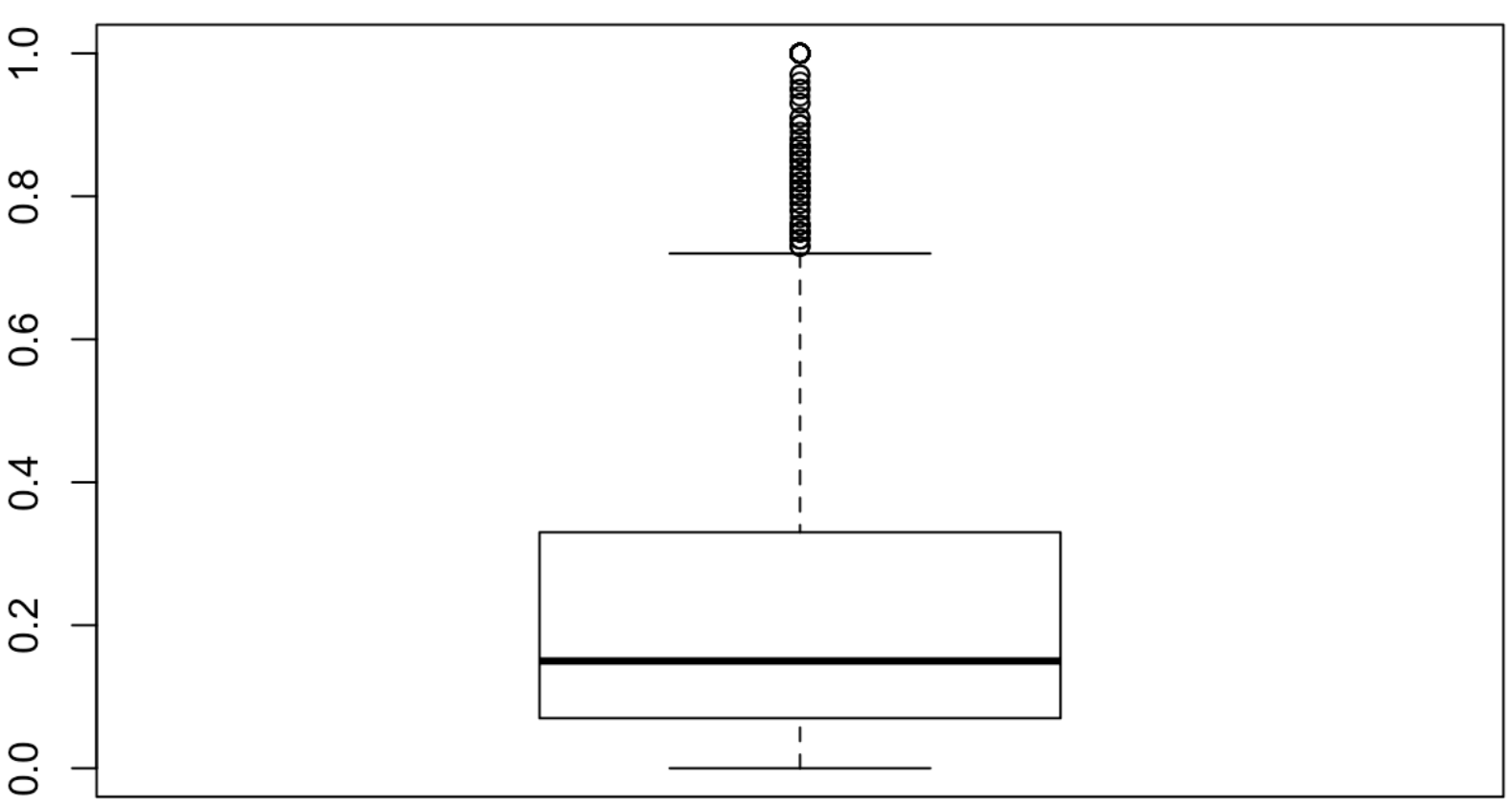
Likun Cui(41725041)

06 August, 2018

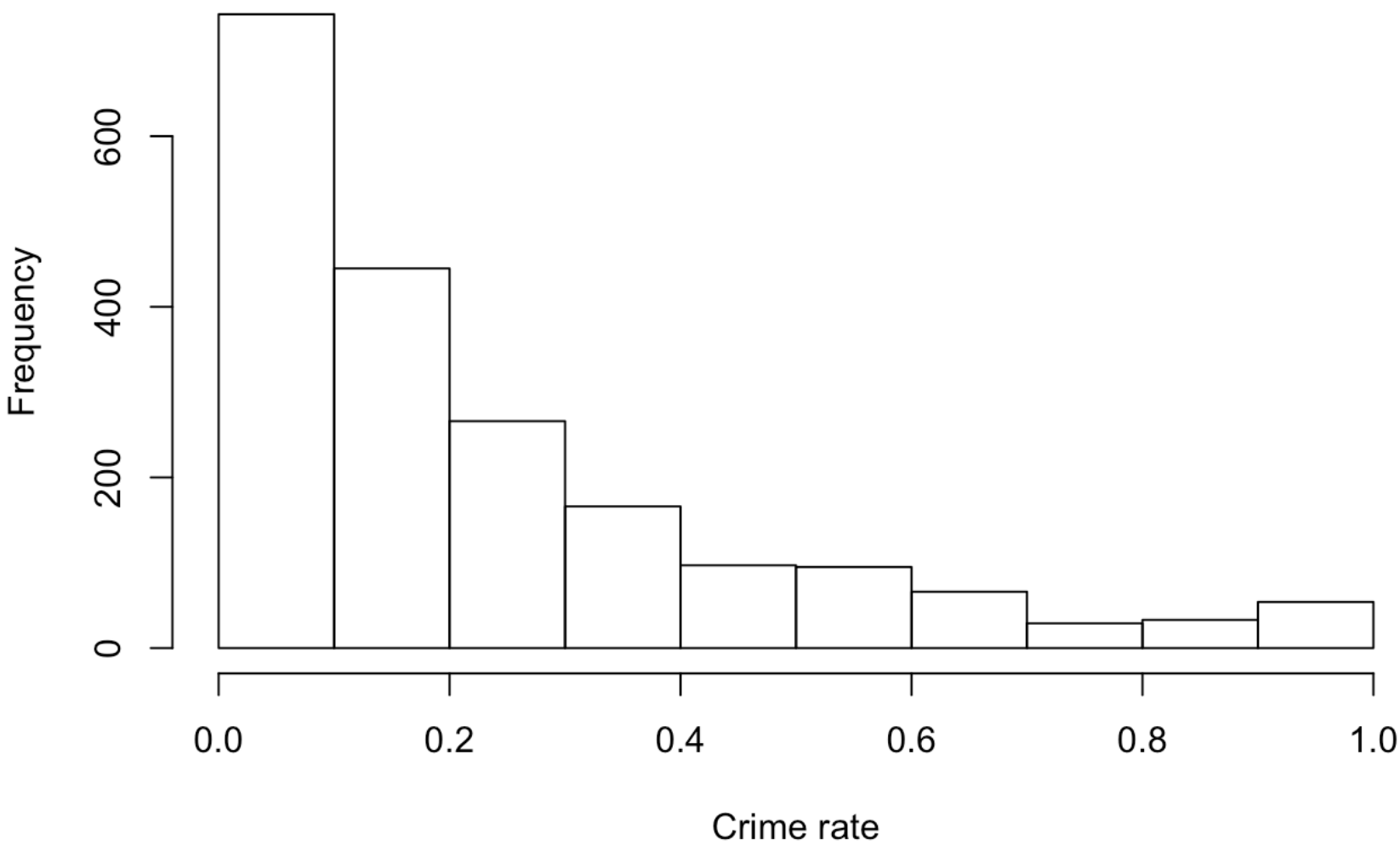
```
setwd("/Users/likuncui/Downloads/5003/Data_w1/")
#(2) read data
data<-read.delim("communities.data",sep=",",header=FALSE);
names<-read.delim("communities.names",head= FALSE);

#generating response
response<-data[,ncol(data)]
summary (response)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.070   0.150   0.238   0.330   1.000
```

```
boxplot(response)
```




```
hist(response,xlab="Crime rate")
```




```
sd(response)
```

```
## [1] 0.2329849
```

```
#selecting variables containing no missing values
selected<-(colSums(data=="?")==0)
datComplete<-data[,selected]
names[selected,]
```

```
##   [1] @attribute state numeric
##   [2] @attribute communityname string
##   [3] @attribute fold numeric
##   [4] @attribute population numeric
##   [5] @attribute householdsize numeric
##   [6] @attribute racepctblack numeric
##   [7] @attribute racePctWhite numeric
##   [8] @attribute racePctAsian numeric
##   [9] @attribute racePctHisp numeric
##  [10] @attribute agePct12t21 numeric
##  [11] @attribute agePct12t29 numeric
##  [12] @attribute agePct16t24 numeric
##  [13] @attribute agePct65up numeric
##  [14] @attribute numbUrban numeric
##  [15] @attribute pctUrban numeric
```

```
##  [16] @attribute medIncome numeric
##  [17] @attribute pctWWage numeric
##  [18] @attribute pctWFarmSelf numeric
##  [19] @attribute pctWInvInc numeric
##  [20] @attribute pctWSocSec numeric
##  [21] @attribute pctWPubAsst numeric
##  [22] @attribute pctWRetire numeric
##  [23] @attribute medFamInc numeric
##  [24] @attribute perCapInc numeric
##  [25] @attribute whitePerCap numeric
##  [26] @attribute blackPerCap numeric
##  [27] @attribute indianPerCap numeric
##  [28] @attribute AsianPerCap numeric
##  [29] @attribute HispPerCap numeric
##  [30] @attribute NumUnderPov numeric
##  [31] @attribute PctPopUnderPov numeric
##  [32] @attribute PctLess9thGrade numeric
##  [33] @attribute PctNotHSGrad numeric
##  [34] @attribute PctBSorMore numeric
##  [35] @attribute PctUnemployed numeric
##  [36] @attribute PctEmploy numeric
##  [37] @attribute PctEmplManu numeric
##  [38] @attribute PctEmplProfServ numeric
##  [39] @attribute PctOccupManu numeric
##  [40] @attribute PctOccupMgmtProf numeric
##  [41] @attribute MalePctDivorce numeric
##  [42] @attribute MalePctNevMarr numeric
##  [43] @attribute FemalePctDiv numeric
##  [44] @attribute TotalPctDiv numeric
##  [45] @attribute PersPerFam numeric
##  [46] @attribute PctFam2Par numeric
##  [47] @attribute PctKids2Par numeric
##  [48] @attribute PctYoungKids2Par numeric
##  [49] @attribute PctTeen2Par numeric
##  [50] @attribute PctWorkMomYoungKids numeric
##  [51] @attribute PctWorkMom numeric
##  [52] @attribute NumIlleg numeric
##  [53] @attribute PctIlleg numeric
##  [54] @attribute NumImmig numeric
##  [55] @attribute PctImmigRecent numeric
##  [56] @attribute PctImmigRec5 numeric
##  [57] @attribute PctImmigRec8 numeric
##  [58] @attribute PctImmigRec10 numeric
##  [59] @attribute PctRecentImmig numeric
##  [60] @attribute PctRecImmig5 numeric
##  [61] @attribute PctRecImmig8 numeric
##  [62] @attribute PctRecImmig10 numeric
##  [63] @attribute PctSpeakEnglOnly numeric
##  [64] @attribute PctNotSpeakEnglWell numeric
##  [65] @attribute PctLargHouseFam numeric
##  [66] @attribute PctLargHouseOccup numeric
##  [67] @attribute PersPerOccupHous numeric
##  [68] @attribute PersPerOwnOccHous numeric
##  [69] @attribute PersPerRentOccHous numeric
```

```
##  [70] @attribute PctPersOwnOccup numeric
##  [71] @attribute PctPersDenseHous numeric
##  [72] @attribute PctHousLess3BR numeric
##  [73] @attribute MedNumBR numeric
##  [74] @attribute HousVacant numeric
##  [75] @attribute PctHousOccup numeric
##  [76] @attribute PctHousOwnOcc numeric
##  [77] @attribute PctVacantBoarded numeric
##  [78] @attribute PctVacMore6Mos numeric
##  [79] @attribute MedYrHousBuilt numeric
##  [80] @attribute PctHousNoPhone numeric
##  [81] @attribute PctWOFullPlumb numeric
##  [82] @attribute OwnOccLowQuart numeric
##  [83] @attribute OwnOccMedVal numeric
##  [84] @attribute OwnOccHiQuart numeric
##  [85] @attribute RentLowQ numeric
##  [86] @attribute RentMedian numeric
##  [87] @attribute RentHighQ numeric
##  [88] @attribute MedRent numeric
##  [89] @attribute MedRentPctHousInc numeric
##  [90] @attribute MedOwnCostPctInc numeric
##  [91] @attribute MedOwnCostPctIncNoMtg numeric
##  [92] @attribute NumInShelters numeric
##  [93] @attribute NumStreet numeric
##  [94] @attribute PctForeignBorn numeric
##  [95] @attribute PctBornSameState numeric
##  [96] @attribute PctSameHouse85 numeric
##  [97] @attribute PctSameCity85 numeric
##  [98] @attribute PctSameState85 numeric
##  [99] @attribute LandArea numeric
## [100] @attribute PopDens numeric
## [101] @attribute PctUsePubTrans numeric
## [102] @attribute LemasPctOfficDrugUn numeric
## [103] @attribute ViolentCrimesPerPop numeric
## 128 Levels: @attribute agePct12t21 numeric ...
```

```
# selecting variables that are numeric
darNumeric.raw<-datComplete[,-2]
datNumeric<-apply(darNumeric.raw, 2, as.numeric)

# use a loop to calculate correlation of each variable to the response variable
correlationVector <- c()
for(i in 1:ncol(datNumeric)) {
  correlationVector <- c(correlationVector, cor(datNumeric[,i], response))
}
names(correlationVector) <- colnames(datNumeric)

# sort the variable by correlation from high to low and select the top 9
newNames<- names[selected,]
newNames[-2][order(abs(correlationVector), decreasing = TRUE)[1:9]]
```

```
## [1] @attribute ViolentCrimesPerPop numeric
## [2] @attribute PctKids2Par numeric
## [3] @attribute PctIlleg numeric
## [4] @attribute PctFam2Par numeric
## [5] @attribute racePctWhite numeric
## [6] @attribute PctYoungKids2Par numeric
## [7] @attribute PctTeen2Par numeric
## [8] @attribute racepctblack numeric
## [9] @attribute pctWInvInc numeric
## 128 Levels: @attribute agePct12t21 numeric ...
```