Predicting Behavioral Misconduct in Chicago Public Schools
Technical Report

Overview

The goal of this project was to utilize data from a CPS annual report to predict the rate of behavioral misconduct at each school. Our methods revolved around the application of multiple linear regression to uncover relationships between independent variables and misconduct rate. Independent variables used were restricted to the columns available in the CPS report. Models were testing using simple linear regression as well as several transforms of the dependent variable. The final model generated a fit that proved relatively useful for predictions, though opportunities for improvement existed as well.

Quantitative Variables

- Safety Score: the safety score reflects how secure students and faculty feel in their school(1-100), where a score of 100 is the best.
- Environment Score: An evaluation of the schools environment(1-100), where a score of 100 is the best.
- Instruction Score: Evaluation of instruction practices (1-100) , where a score of 100 is the best.
- Teachers Score: evaluation of the teacher performance(1-100), where a score of 100 is the best.
- Parent Engagement Score:Parent Perception/Engagement score from parent survey (1-100), where a score of 100 is the best.
- Rate of Misconducts (per 100 students): Acts of behavioral misconduct that occur per 100 students

Qualitative Variables

- CPS Performance Policy Level **:** How well the school is doing overall accordance to all variables in the study and other variables left out of the survey such as administrator assessments and other assessments not calculated in the survey. Level 1 being in good standing- level 3 being the lowest evaluation

Exploratory Analysis

To start the exploratory data analysis, a matrix of pairwise scatter plots were created. These plots were used to judge if any obvious correlation existed between the predictors and the dependent variable. They were also used to judge obvious relationships between independent variables themselves. There appeared to be a relationship between the dependent variable and at least one predictor, Safety. In addition, there was evidence of a nonlinear relationship observed between these variables. This implied that a transform of variables might be

necessary. In response to this, models that featured a square root transform of the dependent variable were compared against linear models with no transforms.

Evidence of a transform needed was also supported by univariate plots of the dependent variable. A histogram and box plot of misconduct that can be found in the appendix show a strong skew to the right in the data. For the box plot, there are several outliers to the upper extreme. We did not choose to remove any of these observations since the population we are working with is finite. In other words, there is no possibility that these outliers belong to a population other than the one of interest.

Collinearity of Variables

Only a few of the independent variables showed much correlation between each other. In particular, Environment and Instruction shared an r-value of 0.8. This did not end up being an issue in model selection since these variables were not shown as being significant predictors. Outside of these two independent variables, no others were related strongly enough to warrant removal from our model based on collinearity.

Model Selection

The final model was chosen based on a backward selection method with all available independent variables. A p-value of 0.05 was chosen to determine if a parameter was significant in predicting the dependent variable. The adjusted-R^2 statistic was also assessed for each iteration of backward selection to assess the effect of adding or removing variables. To execute this method, an initial model was created with every independent variable included as a predictor. The output of this initial model showed several predictors as not being significant in the model. The predictor with the least significant p-value was removed from the model, and the regression model was re-run. This methodology was repeated until all independent variables were shown as significant. In the case of the model with no transforms, this resulted in the inclusion of the variables for Safety and Level 3 (dummy).

The backward selection method was also applied to models incorporating a square-root transform on the dependent variable. These models can be found in the appendix under Regression Models. The iterative process of removing predictors was followed again, resulting in a model that also included Safety and Level 3 (dummy). The transformed model results in general reflected a better fit to the data, judging by the Adjusted-R^2 statistic and the p-values of each independent variable. For this reason, we chose to continue our analysis with the model including the square root transform. The regression equation from this model can be found below:

$$\text{sqrt(Misconduct\_Rate)} = -0.05177 * \text{Safety} + 0.94631 * \text{Level 3} + 6.15981$$

There were thoughts to also test a model that included a log transform of the dependent variable. However, since 0 was included in the range of misconduct rates, a log transform would not work as the function is undefined at that point.

Residual Analysis

The residual plot depicting residual values vs the predicted values (found in the appendix) shows that the points are Homoscedastic and unbiased.

For the first partial residual plot of Level 3, we do not learn anything significant since the predictor is a dummy variable.

In the second partial residual plot for safety, we see that the points are relatively randomly scattered or Homoscedastic but they are biased because the points are moving in a downward trend. This means that the variable safety score predicts lower rates of misconduct as the safety score increase.

The normal plot shows a linear relationship in the data. This suggests that the data is normally distributed. The tail of the plot at the upward end does lie above the linear trend line, suggesting that the distribution of residuals is skewed slightly to the right.

Influential Observations

We ran an influential points test in R and found out that there are several points of interest. The full table of influence points can be found in the appendix. Based on the output, the only point we felt warranted possible removal was observation 47. This point was influential based on the DFFITS measure, and also had the highest Cook's Distance of any influential point by a significant margin. The model was tested briefly without this point, and there were no significant changes to the adjusted r-squared values or variable selection. For this reason, we decided to continue with the observation included in the dataset.

Improvement Opportunities

We are satisfied with our model with the data that we utilized but there are adjustments that could result in a better model predicting misconduct in Chicago public schools. We had to remove many observations because they were missing so much data. Not only that but most of the schools that were removed were the High Schools and Middle Schools. This made the proportion of elementary schools higher than High Schools and Middle Schools.

If we could get more accurate data, there would be opportunity to split the analysis into three models for elementary, middle school and high school, respectively. Minors in middle school could be different from minors in highschool and they may be more sensitive to some variables compared to others depending on what class of school they are in.

There is also opportunity to test additional interaction terms and judge how they effect the model. This in particular could be aided by gathering additional attribute data concerning each school. Standard statistics such as size, location, etc. were not included in our model, and could potentially correlate with rates of misconduct.

Strength of Model Predictors

Since our final model only included a dummy variable and one continuous independent variable, judging the strength of predictors was fairly simple. In this case, the safety score has a much more significant effect on the model's prediction of misconduct than the dummy variable. Although the coefficient for safety is smaller in magnitude than that of Level 3, the wider range of safety scores (0-100) allows it to influence the model to a much greater degree.

Assessing Predictive Usefulness with Cross-Validation

To assess our model's predictive capabilities, we applied a ten-fold cross validation technique. In this process, the model was trained and tested ten times. In addition, the data set was split up such that each observation was only used in testing exactly once. For each run, or fold, the sum of squares and mean sum of squares of the predicted residuals were calculated and stored. A summary table of both statistics can be found in the appendix. Although the PRESS statistic shows that the model technically holds up under cross-validation, there is a high amount of variability between the sum of squares predicted residual for each fold. This suggests that there is either evidence of over-fitting, or that some observations are influencing the model to a high degree. In either case, this leads to the conclusion that the model might not be consistent in predicting for a brand new, independent set of data.

APPENDIX

Data set

SchoolID	Level2	Level3	Safety	Environment		Instruction		Teachers	
ParentEngagement		Misconduct_Rate							
609679	1	0	59	53	51	44	46	9.7	
609682	1	0	30	32	19	34	43	17.3	
609694	0	1	51	44	41	42	50	7.5	
609695	0	1	51	43	42	41	43	8.7	
609711		1	0	22	39	42	31	49	63.6
609733	0	1	30	34	34	33	44	14.7	
609735	1	0	34	45	37	35	51	13	
609753	0	0	87	49	47	40	52	5.2	
609755	0	0	95	69	67	46	53	1.2	
609772	1	0	56	44	46	44	53	5	
609775	0	1	48	64	58	35	46	32.9	
609780	1	0	38	27	35	26	50	5.3	
609782	0	0	95	75	60	65	61	4.5	
609786	1	0	32	38	32	16	46	52.1	
609789	1	0	35	70	81	41	51	5.6	
609790	0	1	25	28	34	59	51	25.3	
609792	1	0	55	50	47	34	52	0.2	
609798	1	0	64	67	69	51	53	21.3	
609800	0	1	46	55	66	56	47	6.5	
609804	0	0	52	47	48	53	50	10.5	
609805	0	1	35	44	53	34	52	44.5	
609806	0	1	31	41	50	54	49	75.9	
609812	1	0	36	58	78	73	46	59	
609813	1	0	20	79	82	19	48	14.1	
609815	0	1	33	55	50	58	50	28	
609817	1	0	59	33	38	41	50	3.9	
609821	0	0	61	85	99	71	54	29.4	
609826	1	0	36	34	20	31	46	24.3	
609828	0	0	71	68	51	63	56	4.3	
609829	1	0	50	39	38	53	50	0.2	
609832	0	1	63	35	38	57	49	51.8	
609833	0	1	25	40	44	25	44	30.7	
609834	0	0	45	52	66	43	47	3.2	
609836	0	0	87	70	64	47	66	2.3	
609837	0	1	48	37	63	31	49	6.2	
609842	1	0	38	45	56	36	50	3.9	
609844	0	1	13	33	35	62	46	27	
609849	1	0	76	57	36	61	52	13.8	

609852	0	0	70	53	51	48	50	2.9
609854	0	0	86	99	99	99	49	1.7
609857	0	0	64	58	47	44	49	6.1
609861	1	0	45	39	61	51	52	8.4
609862	1	0	22	40	54	53	50	24.9
609865	1	0	61	57	49	15	52	16.4
609866	0	0	87	65	39	56	58	1.4
609869	1	0	31	54	32	50	53	29
609873	1	0	27	51	52	31	49	100.5
609875	1	0	58	44	34	46	49	33.8
609879	1	0	44	27	30	54	50	35.6
609880	0	0	99	99	99	99	52	2.8
609881	1	0	38	66	52	79	55	20.3
609884	0	0	67	60	78	36	50	2.1
609888	1	0	57	68	73	83	56	12.7
609893	0	0	74	69	79	51	58	10.8
609894	0	1	59	57	46	16	50	43.3
609899	0	0	78	66	50	62	52	9.3
609901	0	0	99	51	53	70	56	3.4
609903	1	0	48	38	41	49	49	10.3
609910	1	0	54	48	27	30	49	5.7
609912	1	0	55	40	38	59	55	1.7
609913	0	1	44	52	56	28	46	47.5
609917	0	1	29	17	43	50	48	29
609918	1	0	42	47	37	8	44	10
609919	0	1	13	37	55	35	47	37.1
609920	0	1	44	48	48	43	47	3.3
609921	0	1	42	19	22	69	55	2.7
609922	1	0	75	67	66	53	55	1.3
609925	1	0	58	42	45	50	59	4.4
609927	0	1	11	20	22	46	46	33.6
609929	1	0	46	63	63	39	45	10.1
609933	0	1	39	35	27	29	45	34.8
609935	1	0	49	37	32	40	50	5.9
609937	0	0	60	50	51	55	56	10.2
609938	1	0	32	29	28	17	45	16.8
609942	1	0	75	68	52	69	57	5.6
609943	1	0	34	48	52	59	49	54.2
609947	1	0	49	31	33	41	48	27.4
609954	0	0	44	39	43	59	49	3.2
609956	0	0	79	56	59	71	53	5
609958	0	0	46	30	45	14	49	87.3
609959	0	0	32	37	37	48	43	0.3

609960	1	0	55	51	56	65	55	9.8
609963	1	0	73	60	59	48	57	12.5
609964	1	0	33	29	43	33	44	7.5
609967	1	0	56	50	52	56	49	6.7
609971	1	0	51	61	61	36	47	8.7
609972	1	0	43	57	49	30	46	23.4
609973	1	0	60	43	42	79	46	2.8
609974	0	0	92	61	56	84	58	1.2
609978	0	1	22	1	1	29	42	90
609985	1	0	29	26	63	52	51	13.4
609988	0	0	45	46	43	45	52	4.3
609990	0	0	84	34	36	46	54	12.7
609993	0	0	56	45	37	48	53	10.4
609995	0	0	64	55	58	70	55	3.7
609996	1	0	53	50	35	51	54	4.7
610003	0	1	51	45	54	83	48	46.9
610006	0	0	77	50	49	54	49	12.2
610009	0	0	65	40	42	33	49	3.1
610011	0	0	64	62	66	41	52	8.1
610012	0	1	50	62	41	26	49	16
610015	1	0	36	56	53	43	48	5.3
610019	1	0	62	69	56	59	58	8.2
610028	0	1	19	22	13	19	47	31.2
610031	0	1	44	43	38	72	52	14.3
610033	0	0	99	62	52	43	53	7
610036	0	1	21	36	64	48	42	37.7
610038	0	0	99	74	66	70	56	2
610039	0	0	68	66	75	44	50	1.8
610041	0	0	57	49	56	45	47	2.2
610052	0	1	28	31	30	24	44	44.4
610054	1	0	72	53	39	48	50	5.4
610056	0	1	37	44	58	36	51	31
610061	1	0	52	62	65	47	50	26
610065	0	1	31	47	40	42	47	95.7
610067	0	1	23	37	54	61	48	9.1
610072	0	1	28	58	55	36	47	10
610073	0	0	99	95	95	90	56	12.2
610077	0	1	28	32	39	32	44	19.6
610081	0	0	74	53	63	63	55	1.3
610086	1	0	31	42	44	19	47	76.6
610091	1	0	31	52	33	45	52	35.6
610093	0	1	42	27	28	36	52	5
610096	1	0	35	40	48	72	48	9.4

610098	1	0	37	37	35	36	48	24.8
610099	0	0	99	69	68	32	57	0.3
610102	0	1	44	49	49	32	46	25.4
610104	0	0	64	29	27	39	47	6.9
610107	0	0	83	67	58	94	54	3.2
610119	1	0	51	50	61	55	52	7.6
610120	0	0	51	43	47	37	44	1.2
610125	1	0	57	57	62	56	50	15.9
610129	1	0	50	35	37	34	48	5.1
610132	0	0	99	66	88	41	55	0
610138	0	0	68	54	57	23	47	1.7
610143	0	1	48	37	26	15	45	37.2
610147	0	0	81	68	62	64	53	3
610148	0	0	54	63	52	50	45	11.1
610154	0	0	40	48	43	37	45	95.1
610157	1	0	57	41	39	29	49	7.4
610167	1	0	40	38	41	63	49	14.4
610170	1	0	59	59	63	72	52	8.9
610175	1	0	45	40	39	41	49	0.3
610176	0	1	32	43	51	42	52	40.8
610182	0	0	78	67	62	62	52	0.8
610183	1	0	35	57	71	56	50	24.1
610184	1	0	66	70	67	43	46	5.9
610188	0	0	46	65	71	59	49	22.8
610193	0	1	27	35	40	33	47	27.3
610194	0	1	41	44	48	34	48	12.3
610196	1	0	61	62	60	59	49	6
610197	0	0	63	59	71	55	51	3.2
610201	0	0	70	47	40	57	50	25.8
610202	0	1	6	30	41	44	51	65.5
610206	0	0	67	46	44	71	60	5.3
610208	0	1	15	41	46	30	48	39.1
610209	1	0	43	28	37	56	51	22.5
610212	0	0	66	66	71	50	46	2.3
610215	1	0	49	48	47	34	47	7.1
610217	0	0	49	50	46	56	48	1.3
610219	0	0	58	33	43	76	52	5.6
610221	0	0	48	81	66	44	45	5.7
610225	0	1	45	41	28	21	52	11.1
610230	0	0	99	77	67	45	55	1
610231	0	0	67	65	60	46	56	9.2
610238	1	0	54	67	63	32	53	18.8
610239	1	0	29	37	46	42	46	44.2

610242	0	0	50	55	59	54	53	13.7
610243	0	0	52	44	54	88	43	4.6
610249	0	0	99	99	99	81	57	9.2
610251	0	0	69	78	81	83	56	29.5
610257	0	0	53	66	63	95	52	73.9
610263	1	0	42	53	42	21	51	28.2
610269	1	0	56	67	41	42	52	8.8
610271	1	0	35	42	67	40	46	17.2
610274	1	0	67	61	75	66	51	1.4
610281	0	0	54	74	84	76	46	15.7
610282	0	1	31	56	59	20	47	52.8
610287	1	0	48	52	40	70	51	28.5
610298	0	0	79	51	67	63	52	0.3
610300	1	0	30	48	54	47	46	31.5
610313	0	1	57	37	40	64	53	29
610319	1	0	60	47	44	65	43	5.5
610339	0	1	23	35	42	42	44	34.3
610345	0	1	26	2	1	41	44	70.5
610347	1	0	36	50	51	59	46	64.2
610352	0	0	61	57	58	74	53	0.7
610353	1	0	58	59	72	55	51	1.9
610354	1	0	67	53	51	47	50	20.1
610365	0	1	53	53	39	46	56	25.8
610368	0	1	28	16	30	48	46	27.5
610381	0	1	41	42	43	15	51	16.2
610383	0	1	43	46	44	65	53	16.4
610392	1	0	51	49	47	49	50	4
610499	1	0	60	51	46	48	54	2.9
610502	1	0	41	38	34	35	54	4.5
610503	0	0	60	84	72	55	66	7.2
610506	1	0	45	55	53	68	58	8.2
610520	0	0	66	36	52	79	55	5.6

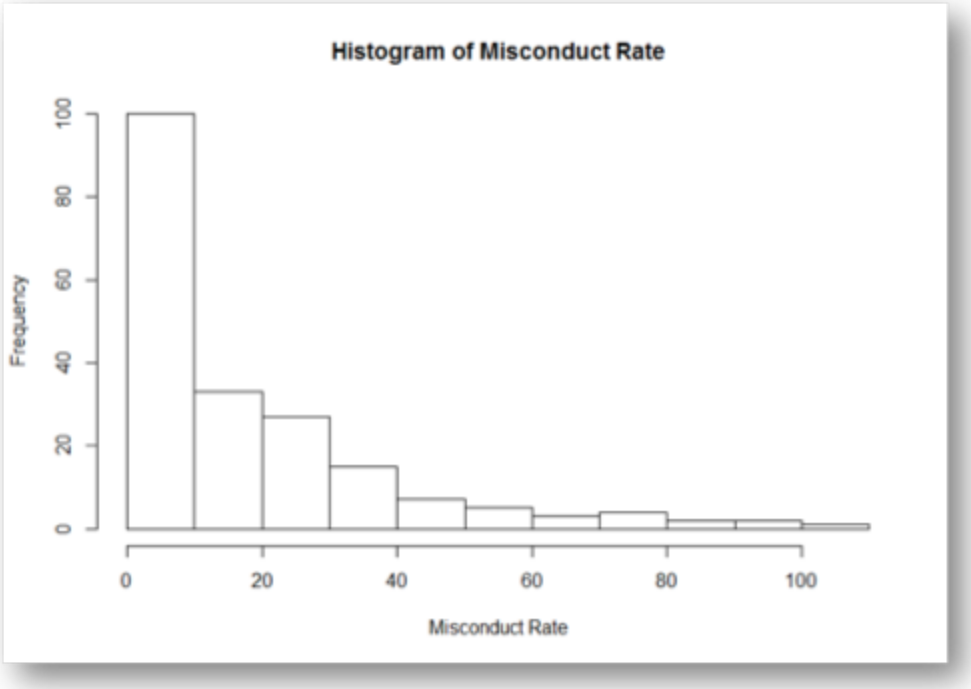


Figure 1: Histogram of Misconduct Rate

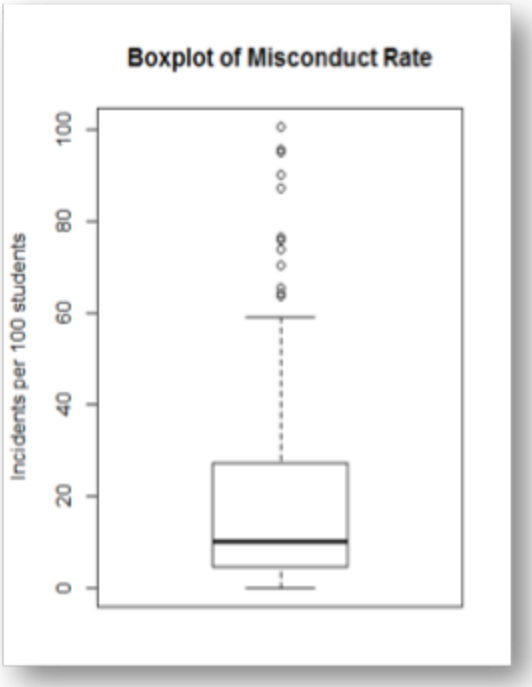


Figure 2: Boxplot of Misconduct Rate

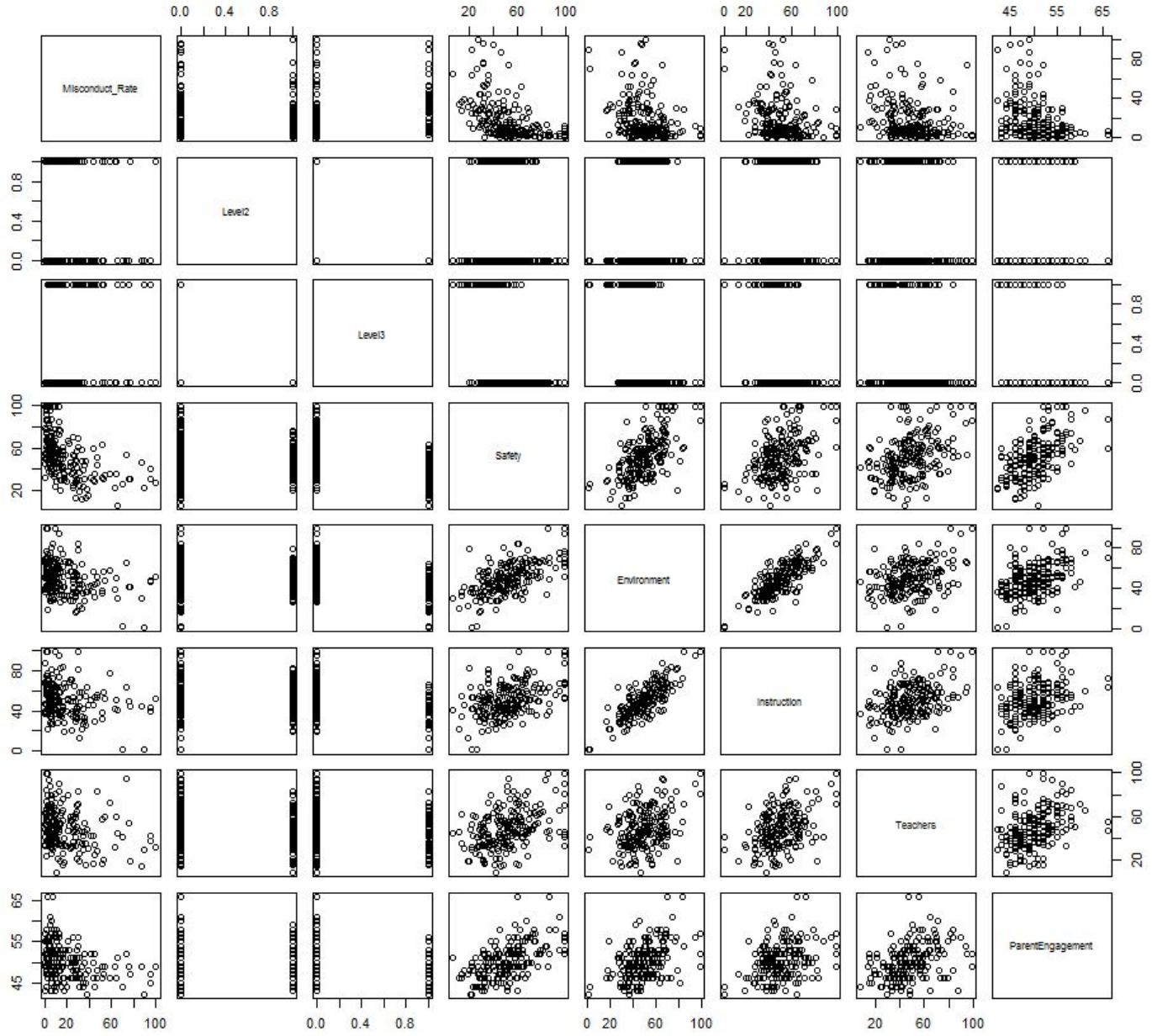


Figure 3: Pairwise Scatterplot Matrix of Variables

	SchoolID	Level2	Level3	Safety	Environment	Instruction	Teachers	ParentEngagement	Misconduct_Rate
SchoolID	1.000000000	-0.05701635	-0.02783134	0.0186476	0.03621627	0.05688891	0.1364474	0.07321530	-0.008145182
Level2	-0.057016354	1.00000000	-0.49000659	-0.1826030	-0.05116098	-0.09738581	-0.1097906	-0.04465002	-0.043506934
Level3	-0.027831339	-0.49000659	1.00000000	-0.4740345	-0.40953721	-0.32233272	-0.2564666	-0.30507401	0.365133151
Safety	0.018647599	-0.18260302	-0.47403448	1.0000000	0.62099896	0.45451426	0.4090436	0.57605397	-0.508882671
Environment	0.036216270	-0.05116098	-0.40953721	0.6209990	1.00000000	0.80899624	0.3790526	0.44558608	-0.306010367
Instruction	0.056888913	-0.09738581	-0.32233272	0.4545143	0.80899624	1.00000000	0.4145813	0.31386494	-0.249406233
Teachers	0.136447392	-0.10979059	-0.25646657	0.4090436	0.37905260	0.41458127	1.0000000	0.39490504	-0.226416255
ParentEngagement	0.073215299	-0.04465002	-0.30507401	0.5760540	0.44558608	0.31386494	0.3949050	1.00000000	-0.326233751
Misconduct_Rate	-0.008145182	-0.04350693	0.36513315	-0.5088827	-0.30601037	-0.24940623	-0.2264163	-0.32623375	1.000000000

Figure 4: Correlation Matrix of Variables

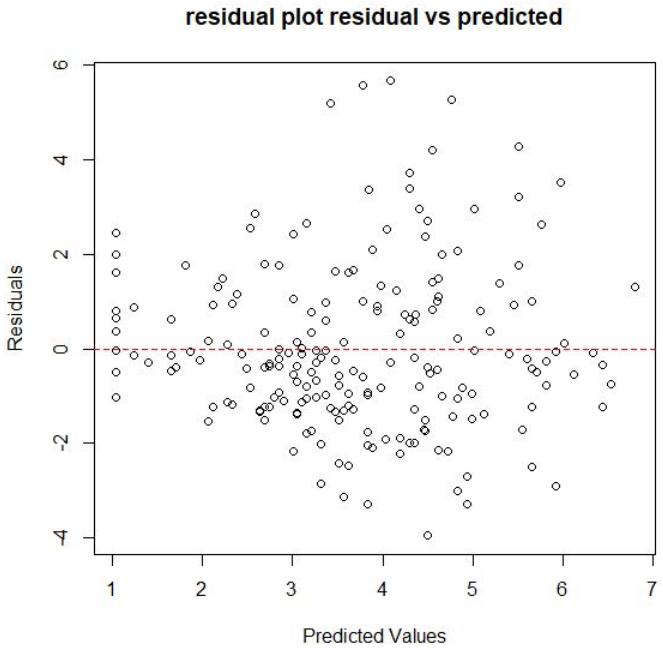


Figure 5: Model Residual Plot

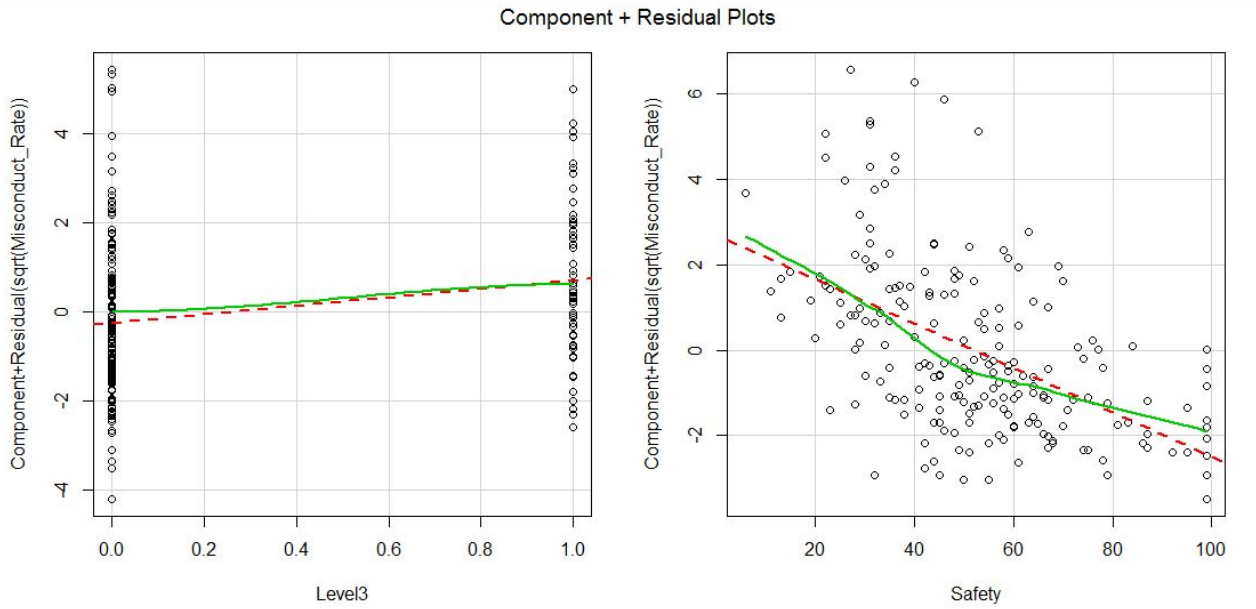


Figure 6: Partial Residual Plots

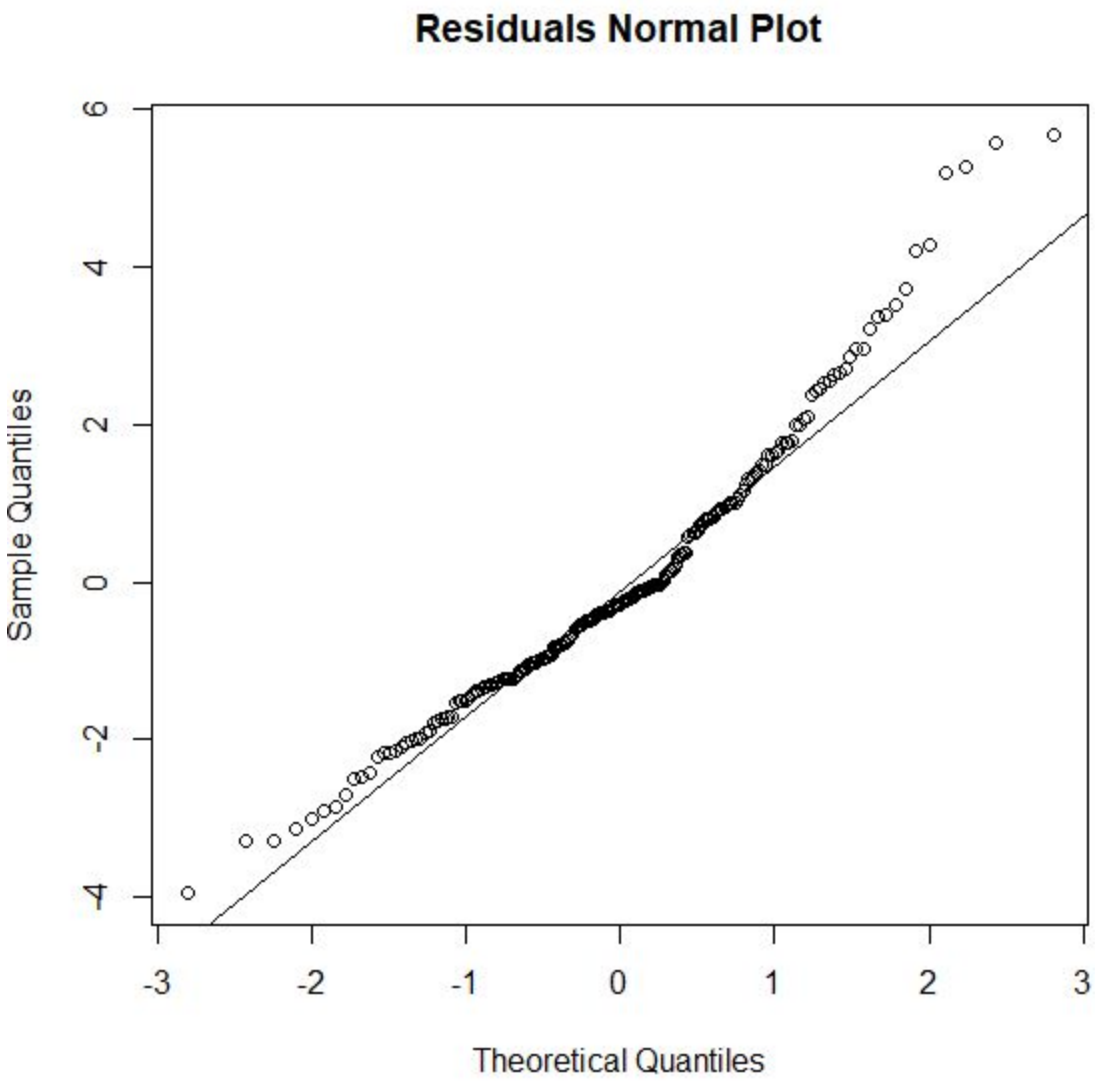


Figure 7: Normal Plot of Model Residuals

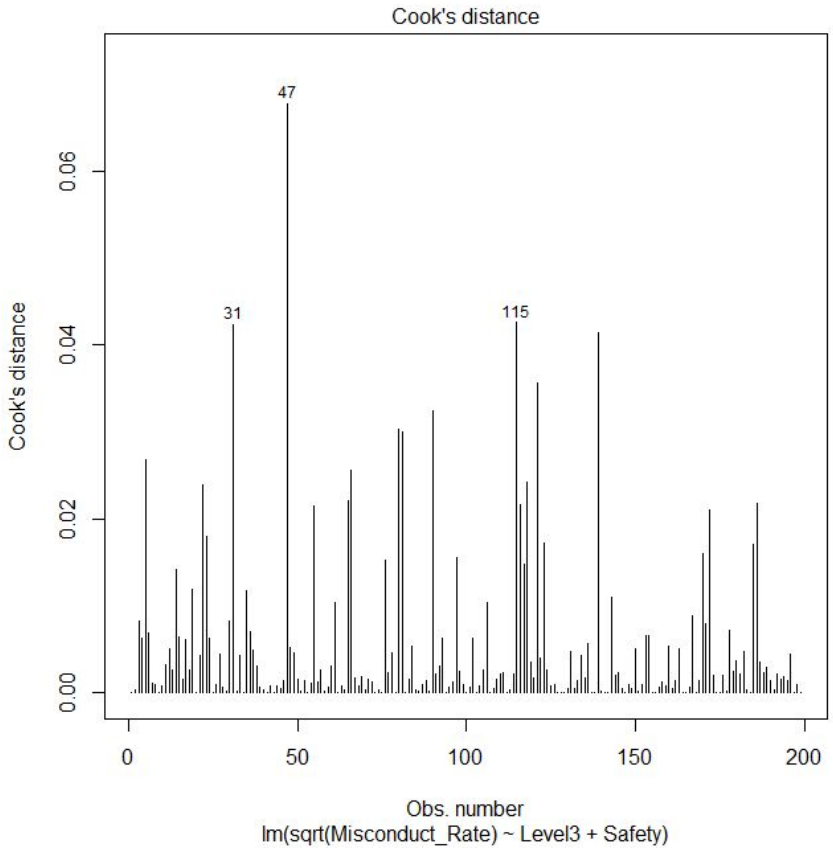
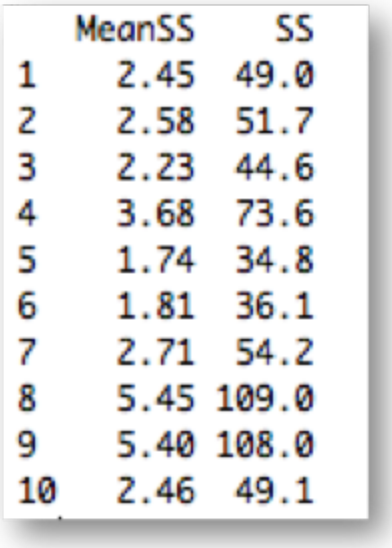


Figure 8: Cook's Distance D

```
> print(summary(influence.measures(lm.cps6)))
Potentially influential observations of
        lm(formula = sqrt(Misconduct_Rate) ~ Level3 + Safety, data = cps_data) :

    dfb.1_ dfb.Lvl3 dfb.Sfty dffit   cov.r   cook.d hat
47   0.45  -0.29    -0.38     0.46_*  0.90_*  0.07   0.02
50  -0.05   0.02     0.06     0.07    1.05_*  0.00   0.03
57  -0.06   0.02     0.08     0.09    1.05_*  0.00   0.03
80   0.23  -0.19    -0.15     0.31    0.87_*  0.03   0.01
81  -0.29   0.20     0.24    -0.30    0.95_*  0.03   0.02
108 -0.03   0.01     0.04     0.04    1.05_*  0.00   0.03
115  0.04   0.25    -0.04     0.36    0.94_*  0.04   0.02
121  0.31  -0.21    -0.26     0.33    0.94_*  0.04   0.02
126  0.04  -0.01    -0.05    -0.05    1.05_*  0.00   0.03
139  0.31  -0.23    -0.23     0.36    0.87_*  0.04   0.01
164  0.00   0.00     0.00     0.00    1.05_*  0.00   0.03
172  0.14  -0.14    -0.05     0.26    0.89_*  0.02   0.01
```

Figure 9: Table of Influential Points



	MeanSS	SS
1	2.45	49.0
2	2.58	51.7
3	2.23	44.6
4	3.68	73.6
5	1.74	34.8
6	1.81	36.1
7	2.71	54.2
8	5.45	109.0
9	5.40	108.0
10	2.46	49.1

Figure 10: 10-Fold Cross Validation Output

Regression Models

```
Call:
lm(formula = sqrt(Misconduct_Rate) ~ Level2 + Level3 + Safety +
    Environment + Instruction + Teachers + ParentEngagement,
    data = cps_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0725 -1.1110 -0.2352  0.8993  6.0362

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.636072   1.705313   3.891 0.000138 ***
Level2          -0.045358   0.350618  -0.129 0.897205
Level3           0.973637   0.446189   2.182 0.030321 *
Safety          -0.058069   0.010204  -5.691  4.7e-08 ***
Environment      0.027757   0.015311   1.813 0.071431 .
Instruction     -0.016784   0.013367  -1.256 0.210790
Teachers         0.002280   0.008165   0.279 0.780395
ParentEngagement -0.015653   0.037064  -0.422 0.673267
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.749 on 191 degrees of freedom
Multiple R-squared:  0.3704,  Adjusted R-squared:  0.3474
F-statistic: 16.06 on 7 and 191 DF,  p-value: < 2.2e-16

> print(summary(lm.cps2))

Call:
lm(formula = sqrt(Misconduct_Rate) ~ Level3 + Safety + Environment +
    Instruction + Teachers + ParentEngagement, data = cps_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0334 -1.1232 -0.2546  0.9014  6.0694

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.586508   1.657456   3.974  0.00010 ***
Level3           1.012526   0.328879   3.079  0.00238 **
Safety          -0.057471   0.009073  -6.334 1.66e-09 ***
```

```
Environment       0.027586   0.015215   1.813  0.07139 .
Instruction      -0.016538   0.013197  -1.253  0.21167
Teachers          0.002381   0.008107   0.294  0.76932
ParentEngagement -0.016031   0.036854  -0.435  0.66405
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.745 on 192 degrees of freedom
Multiple R-squared:  0.3704,  Adjusted R-squared:  0.3507
F-statistic: 18.83 on 6 and 192 DF,  p-value: < 2.2e-16

> print(summary(lm.cps3))

Call:
lm(formula = sqrt(Misconduct_Rate) ~ Level3 + Safety + Environment +
    Instruction + ParentEngagement, data = cps_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0043 -1.1259 -0.2485  0.8969  5.9880

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.540978   1.646278   3.973   0.0001 ***
Level3           1.008285   0.327784   3.076   0.0024 **
Safety          -0.057082   0.008955  -6.374 1.32e-09 ***
Environment      0.027177   0.015116   1.798   0.0737 .
Instruction     -0.015587   0.012763  -1.221   0.2235
ParentEngagement -0.013748   0.035939  -0.383   0.7025
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.74 on 193 degrees of freedom
Multiple R-squared:  0.3701,  Adjusted R-squared:  0.3538
F-statistic: 22.68 on 5 and 193 DF,  p-value: < 2.2e-16

> print(summary(lm.cps4))

Call:
lm(formula = sqrt(Misconduct_Rate) ~ Level3 + Safety + Environment +
    Instruction, data = cps_data)

Residuals:
```

```
   Min      1Q  Median      3Q     Max
-3.9373 -1.1109 -0.2678  0.9019  5.9847

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.944137   0.524252  11.338  < 2e-16 ***
Level3       1.011159   0.326976   3.092  0.00228 **
Safety      -0.058449   0.008193  -7.134 1.89e-11 ***
Environment  0.026459   0.014965   1.768  0.07864 .
Instruction -0.015345   0.012720  -1.206  0.22913
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.737 on 194 degrees of freedom
Multiple R-squared:  0.3696,  Adjusted R-squared:  0.3566
F-statistic: 28.44 on 4 and 194 DF,  p-value: < 2.2e-16

> print(summary(lm.cps5))

Call:
lm(formula = sqrt(Misconduct_Rate) ~ Level3 + Safety + Environment,
   data = cps_data)

Residuals:
   Min      1Q  Median      3Q     Max
-3.8820 -1.1286 -0.2275  0.8712  5.8082

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.788704   0.508767  11.378  < 2e-16 ***
Level3       1.017484   0.327315   3.109  0.00216 **
Safety      -0.057423   0.008158  -7.039 3.23e-11 ***
Environment  0.012934   0.009925   1.303  0.19407
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.739 on 195 degrees of freedom
Multiple R-squared:  0.3649,  Adjusted R-squared:  0.3551
F-statistic: 37.35 on 3 and 195 DF,  p-value: < 2.2e-16
```

> print(summary(lm.cps6)) #cps6 = Best Model.

Call:

lm(formula = sqrt(Misconduct_Rate) ~ Level3 + Safety, data = cps_data)

Residuals:

Min	1Q	Median	3Q	Max
-3.955	-1.191	-0.294	0.948	5.663

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.159811	0.422345	14.585	< 2e-16	***
Level3	0.946311	0.323300	2.927	0.00383	**
Safety	-0.051774	0.006923	-7.478	2.46e-12	***

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 1.742 on 196 degrees of freedom
Multiple R-squared: 0.3594, Adjusted R-squared: 0.3528
F-statistic: 54.97 on 2 and 196 DF, p-value: < 2.2e-16