School Crimes

Daniel Gutstein, Ki Young, Yusheng Zhu
and Saravana Kumar Loganathan

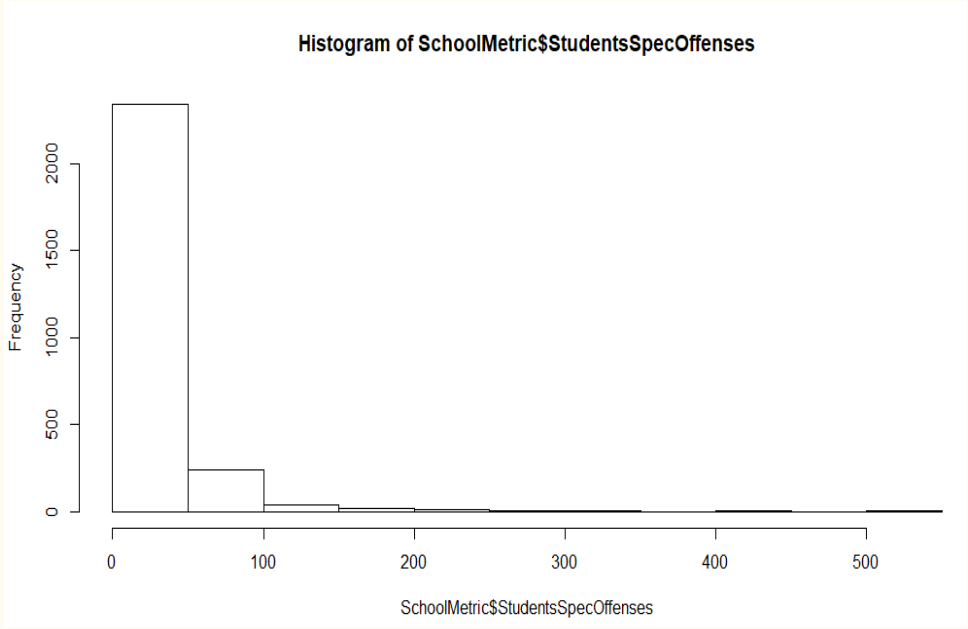
Dataset Background Information

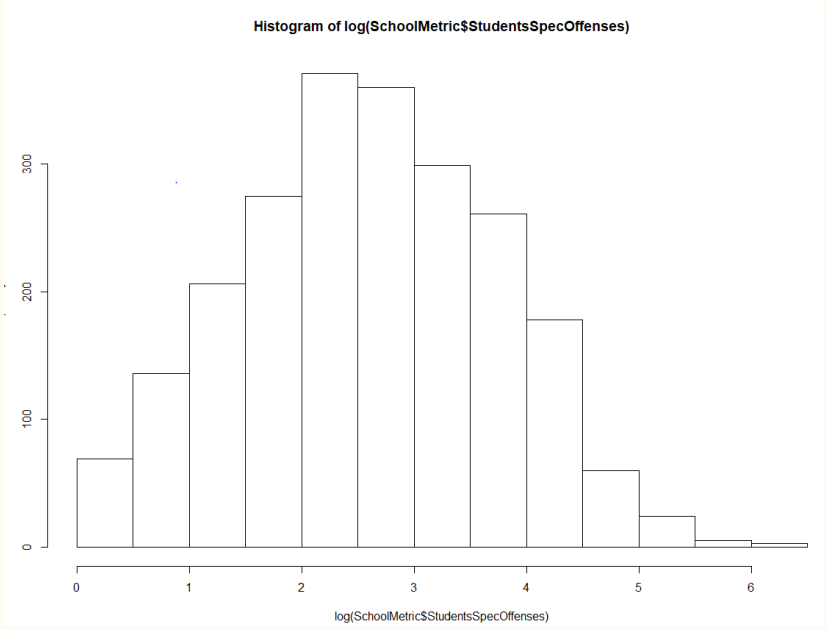
- Dataset: School Survey on Crime and Safety, 2010
  - Obtained from U.S Department of Education Data Portal
  - Survey of nation's schools designed to provide estimates on school crime, discipline, disorder, programs, and policies.
  - Data obtained with questionnaire and telephone follow-ups of school principals from School staffs
- 192 Variables (2648 cases)
  - 35 Ordinal
  - 34 Metric
  - 123 Categorical Variables
    - Binary and Nominal
      - Legitimate Skip connoted with entry of -1 (instead of missing value)

Goal & Response Variable

- Analyze data to build model that predict number of students involved in school crime and find statistically significant factors that contributed in school crime.

- StudentsSpecOffenses = Total students involved in specified offenses



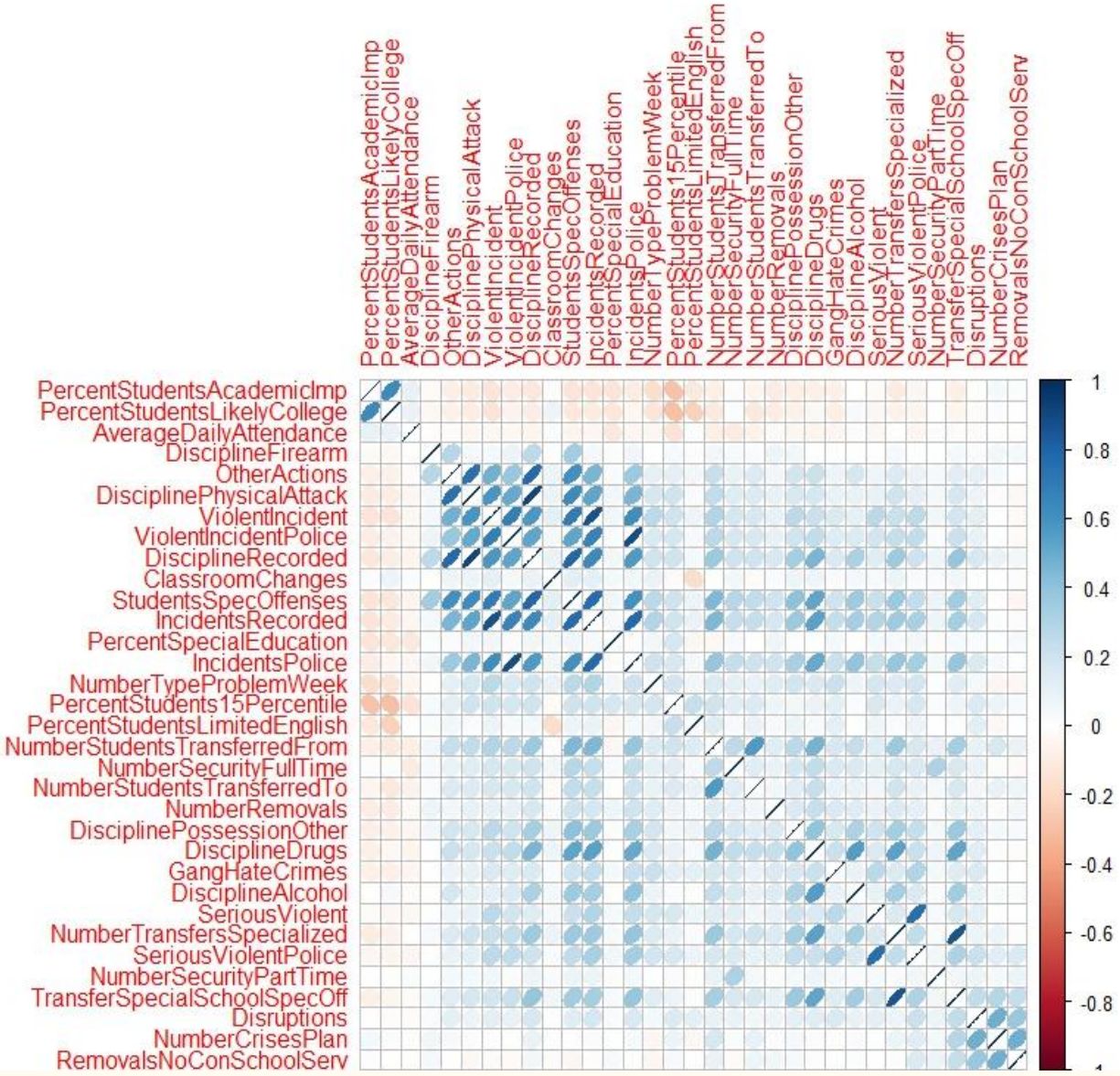


Binary Data examples

- VisCheckReq =School practice require vistor check in
- Access_Doors_MonitLocked = Access controlled locked/monitored doors
- Security = Security guard, security personnel, or sworn law enforcement officer
- TeachTrainStudentsDrugs = Teach or train students (e/g/, drug-related education)
- TeacherAlcDrugAbus = Teacher training-student alcohol/drug abuse
- CorporalPunishment = Corporal punishment available
- SchoolProbation = School probation available
- MetalDetectors = Students pass through metal detector
- DogSniffsDrugs = Practice random dog sniffs for drugs
- SecurityCameras = Security camera(s) monitor the school
- CellPhoneBan = Prohibit use of cell phones and text messaging devices
- ShootingPlan = School has written plan for shootings
- BombThreatPlan = Written plan for bomb threats
- DetentionSaturday =Detention/Saturday school available
- TransferSpecialAvailable = Transfer to specialized school available
- SchoolProbation = School probation available
    - -1 = “Legitmate Skip”
    - 1 = “Yes”
    - 2 = “No”

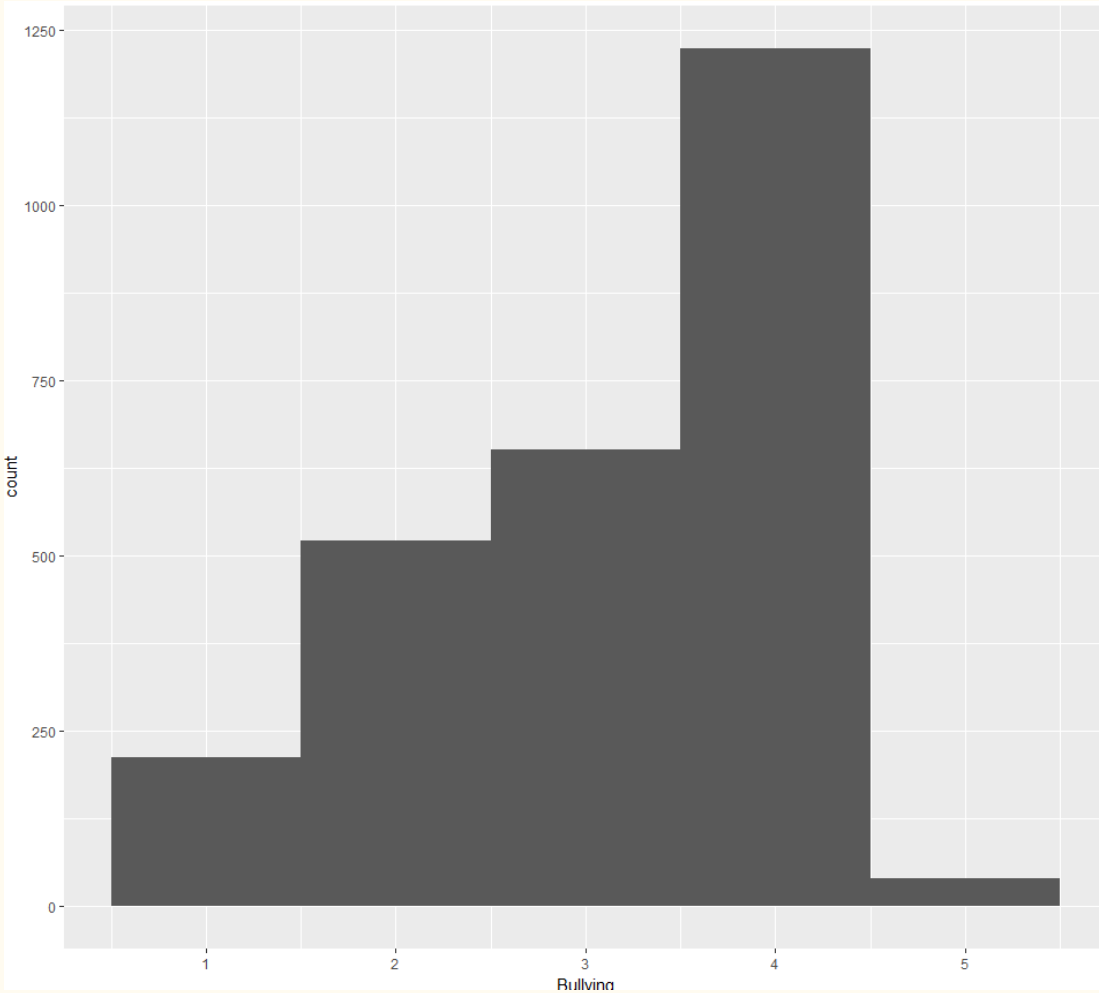
Metric Data Examples

1) StudentsSpecOffenses = Total students involved in specified offenses
2) SeriousViolent = Total number of serious violent incidents recorded
3) SeriousViolentPolice = Total number of serious violent incidents reported to police
4) TransferSpecialSchoolSpecOff = Total transfers to specialized schools for specified offenses
5) ViolentIncident = Total number of violent incidents recorded
6) ViolentIncidentPolice = Total number of violent incidents reported to Police
7) DisciplineFirearm = Total Number of disciplinary actions recorded for use or possession of a firearm or explosive device
8) DisciplineDrugs = Total number of disciplinary actions recorded for distribution, possession, or use of illegal drugs
9) DisciplinePossessionOther = Total number of disciplinary actions recorded for use or possession of a weapon other than a firearm or explosive device
10) GangHateCrimes = Total number of gang-related and hate crimes
11) Disruptions = Total number of disruptions
12) DisciplinePhysicalAttack = Total number of disciplinary actions recorded for physical attacks or fights.
13) DisciplineAlcohol = Total number of disciplinary actions recorded for distribution, possession, or use of alcohol
14) NumberSecurityFullTime = Total number of full-time security guards, SROs, or sworn law enforcement officers
15) NumberSecurityPartTime == Total number of part-time security guards, SROs, or sworn law enforcement officers
16) PercentStudentsLikelyCollege = Percent students likely to go to college
17) PercentStudentsAcademicImp = Percent students academic achievement important


PercentStudentsAcademicImp
PercentStudentsLikelyCollege
AverageDailyAttendance
DisciplineFirearm
OtherActions
DisciplinePhysicalAttack
ViolentIncident
ViolentIncidentPolice
DisciplineRecorded
ClassroomChanges
StudentsSpecOffenses
IncidentsRecorded
PercentSpecialEducation
IncidentsPolice
NumberTypeProblemWeek
PercentStudents15Percentile
PercentStudentsLimitedEnglish
NumberStudentsTransferredFrom
NumberSecurityFullTime
NumberStudentsTransferredTo
NumberRemovals
DisciplinePossessionOther
DisciplineDrugs
GangHateCrimes
DisciplineAlcohol
SeriousViolent
NumberTransfersSpecialized
SeriousViolentPolice
NumberSecurityPartTime
TransferSpecialSchoolSpecOff
Disruptions
NumberCrisesPlan
RemovalsNoConSchoolServ
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

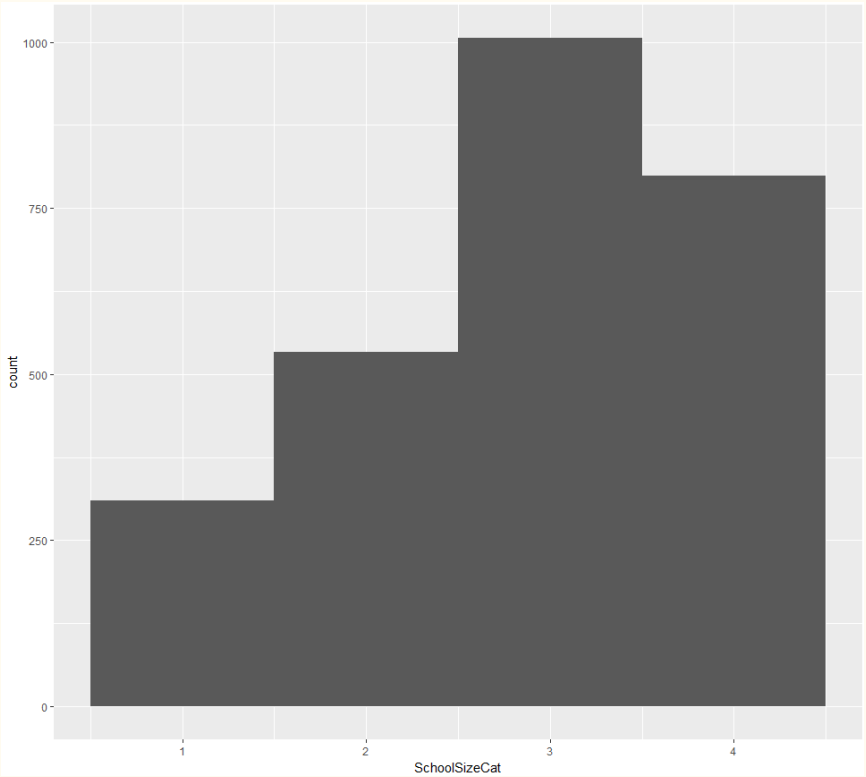
Ordinal Data

- Example Question1:
- 103. Bullying = How often student bullying occurs
  - 1 = “Happens Daily”
  - 2 = “Happens at least once a week”
  - 3 = “Happens at least once a month”
  - 4 = “Happens on occasion”
  - 5 = “Never happens”
- Flipped the data.



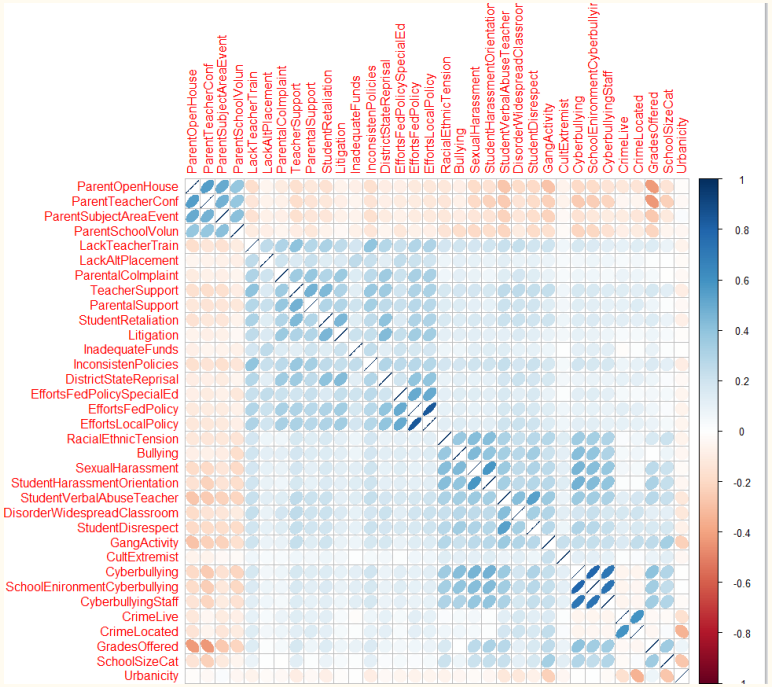
Ordinal Data

- Example Question 2:
  - SchoolSizeCat = School size categories - based on 07-08 CCD frame variables(School)
    - 1 = “< 300”
    - 2 = “300 - 499”
    - 3 = “500 - 999”
    - 4 = “1000+”
  - LackTeacherTrain = Efforts limited by inadequate/lack of teacher training
    - 1 = “Limits in major way”
    - 2 = “Limits in minor way”
    - 3 = “Does not limit”
  - 52. ParentOpenHouse = Parent participates in open house or back to school night
    - 1 = 0-25%
    - 2 = 26-50%
    - 3 = 51-75%
    - 4 = 76 - 100%
    - 5 = “School does not offer”

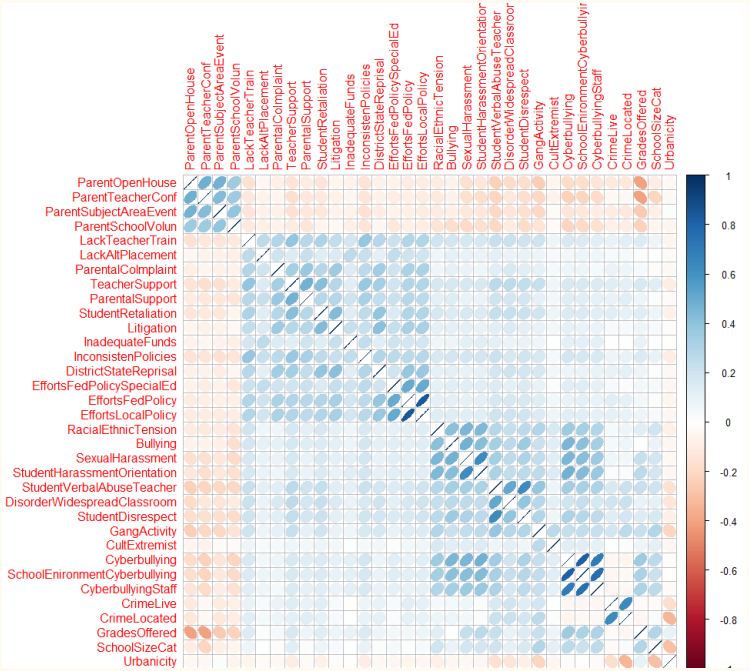


Ordinal Data

- Spearman Correlation



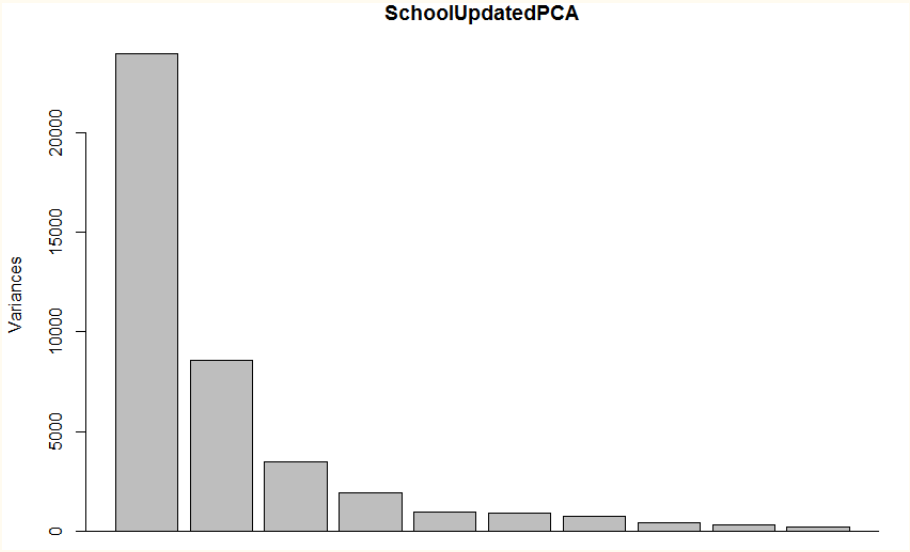
- Pearson Correlation

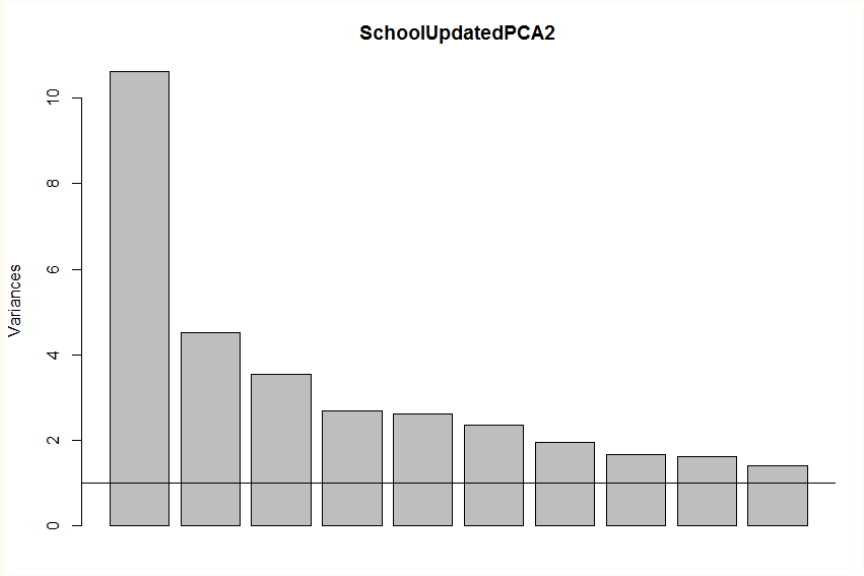


METHOD 1: Principal Component Analysis

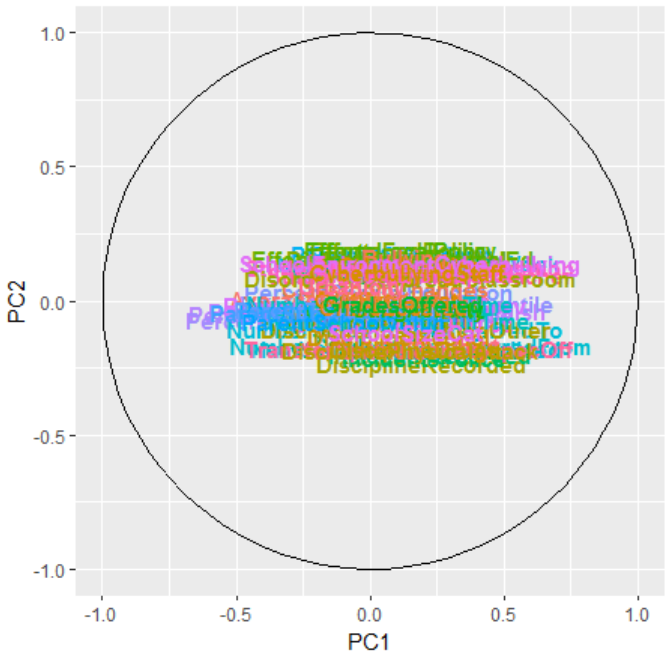
- Use 6 Principal Components (from combined metric and ordinal variables; 68)

```
> summary(SchoolUpdatedPCA)
Importance of components:
                          PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8
Standard deviation     3.3573 2.20028 1.94113 1.64675 1.60987 1.52230 1.39762 1.31306
Proportion of Variance 0.1658 0.07119 0.05541 0.03988 0.03811 0.03408 0.02873 0.02535
Cumulative Proportion  0.1658 0.23695 0.29236 0.33224 0.37036 0.40444 0.43316 0.45852
```





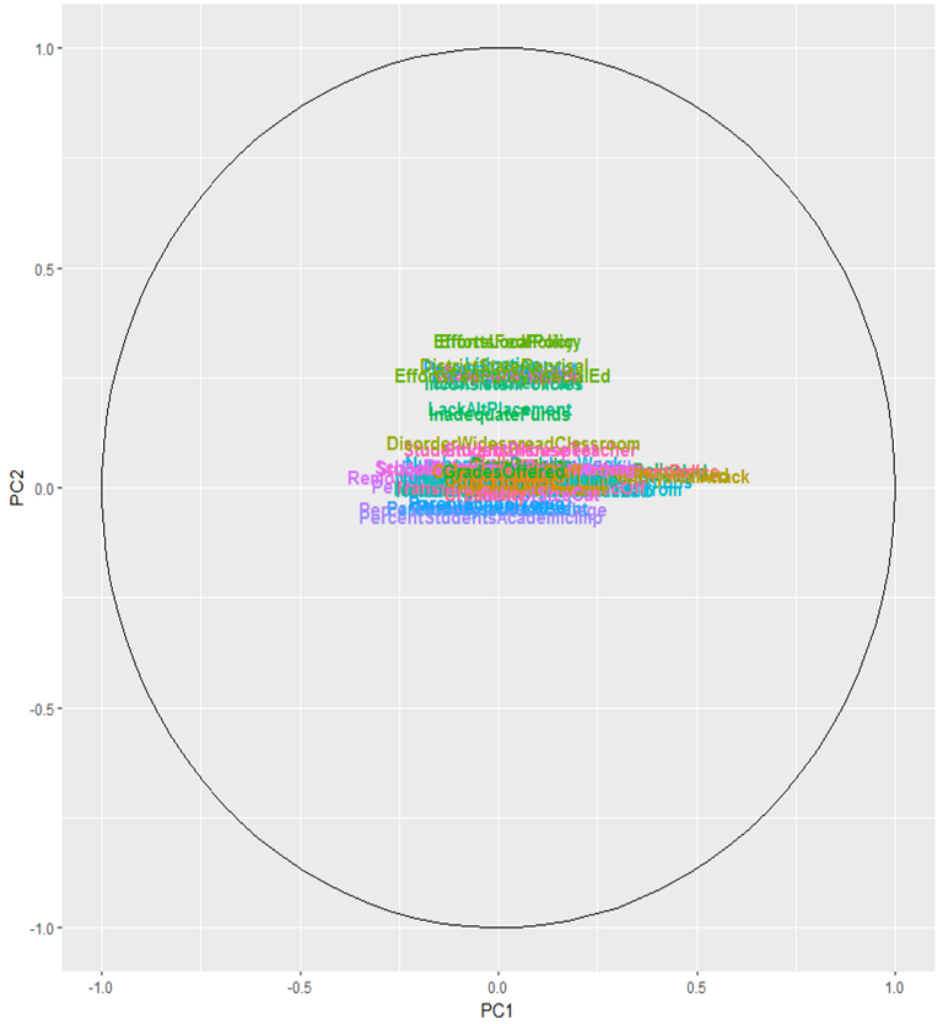
PCA Plot (No Rotation)



.names

AverageDailyAttendance	EffortsFedPolicy	NumberStudentsTransferredFrom	RemovalsNoConSchoolServ
Bullying	EffortsFedPolicySpecialEd	NumberStudentsTransferredTo	SchoolEnironmentCyberbullying
ClassroomChanges	EffortsLocalPolicy	NumberTransfersSpecialized	SchoolSizeCat
CrimeLive	GangActivity	NumberTypeProblemWeek	SeriousViolent
CrimeLocated	GangHateCrimes	OtherActions	SeriousViolentPolice
CultExtremist	GradesOffered	ParentalCoImplaint	SexualHarassment
Cyberbullying	InadequateFunds	ParentalSupport	StudentDisrespect
CyberbullyingStaff	IncidentsPolice	ParentOpenHouse	StudentHarassmentOrientation
DisciplineAlcohol	IncidentsRecorded	ParentSchoolVolun	StudentRetaliation
DisciplineDrugs	InconsistenPolicies	ParentSubjectAreaEvent	StudentVerbalAbuseTeacher
DisciplineFirearm	LackAltPlacement	ParentTeacherConf	TeacherSupport
DisciplinePhysicalAttack	LackTeacherTrain	PercentSpecialEducation	TransferSpecialSchoolSpecOff
DisciplinePossessionOther	Litigation	PercentStudents15Percentile	Urbanicity
DisciplineRecorded	NumberCrisesPlan	PercentStudentsAcademicImp	ViolentIncident
DisorderWidespreadClassroom	NumberRemovals	PercentStudentsLikelyCollege	ViolentIncidentPolice
Disruptions	NumberSecurityFullTime	PercentStudentsLimitedEnglish	
DistrictStateReprisal	NumberSecurityPartTime	RacialEthnicTension	

PCA Plot (Rotation with Varimax)



.names

AverageDailyAttendance	EffortsFedPolicy	NumberStudentsTransferredFrom	RemovalsNoConSchoolServ
Bullying	EffortsFedPolicySpecialEd	NumberStudentsTransferredTo	SchoolEnironmentCyberbullying
ClassroomChanges	EffortsLocalPolicy	NumberTransfersSpecialized	SchoolSizeCat
CrimeLive	GangActivity	NumberTypeProblemWeek	SeriousViolent
CrimeLocated	GangHateCrimes	OtherActions	SeriousViolentPolice
CultExtremist	GradesOffered	ParentalCoImplaint	SexualHarassment
Cyberbullying	InadequateFunds	ParentalSupport	StudentDisrespect
CyberbullyingStaff	IncidentsPolice	ParentOpenHouse	StudentHarassmentOrientation
DisciplineAlcohol	IncidentsRecorded	ParentSchoolVolun	StudentRetaliation
DisciplineDrugs	InconsistenPolicies	ParentSubjectAreaEvent	StudentVerbalAbuseTeacher
DisciplineFirearm	LackAltPlacement	ParentTeacherConf	TeacherSupport
DisciplinePhysicalAttack	LackTeacherTrain	PercentSpecialEducation	TransferSpecialSchoolSpecOff
DisciplinePossessionOther	Litigation	PercentStudents15Percentile	Urbanicity
DisciplineRecorded	NumberCrisesPlan	PercentStudentsAcademicImp	ViolentIncident
DisorderWidespreadClassroom	NumberRemovals	PercentStudentsLikelyCollege	ViolentIncidentPolice
Disruptions	NumberSecurityFullTime	PercentStudentsLimitedEnglish	
DistrictStateReprisal	NumberSecurityPartTime	RacialEthnicTension	

```
> p2 = psych::principal(SchoolMetricOrdinal, rotate="varimax", nfactors=6, scores=TRUE)
> print(p2$loadings, cutoff=.47, sort=T)

Loadings:
                           RC3    RC5    RC2    RC1    RC4    RC6
NumberTypeProblemWeek       0.786
RacialEthnicTension         0.596
Bullying                    0.695
SexualHarassment            0.690
StudentHarassmentOrientation 0.679
StudentVerbalAbuseTeacher   0.556
StudentDisrespect           0.528
Cyberbullying               0.741
SchoolEnironmentCyberbullying 0.720
CyberbullyingStaff          0.645
NumberTransfersSpecialized         0.657
NumberStudentsTransferredFrom      0.556
TransferSpecialSchoolSpecOff       0.742
DisciplineDrugs                    0.680
DisciplineAlcohol                  0.528
LackTeacherTrain                          0.526
ParentalColmplaint                        0.598
TeacherSupport                            0.572
ParentalSupport                           0.563
StudentRetaliation                        0.552
Litigation                                0.617
InconsistenPolicies                       0.523
DistrictStateReprisal                     0.607
EffortsFedPolicySpecialEd                 0.557
EffortsFedPolicy                          0.730
EffortsLocalPolicy                        0.727
DisciplineRecorded                               0.821
IncidentsRecorded                                0.719
IncidentsPolice                                  0.670
OtherActions                                     0.768
ViolentIncident                                  0.770
ViolentIncidentPolice                            0.735
DisciplinePhysicalAttack                         0.837
PercentStudentsLimitedEnglish                           0.557
CrimeLive                                               0.596
CrimeLocated                                            0.663
PercentStudentsLikelyCollege                                  -0.556
PercentStudentsAcademicImp                                    -0.590
ParentOpenHouse                                               -0.672
ParentTeacherConf                                             -0.578
ParentSubjectAreaEvent                                        -0.639
ParentSchoolVolun                                             -0.596
NumberRemovals
PercentSpecialEducation
PercentStudents15Percentile

                      RC3   RC1   RC5   RC2   RC4   RC6
SS loadings         5.553 5.377 4.921 4.681 3.314 3.129
Proportion Var      0.082 0.079 0.072 0.069 0.049 0.046
Cumulative Var      0.082 0.161 0.233 0.302 0.351 0.397
```

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
NumberTypeProblemWeek	0.622					
Bullying	0.575					
SexualHarassment	0.579					
StudentHarassmentOrientation	0.574					
Cyberbullying	0.842					
SchoolEnironmentCyberbullying	0.830					
CyberbullyingStaff	0.726					
NumberTransfersSpecialized		0.632				
TransferSpecialSchoolSpecOff		0.689				
DisciplineDrugs		0.815				
DisciplineAlcohol		0.625				
EffortsFedPolicySpecialEd			0.561			
EffortsFedPolicy			0.896			
EffortsLocalPolicy			0.892			
CrimeLocated				0.538		
IncidentsRecorded					0.582	
IncidentsPolice					0.836	
ViolentIncident					0.580	
ViolentIncidentPolice					0.896	
DisciplineRecorded						0.846
OtherActions						0.748
DisciplinePhysicalAttack						0.897
NumberRemovals						
PercentStudentsLimitedEnglish						
PercentSpecialEducation						
PercentStudents15Percentile						
PercentStudentsLikelyCollege						
PercentStudentsAcademicImp						
ClassroomChanges						
AverageDailyAttendance						
NumberStudentsTransferredTo						
NumberStudentsTransferredFrom		0.475				
NumberCrisesPlan						

Method 2: Factor Rotation

Numeric and Ordinal Data Divide into 6 Components/factors(underlying factor)

1) MisconductType
2) Discipline drug&alcohol
3) Limitation
4) Amount of Crime
5) Student Information
6) Parental Involvement

Factors explained

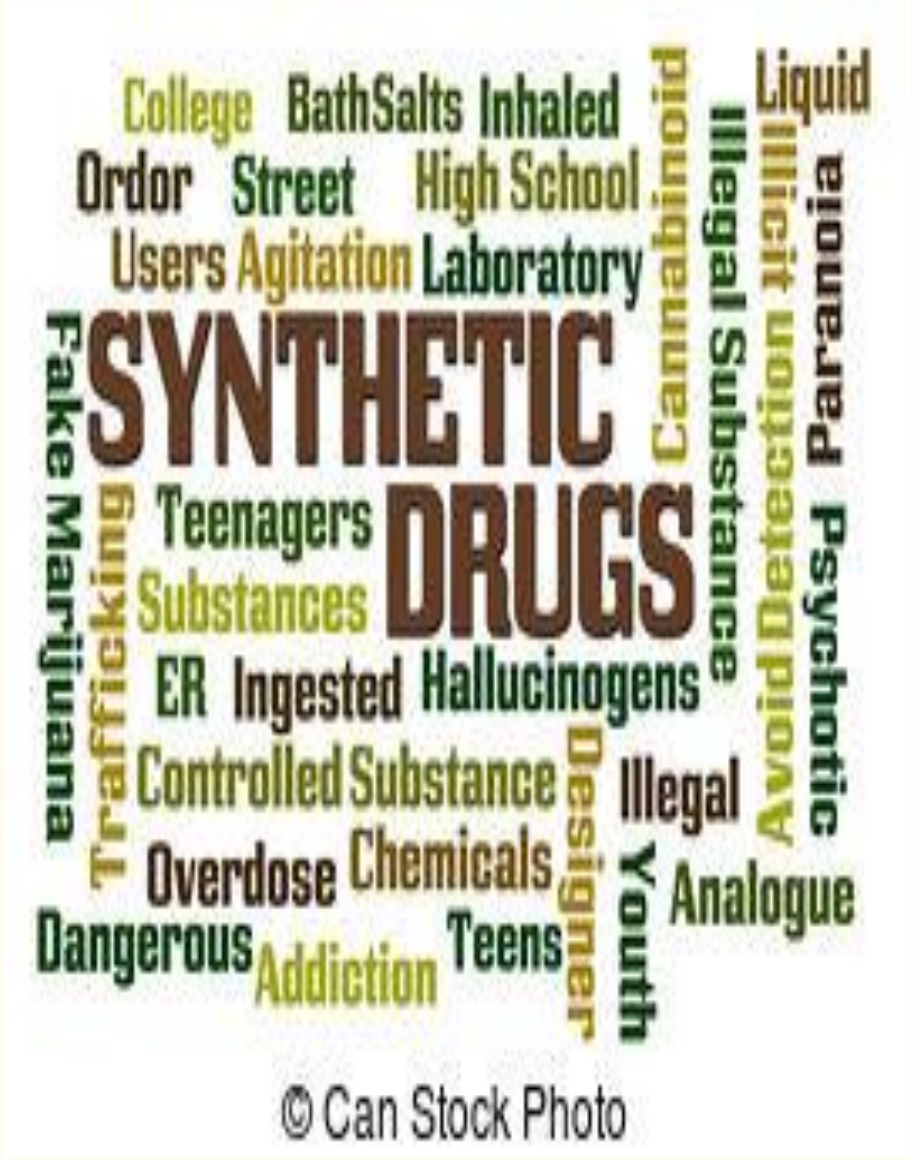
1)MisconductType

- NumberTypeProblemWeek
- RacialEthnicTension
- Bullying
- SexualHarassment
- StudentHarassmentOrientation
- StudentVerbalAbuseTeacher
- StudentDisrespect
- Cyberbullying
- SchoolEnironmentCyberbullying

Factors explained

2) Discipline (Drug&Alcohol)

- NumberTransfersSpecialized
- NumberStudentsTransferredFrom
- TransferSpecialSchoolSpecOff
- DisciplineDrugs
- DisciplineAlcohol



Factors explained

3) Limitation

- LackTeacherTrain
- ParentalColmplaint
- TeacherSupport
- ParentalSupport
- StudentRetaliation
- Litigation
- InconsistenPolicies
- DistrictStateReprisal
- EffortsFedPolicySpecialEd
- EffortsFedPolicy
- EffortsLocalPolicy

Factors explained

4) Amount of Crime

- DisciplineRecorded
- IncidentsRecorded
- IncidentsPolice
- OtherActions
- ViolentIncident
- ViolentIncidentPolice
- DisciplinePhysicalAttack

Factors explained

5)Student Information (location crime level and minority)

- PercentStudentsLimitedEnglish
- CrimeLive
- CrimeLocated

Factors explained

6) Parental Involvement (parents' and students' attitude toward school)

- PercentStudentsLikelyCollege
- PercentStudentsAcademicImp
- ParentOpenHouseb (percentage)
- ParentTacherConf (percentage)
- ParentSubjectAreaEvent (percentage)
- ParentSchoolVolun (percentage)
  - 1 = 0-25%
  - 2 = 26-50%
  - 3 = 51-75%
  - 4 = 76 - 100%
  - 5 = "School does not offer"

Method of Multiple Regression model

1) Summated scales of individual factors by average
2) Combine these factors with 79 binary variables(0 & 1) to make new data set.
3) Used new set to perform multiple regression along with the parameter of interest(StudentsSpecOffenses)

```
> MisconductType = (SchoolMetricOrdinal$NumberTypeProblemWeek +SchoolMetricOrdinal$RacialEthnicTension +
+                     SchoolMetricOrdinal$Bullying +SchoolMetricOrdinal$SexualHarassment +
+                     SchoolMetricOrdinal$StudentHarassmentOrientation +SchoolMetricOrdinal$StudentVerbalAbuseTeacher +
+                     SchoolMetricOrdinal$StudentDisrespect +SchoolMetricOrdinal$Cyberbullying +
+                     SchoolMetricOrdinal$SchoolEnironmentCyberbullying +SchoolMetricOrdinal$CyberbullyingStaff)/ 10
> Discipline = (SchoolMetricOrdinal$NumberTransfersSpecialized +SchoolMetricOrdinal$NumberStudentsTransferredFrom
+               +SchoolMetricOrdinal$TransferSpecialSchoolSpecOff +SchoolMetricOrdinal$DisciplineDrugs
+               +SchoolMetricOrdinal$DisciplineAlcohol)/5
> AmountOfCrime = (SchoolMetricOrdinal$DisciplineRecorded +SchoolMetricOrdinal$IncidentsRecorded
+                  +SchoolMetricOrdinal$IncidentsPolice +SchoolMetricOrdinal$OtherActions
+                  +SchoolMetricOrdinal$ViolentIncident +SchoolMetricOrdinal$ViolentIncidentPolice
+                  +SchoolMetricOrdinal$DisciplinePhysicalAttack)/7
> Limitation = (SchoolMetricOrdinal$LackTeacherTrain +SchoolMetricOrdinal$ParentalColmplaint
+               +SchoolMetricOrdinal$TeacherSupport +SchoolMetricOrdinal$ParentalSupport
+               +SchoolMetricOrdinal$StudentRetaliation +SchoolMetricOrdinal$Litigation
+               +SchoolMetricOrdinal$InconsistenPolicies +SchoolMetricOrdinal$DistrictStateReprisal
+               +SchoolMetricOrdinal$EffortsFedPolicySpecialEd +SchoolMetricOrdinal$EffortsFedPolicy
+               +SchoolMetricOrdinal$EffortsLocalPolicy)/11
> StudentInformation =  (SchoolMetricOrdinal$PercentStudentsLimitedEnglish +SchoolMetricOrdinal$CrimeLive
+                        +SchoolMetricOrdinal$CrimeLocated)/3
> ParentalInvolvement = (SchoolMetricOrdinal$PercentStudentsLikelyCollege +SchoolMetricOrdinal$PercentStudentsAcademicImp
+                        +SchoolMetricOrdinal$ParentOpenHouse +SchoolMetricOrdinal$ParentTeacherConf
+                        +SchoolMetricOrdinal$ParentSubjectAreaEvent +SchoolMetricOrdinal$ParentSchoolVolun )/6
>
```

Fit Factors Into Regression Model

1. Stepwise

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -0.356553   0.500997  -0.712 0.476721
AmountOfCrime            0.044876   0.001118  40.153  < 2e-16 ***
Discipline               0.029721   0.001886  15.759  < 2e-16 ***
MisconductType           0.644558   0.058191  11.077  < 2e-16 ***
Security                 0.501936   0.075153   6.679 2.93e-11 ***
DetentionSaturday        0.391861   0.088781   4.414 1.06e-05 ***
TransferSpecialAvailable 0.252653   0.077051   3.279 0.001055 **
ParentalInvolvement      -0.022613  0.004734  -4.777 1.88e-06 ***
StudentLockers           0.273024   0.079622   3.429 0.000615 ***
SchoolProbation          0.167548   0.072805   2.301 0.021452 *
ExtraCurricDrugTest      -0.283077  0.134652  -2.102 0.035624 *
CommMentalHealth         0.182825   0.068633   2.664 0.007773 **
VisCheckReq              1.369767   0.483657   2.832 0.004659 **
RequireCommService       0.141485   0.069353   2.040 0.041443 *
Access_Doors_MonitLocked -0.285963  0.116069  -2.464 0.013813 *
Ground_Gates             0.184339   0.068689   2.684 0.007328 **
SecurityCameras          0.158664   0.077655   2.043 0.041135 *
CommBus                  -0.142904  0.072882  -1.961 0.050012 .
CorporalPunishment       -0.225424  0.114424  -1.970 0.048934 *
Removal                  0.103594   0.067125   1.543 0.122883
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.652 on 2628 degrees of freedom
Multiple R-squared:  0.669,     Adjusted R-squared:  0.6666
F-statistic: 279.5 on 19 and 2628 DF,  p-value: < 2.2e-16
```

2. Forward

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -0.356553   0.500997  -0.712 0.476721
AmountOfCrime            0.044876   0.001118  40.153  < 2e-16 ***
Discipline               0.029721   0.001886  15.759  < 2e-16 ***
MisconductType           0.644558   0.058191  11.077  < 2e-16 ***
Security                 0.501936   0.075153   6.679 2.93e-11 ***
DetentionSaturday        0.391861   0.088781   4.414 1.06e-05 ***
TransferSpecialAvailable 0.252653   0.077051   3.279 0.001055 **
ParentalInvolvement      -0.022613  0.004734  -4.777 1.88e-06 ***
StudentLockers           0.273024   0.079622   3.429 0.000615 ***
SchoolProbation          0.167548   0.072805   2.301 0.021452 *
ExtraCurricDrugTest      -0.283077  0.134652  -2.102 0.035624 *
CommMentalHealth         0.182825   0.068633   2.664 0.007773 **
VisCheckReq              1.369767   0.483657   2.832 0.004659 **
RequireCommService       0.141485   0.069353   2.040 0.041443 *
Access_Doors_MonitLocked -0.285963  0.116069  -2.464 0.013813 *
Ground_Gates             0.184339   0.068689   2.684 0.007328 **
SecurityCameras          0.158664   0.077655   2.043 0.041135 *
CommBus                  -0.142904  0.072882  -1.961 0.050012 .
CorporalPunishment       -0.225424  0.114424  -1.970 0.048934 *
Removal                  0.103594   0.067125   1.543 0.122883
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.652 on 2628 degrees of freedom
Multiple R-squared:  0.669,     Adjusted R-squared:  0.6666
F-statistic: 279.5 on 19 and 2628 DF,  p-value: < 2.2e-16
```

Method 2: Regression Model

3. Backward

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -0.273210   0.506171  -0.540 0.589409
VisCheckReq               1.365686   0.484076   2.821 0.004820 **
Access_Doors_MonitLocked -0.286584   0.117113  -2.447 0.014467 *
Ground_Gates              0.173665   0.069115   2.513 0.012040 *
RandomMetalDetect         0.239116   0.133279   1.794 0.072911 .
LunchCloseCampus          0.104996   0.073702   1.425 0.154392
ContrabandSweeps         -0.135939   0.095863  -1.418 0.156292
ExtraCurricDrugTest      -0.293813   0.134866  -2.179 0.029454 *
StudentLockers            0.281597   0.080103   3.515 0.000446 ***
SecurityCameras           0.154282   0.077811   1.983 0.047496 *
CellPhoneBan             -0.147986   0.104481  -1.416 0.156779
CommMentalHealth          0.191369   0.068705   2.785 0.005385 **
CommBus                  -0.143464   0.073051  -1.964 0.049647 *
Security                  0.494820   0.075644   6.541 7.30e-11 ***
Removal                   0.100115   0.067324   1.487 0.137120
TransferSpecialAvailable  0.241301   0.077141   3.128 0.001779 **
CorporalPunishment       -0.200533   0.114799  -1.747 0.080786 .
SchoolProbation           0.166541   0.072969   2.282 0.022549 *
DetentionSaturday         0.394998   0.089138   4.431 9.75e-06 ***
RequireCommService        0.137248   0.069424   1.977 0.048153 *
MisconductType            0.637225   0.058350  10.921  < 2e-16 ***
Discipline                0.029416   0.001891  15.554  < 2e-16 ***
AmountOfCrime             0.044779   0.001119  40.007  < 2e-16 ***
ParentalInvolvement      -0.022367   0.004771  -4.688 2.89e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.651 on 2624 degrees of freedom
Multiple R-squared:  0.6699,    Adjusted R-squared:  0.667
F-statistic: 231.5 on 23 and 2624 DF,  p-value: < 2.2e-16
```

Method 2: Regression Model

1. Stepwise
   a. Y = 0.045*AmountOfCrime +0.030*Discipline +0.64*MisconductType +0.50*Security +0.39*DetentionSaturday +0.25*TransferSpecialAvailable -0.023*ParentalInvolvement +0.27*StudentLockers +0.16*SchoolProbation -0.28*ExtraCurricDrugTest +0.18*CommMentalHealth +1.37*VisCheckReq -0.14*RequireCommService -0.29*Access_Doors_MonitLocked +0.18*Ground_Gates +0.15*SecurityCameras -0.14*CommBus -0.23*CorporalPunishment
2. Backward
   a. Y = 0.045*AmountOfCrime +0.030*Discipline +0.64*MisconductType +0.50Security +0.39*DetentionSaturday +0.25*TransferSpecialAvailable -0.023*ParentalInvolvement +0.28*StudentLockers +0.16*SchoolProbation -0.29*ExtraCurricDrugTest +0.19*CommMentalHealth +1.37*VisCheckReq -0.14*RequireCommService -0.29*Access_Doors_MonitLocked +0.17*Ground_Gates +0.15*SecurityCameras -0.14*CommBus -0.23*CorporalPunishment

Y = StudentsSpecOffenses = Total students involved in specified offenses

Method 2: Regression Model

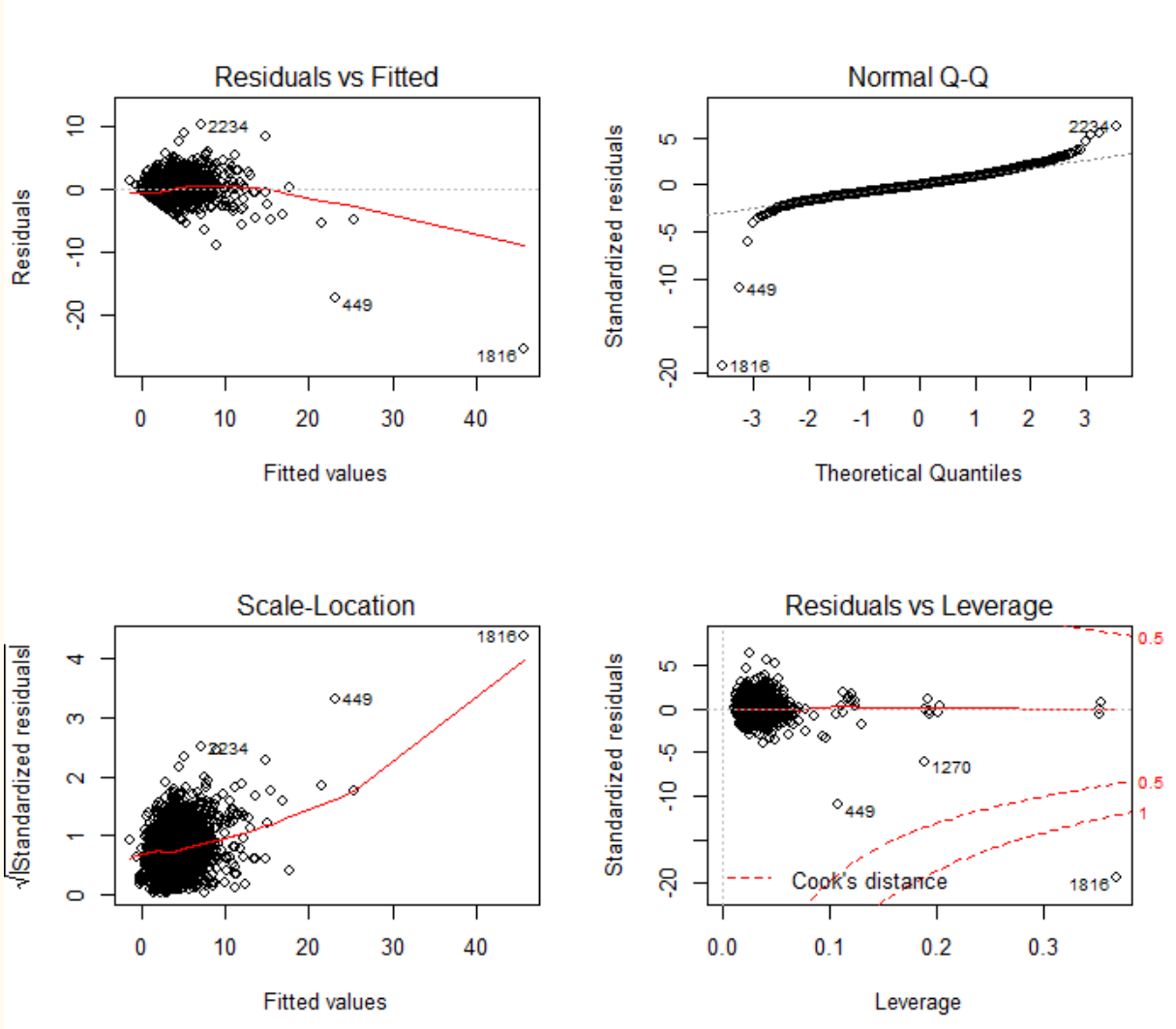
3. cv.glmnet with 10-fold cross validation (set.seed(123456), 75:25 split)

   a. Y = 0.045*AmountOfCrime +0.030*Discipline +0.64*MisconductType +0.48*Security +0.39*DetentionSaturday +0.14*TransferSpecialAvailable -0.023*ParentalInvolvement +0.18*StudentLockers +0.16*SchoolProbation -0.24*ExtraCurricDrugTest +0.12*CommMentalHealth +0*VisCheckReq -0.14*RequireCommService - 0.10*Access_Doors_MonitLocked +0.07*Ground_Gates +0.07*SecurityCameras - 0.003*CommBus -0.13*CorporalPunishment +0.076*InSchoolDisplinaryPlan +0.14*ReferralCounselorAvailable + 0.070*OutsideSuspensionServiceAvailable +0.018*RemovalTutoring + 0.03*CommLawEnforc +0.05* CommJuvenileJustice +0.02*StudentResolvProblems + 0.04*RecreationEnrichmentActiv - 0.03*PreventionCurricInstrucTrain -0.07*NationThreatLevelRedPlan - 0.03*ClassroomTelephones - 0.04*FacultyStaffBadge + 0.05*MetalDetectors

Y = StudentsSpecOffenses = Total students involved in specified offenses

3. Plot of full regression model

```
> yPredict = predict(cvfit, newx=xTest, s="lambda.min")
> dof = length(yTest) - length(c)
> rsePredict = sqrt(sum((yTest - yPredict)^2) / dof)
> rsePredict
[1] 1.888385
```

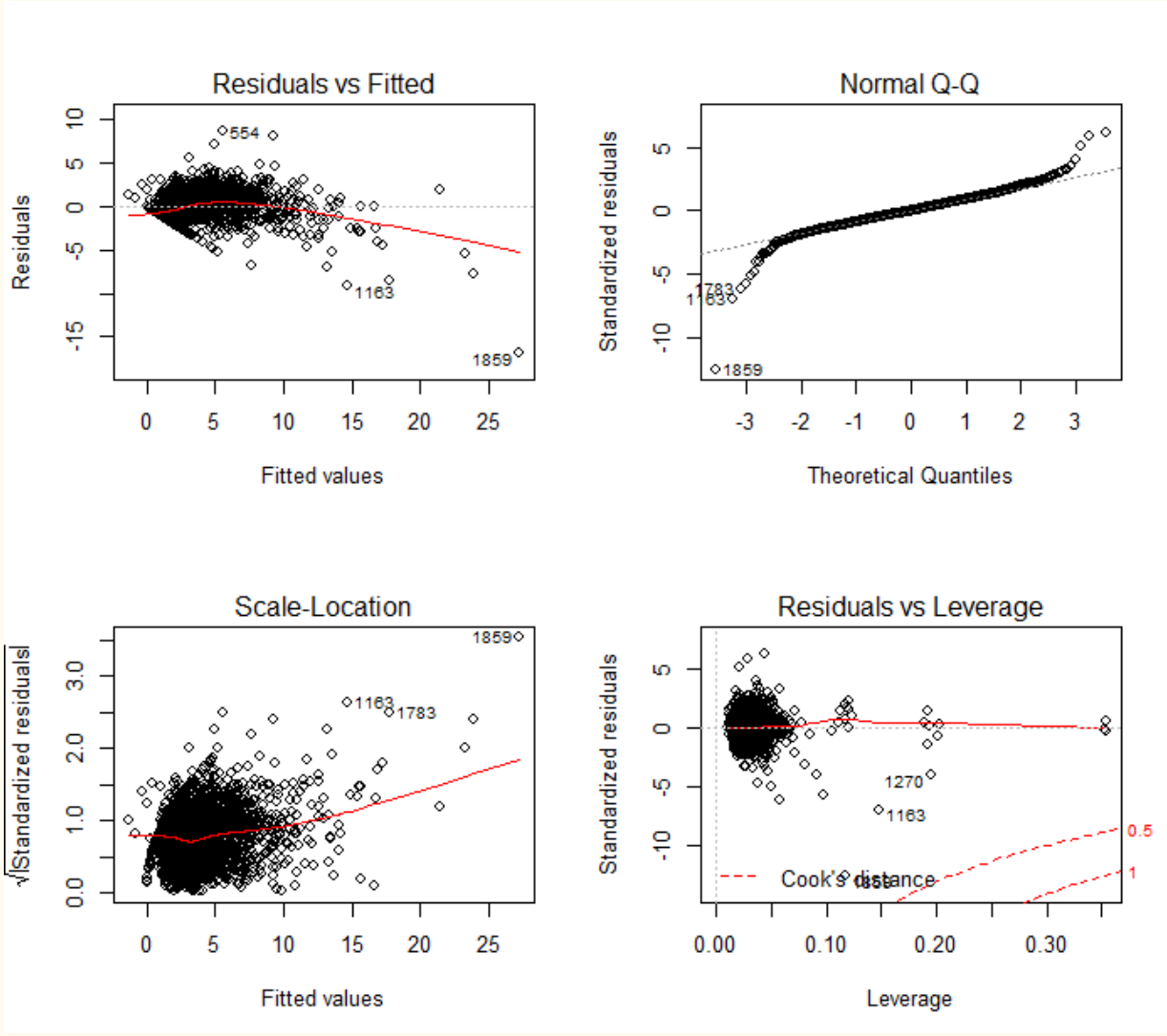


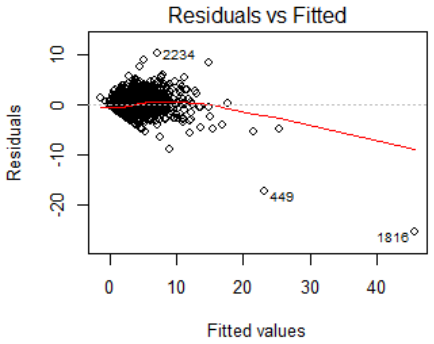
3. Plot after transform

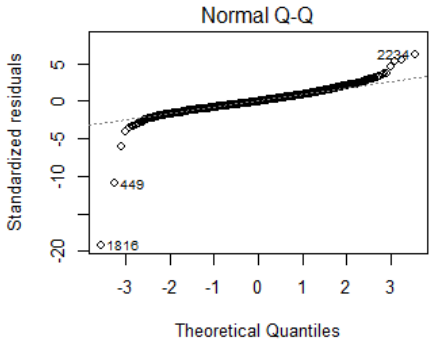
Y into sqrt(Y)

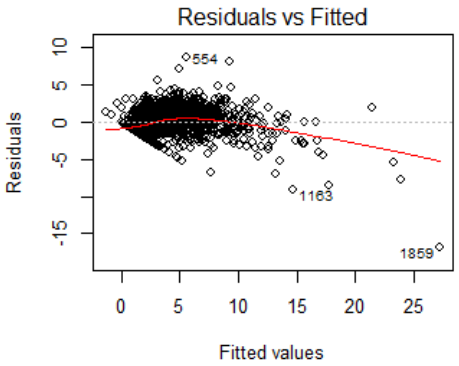
```
> yPredict = predict(cvfit, newx=xTest, s="lambda.min")
> dof = length(yTest) - length(c)
> rsePredict = sqrt(sum((yTest2 - yPredict)^2) / dof)
> rsePredict
```

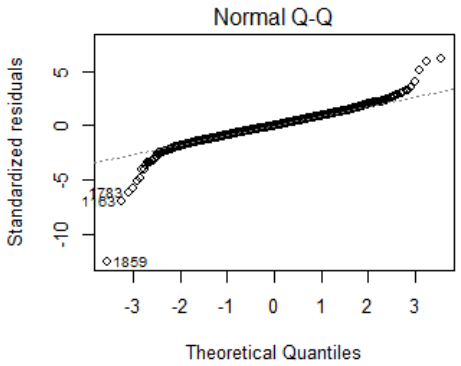
[1] 1.893333

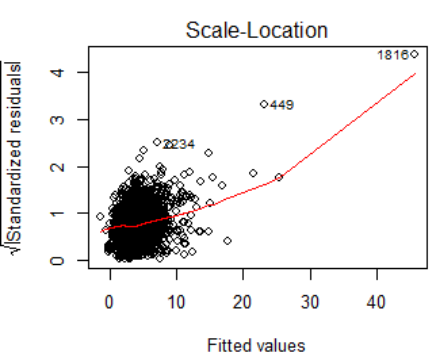


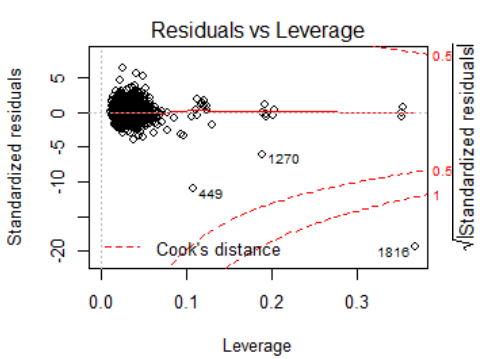

Residuals vs Fitted
Residuals
Fitted values
2234
449
1816

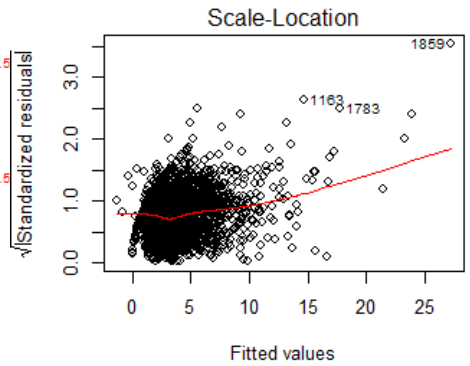

Normal Q-Q
Standardized residuals
Theoretical Quantiles
2234
449
1816

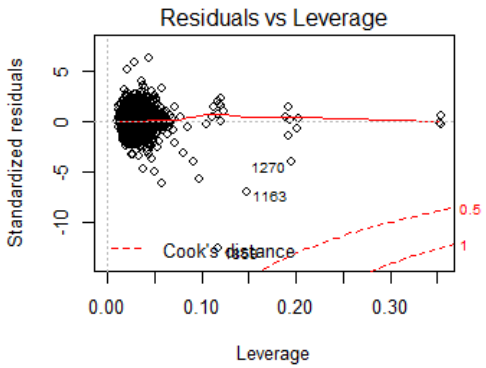

Residuals vs Fitted
Residuals
Fitted values
554
1163
1859


Normal Q-Q
Standardized residuals
Theoretical Quantiles
1783
1163
1859


Scale-Location
√|Standardized residuals|
Fitted values
1816
449
2234


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
1270
449
1816


Scale-Location
√|Standardized residuals|
Fitted values
1859
1163
1783


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
1270
1163

Conclusion

Combination of cv.glmnet, stepwise, backward selection method

a. StudentsSpecOffenses = 0.045*AmountOfCrime +0.030*Discipline +0.64*MisconductType +0.48*Security +0.39*DetentionSaturday +0.14*TransferSpecialAvailable -0.023*ParentalInvolvement +0.18*StudentLockers +0.16*SchoolProbation -0.24*ExtraCurricDrugTest +0.12*CommMentalHealth -0.14*RequireCommService -0.10*Access_Doors_MonitLocked +0.07*Ground_Gates +0.07*SecurityCameras -0.13*CorporalPunishment +0.076*InSchoolDisplinaryPlan +0.14*ReferralCounselorAvailable + 0.070*OutsideSuspensionServiceAvailable

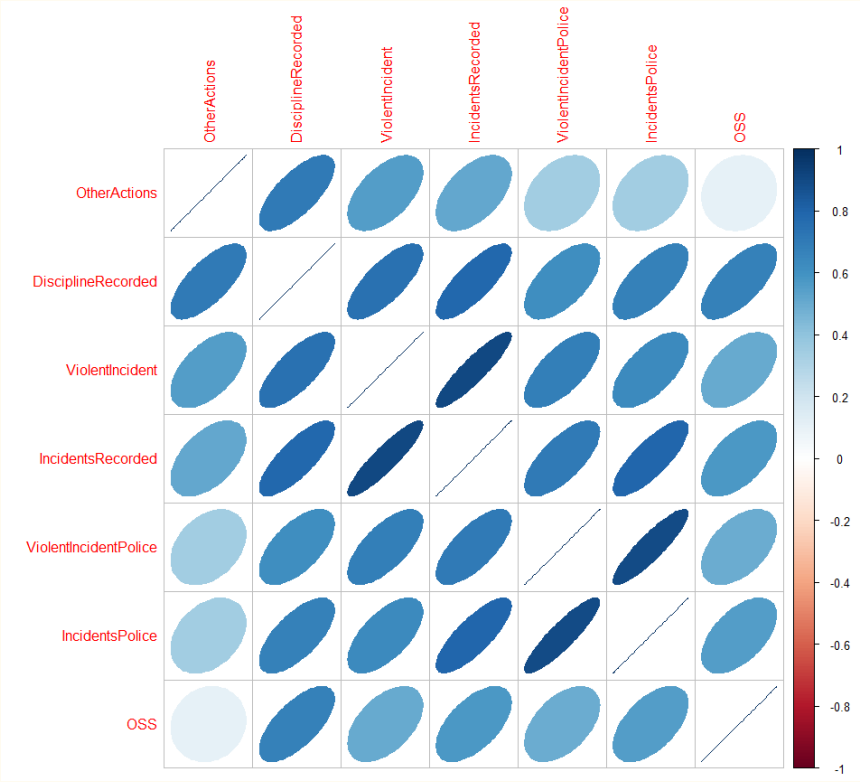
Y = StudentsSpecOffenses = Total students involved in specified offenses

Conclusion(interpretation)

Based on our multiple regression model, in order to reduce the number of student involved crime, the schools needed to emphasize(increase) the Parent's involvement in school, requires drug test in students with extra-curriculum activity, having bad student required for community service, having access controlled locked door, and make corporal punishment available. The bullying, cyberbulling, sexual harassment, abusive attitude toward teachers should be reduced. The drugs are also probromatic. Some of the policies in school includes Detention Saturday, School Probation, have locked monitor gates in ground, security cameras are not effective way to dealing with crimes in school.

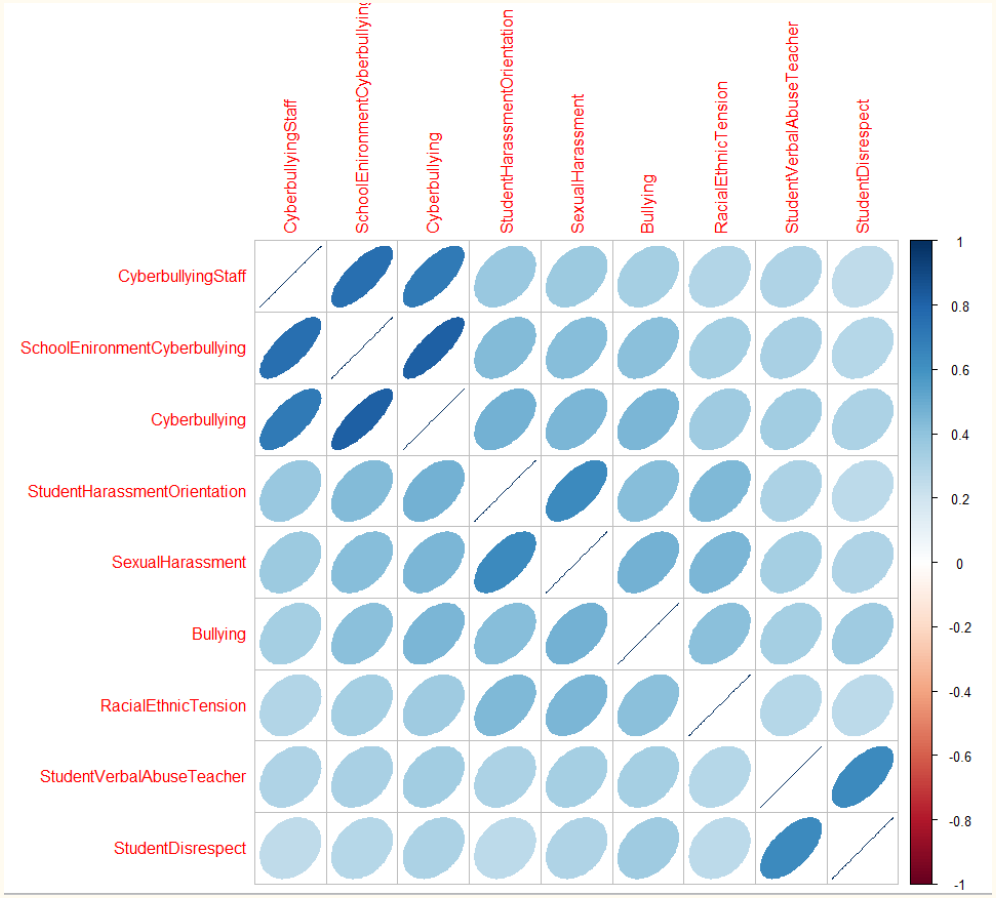
Method 3: Canonical Correlation

sqtVariate_Crime = sqt(OtherActions, DisciplineRecorded,DisciplinePhysicalAttack, ViolentIncident, Incidents Recorded, IncidentsPolice, ViolentIncidentPolice)



Canonical Correlation

Variate_Behavior = sqt(RacialEthnicTension, Bullying, SexualHarassment, StudentHarassmentOrientation, StudentVerbalAbuseTeacher, StudentDisrespect, Cyberbullying, SchoolEnironmentCyberbullying, CyberbullyingStaff)



Select X var in

limitations component as limitationvariate
misconducType component as our behaviour variate
Parental Involvement components as parentalvariate

Canonical Correlation: 4 Variates show significance Behaviour variate

- Correlation of variates:
- X coefficients

	[,1]	[,2]	[,3]	[,4]
RacialEthnicTension	-0.26535096	-0.16941522	-0.2579057	0.64433161
Bullying	0.09201353	0.91466388	-0.4330274	0.05096534
SexualHarassment	-0.20166898	0.09501582	-0.7142202	-0.08056165
StudentHarassmentOrientation	-0.15699487	-0.52990553	-0.0111457	-0.55806691
StudentVerbalAbuseTeacher	-0.56812045	-0.08980886	0.3455336	-0.57323780
StudentDisrespect	-0.18067899	0.43460331	0.4932596	0.40761511
Cyberbullying	-0.15075093	-0.62889858	-0.6918665	0.36992635
SchoolEnironmentCyberbullying	0.07174708	0.34859991	0.5213880	-1.44711029
CyberbullyingStaff	-0.34153959	-0.27036027	0.4111687	1.13448966

Y

	[,1]	[,2]	[,3]	[,4]
DisciplineRecorded	-0.019573180	0.103887038	0.4204339	-0.13228279
IncidentsRecorded	-0.166004477	-0.109252757	-0.3274775	0.15070207
IncidentsPolice	-0.047016219	-0.265152225	-0.3236999	0.10191618
OtherActions	0.008682289	-0.038029684	-0.3954125	-0.23423344
OSS	-0.092571757	-0.194219071	-0.1054968	-0.37405079
ViolentIncident	-0.005421108	0.493757347	0.1538223	0.08846612
ViolentIncidentPolice	-0.048392068	-0.009301732	0.6505142	0.14336725

- Corr.x.scores

	[,1]	[,2]	[,3]	[,4]
RacialEthnicTension	-0.5814495	-0.088748786	0.20284315	0.03689867
Bullying	-0.4796302	0.500553872	0.30521524	0.39447875
SexualHarassment	-0.5814980	-0.067499133	0.33751772	-0.05045682
StudentHarassmentOrientation	-0.5653203	-0.257918998	-0.19749212	0.16398738
StudentVerbalAbuseTeacher	-0.8529910	0.120796393	-0.21327295	0.11071510
StudentDisrespect	-0.7363288	0.395591374	0.04141706	-0.14981320
Cyberbullying	-0.6287177	-0.280246560	0.25178659	0.41780712
SchoolEnironmentCyberbullying	-0.5734180	-0.004777006	-0.11818948	0.25912452
CyberbullyingStaff	-0.6034522	-0.180510851	0.04417807	-0.18879665

	[,1]	[,2]	[,3]	[,4]
DisciplineRecorded	0.6749979	-0.12029764	0.38717173	-0.126625603
IncidentsRecorded	0.9528455	-0.08470838	-0.19676933	0.078456458
IncidentsPolice	0.7739670	0.32724126	-0.03053731	0.428473159
OtherActions	0.3519130	-0.18145318	-0.11838414	-0.211749531
OSS	0.5726405	-0.05030397	0.77098283	-0.003407241
ViolentIncident	0.7689161	-0.50369349	-0.16419458	0.264579975
ViolentIncidentPolice	0.5773096	0.07226947	0.02514911	0.693726344

Significance test

```
> wilksMarsh = ccaWilks(BehaviorVariate,sqtCrimesVariate,ccCrimeBehavior)
> round(wilksMarsh,2)
     WilksL     F df1      df2    p
[1,]   0.66 18.27  63 14829.67 0.00
[2,]   0.91  5.06  48 12959.53 0.00
[3,]   0.97  2.43  35 11082.67 0.00
[4,]   0.98  1.78  24  9193.63 0.01
[5,]   0.99  1.25  15  7277.24 0.23
[6,]   1.00  0.45   8  5274.00 0.89
[7,]   1.00  0.53   3  2638.00 0.66
```

Behavior variate correlates highly with Crime variate

```
> ccCrimeLimitation$cor
[1] 0.36805784 0.13998592 0.10270900
> ccCrimeParents$cor
[1] 0.409520358 0.195610954 0.109452
> ccCrimeBehavior$cor
[1] 0.52989861 0.23990171 0.12603442
```

Conclusion

Combination of cv.glmnet, stepwise, backward selection method

a. StudentsSpecOffenses = 0.045*AmountOfCrime +0.030*Discipline +0.64*MisconductType +0.48*Security +0.39*DetentionSaturday +0.14*TransferSpecialAvailable -0.023*ParentalInvolvement +0.18*StudentLockers +0.16*SchoolProbation -0.24*ExtraCurricDrugTest +0.12*CommMentalHealth -0.14*RequireCommService -0.10*Access_Doors_MonitLocked +0.07*Ground_Gates +0.07*SecurityCameras -0.13*CorporalPunishment +0.076*InSchoolDisplinaryPlan +0.14*ReferralCounselorAvailable + 0.070*OutsideSuspensionServiceAvailable

Y = StudentsSpecOffenses = Total students involved in specified offenses

Conclusion(interpretation)

Based on our multiple regression model, in order to reduce the number of student involved crime, the schools needed to emphasize(increase) the Parent's involvement in school, requires drug test in students with extra-curriculum activity, having bad student required for community service, having access controlled locked door, and make corporal punishment available. The bullying, cyberbulling, sexual harassment, abusive attitude toward teachers should be reduced. Some of the policies in school includes Detention Saturday, School Probation, have locked monitor gates in ground, security cameras are not effective way to dealing with crimes in school.

Future analysis and limitation

Correspondence analysis does not work well.

The k-mean cluster analysis failed miserably. It seemed that we cannot find cluster using metric variables.

If we have more data sample, we may separate our data set based on urbanicity-location of schools and perform analysis based on location of school.

```
   Min     1Q Median     3Q    Max
-63.80 -14.77  -4.83   5.44 519.54

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    2.900e+02  1.897e+01  15.283  < 2e-16 ***
Removal                       -2.200e+01  5.399e+00  -4.075 4.73e-05 ***
RemovalActionUsed             -7.565e+00  1.981e+00  -3.818 0.000138 ***
DateQuestionnnaire             2.621e-06  7.225e-07   3.628 0.000291 ***
PandemicFluPlan                2.964e+00  1.358e+00   2.182 0.029212 *
ExtraCurricDrugTest            9.167e+00  2.469e+00   3.713 0.000209 ***
RequireCommService            -1.800e+01  6.006e+00  -2.997 0.002750 **
RequireCommServiceUsed        -7.320e+00  2.682e+00  -2.730 0.006384 **
SchoolProbation               -3.994e+00  1.381e+00  -2.891 0.003875 **
DetentionSaturday             -2.072e+01  7.526e+00  -2.753 0.005949 **
DetentionSatudayUsed          -7.817e+00  3.578e+00  -2.184 0.029016 *
InSchoolDisplinaryPlan        -1.451e+01  5.919e+00  -2.452 0.014288 *
InSchoolDisciplinaryPlanUsed  -5.900e+00  2.753e+00  -2.143 0.032203 *
OutSchoolDisiplineAvailable   -2.403e+00  1.365e+00  -1.760 0.078442 .
Access_Doors_MonitLocked       5.936e+00  2.171e+00   2.734 0.006295 **
Ground_Gates                  -4.159e+00  1.281e+00  -3.248 0.001179 **
MetalDetectors                -8.334e+00  4.428e+00  -1.882 0.059938 .
RandomMetalDetect             -9.517e+00  2.396e+00  -3.972 7.31e-05 ***
RemovalTutoring               -2.840e+01  4.579e+00  -6.202 6.47e-10 ***
RemovalTutoringAction         -1.072e+01  1.752e+00  -6.119 1.09e-09 ***
TransferSpecialAvailable      -3.548e+01  4.122e+00  -8.608  < 2e-16 ***
TransferSpecialUsed           -1.044e+01  1.572e+00  -6.640 3.82e-11 ***
TransferRegularAvailable      -2.560e+01  5.584e+00  -4.584 4.77e-06 ***
TransferRegularUsed           -9.402e+00  2.175e+00  -4.323 1.60e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.98 on 2587 degrees of freedom
  (36 observations deleted due to missingness)
Multiple R-squared:  0.192,     Adjusted R-squared:  0.1849
F-statistic: 26.73 on 23 and 2587 DF,  p-value: < 2.2e-16
```