DEP Team Project Regression Analysis

Mark Roberts

12/4/2019

Import Report Card data

```
library(readxl)
Rcard<- read_excel('/Users/markroberts/Desktop/Data Engineering Platforms/Chicago_Public
_Schools_-_Progress_Report_Cards__2011-2012_.xlsx')


head(Rcard)
```

```
## # A tibble: 6 x 79
##   `School ID` `Name of School` `Elementary, Mi… `Street Address` City
##         <dbl> <chr>            <chr>            <chr>            <chr>
## 1      609966 Charles G Hammo… ES               2819 W 21st Pl   Chic…
## 2      610539 Marvin Camras E… ES               3000 N Mango Ave Chic…
## 3      609852 Eliza Chappell … ES               2135 W Foster A… Chic…
## 4      609835 Daniel R Camero… ES               1234 N Monticel… Chic…
## 5      610521 Sir Miles Davis… ES               6730 S Paulina … Chic…
## 6      609818 Luther Burbank … ES               2035 N Mobile A… Chic…
## # … with 74 more variables: State <chr>, `ZIP Code` <dbl>, `Phone
## #   Number` <chr>, Link <chr>, `Network Manager` <chr>, `Collaborative
## #   Name` <chr>, `Adequate Yearly Progress Made?` <chr>, `Track
## #   Schedule` <chr>, `CPS Performance Policy Status` <chr>, `CPS
## #   Performance Policy Level` <chr>, `Healthy Schools Certified?` <chr>,
## #   `Safety Icon` <chr>, `Safety Score` <dbl>, `Family Involvement
## #   Icon` <chr>, `Family Involvement Score` <chr>, `Environment
## #   Icon` <chr>, `Environment Score` <dbl>, `Instruction Icon` <chr>,
## #   `Instruction Score` <dbl>, `Leaders Icon` <chr>, `Leaders
## #   Score` <chr>, `Teachers Icon` <chr>, `Teachers Score` <chr>, `Parent
## #   Engagement Icon` <chr>, `Parent Engagement Score` <chr>, `Parent
## #   Environment Icon` <chr>, `Parent Environment Score` <chr>, `Average
## #   Student Attendance` <dbl>, `Rate of Misconducts (per 100
## #   students)` <dbl>, `Average Teacher Attendance` <dbl>, `Individualized
## #   Education Program Compliance Rate` <dbl>, `Pk-2 Literacy %` <chr>,
## #   `Pk-2 Math %` <chr>, `Gr3-5 Grade Level Math %` <chr>, `Gr3-5 Grade
## #   Level Read %` <chr>, `Gr3-5 Keep Pace Read %` <chr>, `Gr3-5 Keep Pace
## #   Math %` <chr>, `Gr6-8 Grade Level Math %` <chr>, `Gr6-8 Grade Level
## #   Read %` <chr>, `Gr6-8 Keep Pace Math%` <chr>, `Gr6-8 Keep Pace Read
## #   %` <chr>, `Gr-8 Explore Math %` <chr>, `Gr-8 Explore Read %` <chr>,
## #   `ISAT Exceeding Math %` <dbl>, `ISAT Exceeding Reading %` <dbl>, `ISAT
## #   Value Add Math` <dbl>, `ISAT Value Add Read` <dbl>, `ISAT Value Add
## #   Color Math` <chr>, `ISAT Value Add Color Read` <chr>, `Students Taking
## #   Algebra %` <chr>, `Students Passing Algebra %` <chr>, `9th Grade
## #   EXPLORE (2009)` <chr>, `9th Grade EXPLORE (2010)` <chr>, `10th Grade
## #   PLAN (2009)` <chr>, `10th Grade PLAN (2010)` <chr>, `Net Change
## #   EXPLORE and PLAN` <chr>, `11th Grade Average ACT (2011)` <chr>, `Net
## #   Change PLAN and ACT` <chr>, `College Eligibility %` <chr>, `Graduation
## #   Rate %` <chr>, `College Enrollment Rate %` <chr>, `College Enrollment
## #   (number of students)` <dbl>, `General Services Route` <dbl>, `Freshman
## #   on Track Rate %` <chr>, `RCDTS Code` <dbl>, X_COORDINATE <dbl>,
## #   Y_COORDINATE <dbl>, Latitude <dbl>, Longitude <dbl>, `Community Area
## #   Number` <dbl>, `Community Area Name` <chr>, Ward <dbl>, `Police
## #   District` <dbl>, Location <chr>
```

Remove all rows containing "NDA" from College Enrollment Rate % Column

```
data<-Rcard[!grepl("NDA", Rcard$`College Enrollment Rate %`),]
head(data)
```

```
## # A tibble: 6 x 79
##   `School ID` `Name of School` `Elementary, Mi… `Street Address` City
##        <dbl> <chr>            <chr>            <chr>           <chr>
## 1      609680 Walter Payton C… HS               1034 N Wells St  Chic…
## 2      609722 Manley Career A… HS               2935 W Polk St   Chic…
## 3      609749 Northside Colle… HS               5501 N Kedzie A… Chic…
## 4      610244 Michele Clark A… HS               5101 W Harrison… Chic…
## 5      610394 Uplift Communit… HS               900 W Wilson Ave Chic…
## 6      609725 Morgan Park Hig… HS               1744 W Pryor Ave Chic…
## # … with 74 more variables: State <chr>, `ZIP Code` <dbl>, `Phone
## #   Number` <chr>, Link <chr>, `Network Manager` <chr>, `Collaborative
## #   Name` <chr>, `Adequate Yearly Progress Made?` <chr>, `Track
## #   Schedule` <chr>, `CPS Performance Policy Status` <chr>, `CPS
## #   Performance Policy Level` <chr>, `Healthy Schools Certified?` <chr>,
## #   `Safety Icon` <chr>, `Safety Score` <dbl>, `Family Involvement
## #   Icon` <chr>, `Family Involvement Score` <chr>, `Environment
## #   Icon` <chr>, `Environment Score` <dbl>, `Instruction Icon` <chr>,
## #   `Instruction Score` <dbl>, `Leaders Icon` <chr>, `Leaders
## #   Score` <chr>, `Teachers Icon` <chr>, `Teachers Score` <chr>, `Parent
## #   Engagement Icon` <chr>, `Parent Engagement Score` <chr>, `Parent
## #   Environment Icon` <chr>, `Parent Environment Score` <chr>, `Average
## #   Student Attendance` <dbl>, `Rate of Misconducts (per 100
## #   students)` <dbl>, `Average Teacher Attendance` <dbl>, `Individualized
## #   Education Program Compliance Rate` <dbl>, `Pk-2 Literacy %` <chr>,
## #   `Pk-2 Math %` <chr>, `Gr3-5 Grade Level Math %` <chr>, `Gr3-5 Grade
## #   Level Read %` <chr>, `Gr3-5 Keep Pace Read %` <chr>, `Gr3-5 Keep Pace
## #   Math %` <chr>, `Gr6-8 Grade Level Math %` <chr>, `Gr6-8 Grade Level
## #   Read %` <chr>, `Gr6-8 Keep Pace Math%` <chr>, `Gr6-8 Keep Pace Read
## #   %` <chr>, `Gr-8 Explore Math %` <chr>, `Gr-8 Explore Read %` <chr>,
## #   `ISAT Exceeding Math %` <dbl>, `ISAT Exceeding Reading %` <dbl>, `ISAT
## #   Value Add Math` <dbl>, `ISAT Value Add Read` <dbl>, `ISAT Value Add
## #   Color Math` <chr>, `ISAT Value Add Color Read` <chr>, `Students Taking
## #   Algebra %` <chr>, `Students Passing Algebra %` <chr>, `9th Grade
## #   EXPLORE (2009)` <chr>, `9th Grade EXPLORE (2010)` <chr>, `10th Grade
## #   PLAN (2009)` <chr>, `10th Grade PLAN (2010)` <chr>, `Net Change
## #   EXPLORE and PLAN` <chr>, `11th Grade Average ACT (2011)` <chr>, `Net
## #   Change PLAN and ACT` <chr>, `College Eligibility %` <chr>, `Graduation
## #   Rate %` <chr>, `College Enrollment Rate %` <chr>, `College Enrollment
## #   (number of students)` <dbl>, `General Services Route` <dbl>, `Freshman
## #   on Track Rate %` <chr>, `RCDTS Code` <dbl>, X_COORDINATE <dbl>,
## #   Y_COORDINATE <dbl>, Latitude <dbl>, Longitude <dbl>, `Community Area
## #   Number` <dbl>, `Community Area Name` <chr>, Ward <dbl>, `Police
## #   District` <dbl>, Location <chr>
```

```
View(data)
```

Convert College Enrollment Rate into a numeric

```
data$`College Enrollment.Rate %`<- as.numeric(as.character(data$`College Enrollment Rate
%`))
data$`9th Grade EXPLORE (2010)`<- as.numeric(as.character(data$`9th Grade EXPLORE (2010)
`))
```

```
## Warning: NAs introduced by coercion
```

```
data$`11th Grade Average ACT (2011)`<-as.numeric(as.character(data$`11th Grade Average A
CT (2011)`))
data$`Average Student Attendance`<-as.numeric(as.character(data$`Average Student Attenda
nce`))
data$`College Enrollment Rate %`<-as.numeric(as.character(data$`College Enrollment Rate
%`))

str(data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    77 obs. of  80 variables:
##  $ School ID                               : num  609680 609722 609749 610244
610394 ...
##  $ Name of School                          : chr  "Walter Payton College Prep
aratory High School" "Manley Career Academy High School" "Northside College Preparatory
High School" "Michele Clark Academic Prep Magnet High School" ...
##  $ Elementary, Middle, or High School      : chr  "HS" "HS" "HS" "HS" ...
##  $ Street Address                          : chr  "1034 N Wells St" "2935 W P
olk St" "5501 N Kedzie Ave" "5101 W Harrison St" ...
##  $ City                                    : chr  "Chicago" "Chicago" "Chicag
o" "Chicago" ...
##  $ State                                   : chr  "IL" "IL" "IL" "IL" ...
##  $ ZIP Code                                : num  60610 60612 60625 60644 606
40 ...
##  $ Phone Number                            : chr  "(773) 534-0034" "(773) 534
-6900" "(773) 534-3954" "(773) 534-6250" ...
##  $ Link                                    : chr  "http://schoolreports.cps.e
du/SchoolProgressReport_Eng/Spring2011Eng_609680.pdf" "http://schoolreports.cps.edu/Scho
olProgressReport_Eng/Spring2011Eng_609722.pdf" "http://schoolreports.cps.edu/SchoolProgr
essReport_Eng/Spring2011Eng_609749.pdf" "http://schoolreports.cps.edu/SchoolProgressRepo
rt_Eng/Spring2011Eng_610244.pdf" ...
##  $ Network Manager                         : chr  "North-Northwest Side High
School Network" "West Side High School Network" "North-Northwest Side High School Networ
k" "West Side High School Network" ...
##  $ Collaborative Name                      : chr  "NORTH-NORTHWEST SIDE COLLA
BORATIVE" "WEST SIDE COLLABORATIVE" "NORTH-NORTHWEST SIDE COLLABORATIVE" "WEST SIDE COLL
ABORATIVE" ...
##  $ Adequate Yearly Progress Made?          : chr  "Yes" "No" "Yes" "No" ...
##  $ Track Schedule                          : chr  "Standard" "Standard" "Stan
dard" "Standard" ...
##  $ CPS Performance Policy Status           : chr  "Not on Probation" "Probati
on" "Not on Probation" "Probation" ...
##  $ CPS Performance Policy Level            : chr  "Level 1" "Level 3" "Level
1" "Level 3" ...
##  $ Healthy Schools Certified?              : chr  "No" "No" "No" "No" ...
##  $ Safety Icon                             : chr  "Very Strong" "Average" "Ve
ry Strong" "NDA" ...
##  $ Safety Score                            : num  98 41 99 NA 50 40 41 20 13
51 ...
##  $ Family Involvement Icon                 : chr  "NDA" "Weak" "NDA" "NDA"
...
##  $ Family Involvement Score                : chr  "NDA" "39" "NDA" "NDA" ...
##  $ Environment Icon                        : chr  "Very Strong" "Average" "Ve
ry Strong" "NDA" ...
##  $ Environment Score                       : num  80 43 99 NA 61 25 42 18 28
49 ...
##  $ Instruction Icon                        : chr  "Strong" "Weak" "Very Stron
g" "NDA" ...
##  $ Instruction Score                       : num  77 31 88 NA 62 38 43 29 28
47 ...
##  $ Leaders Icon                            : chr  "NDA" "Very Weak" "NDA" "ND
A" ...
##  $ Leaders Score                           : chr  "NDA" "19" "NDA" "NDA" ...
```

```
##  $ Teachers Icon                                 : chr  "NDA" "Weak" "NDA" "NDA"
...
##  $ Teachers Score                                : chr  "NDA" "32" "NDA" "NDA" ...
##  $ Parent Engagement Icon                        : chr  "NDA" "NDA" "Strong" "Avera
ge" ...
##  $ Parent Engagement Score                       : chr  "NDA" "NDA" "57" "48" ...
##  $ Parent Environment Icon                       : chr  "NDA" "NDA" "Strong" "Weak"
...
##  $ Parent Environment Score                      : chr  "NDA" "NDA" "62" "36" ...
##  $ Average Student Attendance                    : num  93.4 66.8 95.7 79.1 88.7 8
4.2 92.3 75.2 84.8 91.6 ...
##  $ Rate of Misconducts (per 100 students)        : num  0.7 19.7 2.8 24.4 8.1 11.1
16.2 5.8 47.1 4 ...
##  $ Average Teacher Attendance                    : num  96.1 95.4 96.8 95.4 95.9 9
4.6 95.6 94.3 95.4 96 ...
##  $ Individualized Education Program Compliance Rate: num  100 98.4 98.4 100 98.3 99.5
97.6 99.6 88 100 ...
##  $ Pk-2 Literacy %                               : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Pk-2 Math %                                   : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Gr3-5 Grade Level Math %                      : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Gr3-5 Grade Level Read %                      : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Gr3-5 Keep Pace Read %                        : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Gr3-5 Keep Pace Math %                        : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Gr6-8 Grade Level Math %                      : chr  "NDA" "NDA" "NDA" "19.39999
9999999999" ...
##  $ Gr6-8 Grade Level Read %                      : chr  "NDA" "NDA" "NDA" "22.8"
...
##  $ Gr6-8 Keep Pace Math%                         : chr  "NDA" "NDA" "NDA" "36.5"
...
##  $ Gr6-8 Keep Pace Read %                        : chr  "NDA" "NDA" "NDA" "52.7"
...
##  $ Gr-8 Explore Math %                           : chr  "NDA" "NDA" "NDA" "3.6" ...
##  $ Gr-8 Explore Read %                           : chr  "NDA" "NDA" "NDA" "10.7"
...
##  $ ISAT Exceeding Math %                         : num  NA NA NA 1.7 23.3 52 NA NA
NA NA ...
##  $ ISAT Exceeding Reading %                      : num  NA NA NA 2.6 2.3 38.2 NA NA
NA NA ...
##  $ ISAT Value Add Math                           : num  NA NA NA -1.2 0.8 -1.8 NA N
A NA NA ...
##  $ ISAT Value Add Read                           : num  NA NA NA -0.7 -0.2 -1 NA NA
NA NA ...
##  $ ISAT Value Add Color Math                     : chr  "NDA" "NDA" "NDA" "Red" ...
##  $ ISAT Value Add Color Read                     : chr  "NDA" "NDA" "NDA" "Yellow"
...
##  $ Students Taking  Algebra %                    : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ Students Passing  Algebra %                   : chr  "NDA" "NDA" "NDA" "NDA" ...
##  $ 9th Grade EXPLORE (2009)                      : chr  "21.2" "12.2" "22.4" "14"
...
##  $ 9th Grade EXPLORE (2010)                      : num  21.8 11.9 22.2 13.3 13.9 1
4.6 13.8 13.3 11.2 13.4 ...
##  $ 10th Grade PLAN (2009)                        : chr  "23.1" "13.3" "24.5" "14.4"
...
##  $ 10th Grade PLAN (2010)                        : chr  "23.2" "13" "24.7" "14.5"
```

```
...
##  $ Net Change EXPLORE and PLAN                 : chr  "2" "0.8" "2.29999999999999
98" "0.5" ...
##  $ 11th Grade Average ACT (2011)               : num  27 13.8 28.8 15.7 16.9 17.8
16.3 15.2 14.1 16.2 ...
##  $ Net Change PLAN and ACT                     : chr  "3.9" "0.5" "4.3" "1.3" ...
##  $ College Eligibility %                       : chr  "96.4" "6.7" "98" "13" ...
##  $ Graduation Rate %                           : chr  "96.9" "49" "97.6" "87.4"
...
##  $ College Enrollment Rate %                   : num  82.4 51.9 90.7 60.4 65 78.5
73.2 50 45.1 48.1 ...
##  $ College Enrollment (number of students)     : num  881 599 1053 726 473 ...
##  $ General Services Route                      : num  33 37 31 36 32 49 42 44 46
37 ...
##  $ Freshman on Track Rate %                    : chr  "90.7" "59.3" "95.9" "78.2"
...
##  $ RCDTS Code                                  : num  1.5e+14 1.5e+14 1.5e+14 1.5
e+14 1.5e+14 ...
##  $ X_COORDINATE                                : num  1174485 1156777 1154091 114
2209 1169403 ...
##  $ Y_COORDINATE                                : num  1907490 1896187 1936414 189
6793 1930781 ...
##  $ Latitude                                    : num  41.9 41.9 42 41.9 42 ...
##  $ Longitude                                   : num  -87.6 -87.7 -87.7 -87.8 -8
7.7 ...
##  $ Community Area Number                       : num  8 27 13 25 3 75 38 70 69 30
...
##  $ Community Area Name                         : chr  "NEAR NORTH SIDE" "EAST GAR
FIELD PARK" "NORTH PARK" "AUSTIN" ...
##  $ Ward                                        : num  27 28 40 24 46 19 3 18 8 22
...
##  $ Police District                             : num  18 11 17 15 19 22 2 8 6 10
...
##  $ Location                                    : chr  "(41.90155157, -87.6345374
4)" "(41.87091163, -87.69988652)" "(41.98135196, -87.70867192)" "(41.87285714, -87.75335
467)" ...
##  $ College Enrollment.Rate %                   : num  82.4 51.9 90.7 60.4 65 78.5
73.2 50 45.1 48.1 ...
```

Plot Variables

```
#enrollment is column
plot(data$`Average Student Attendance`,data$`College Enrollment Rate%`)
```

```
## Warning: Unknown or uninitialised column: 'College Enrollment Rate%'.
```
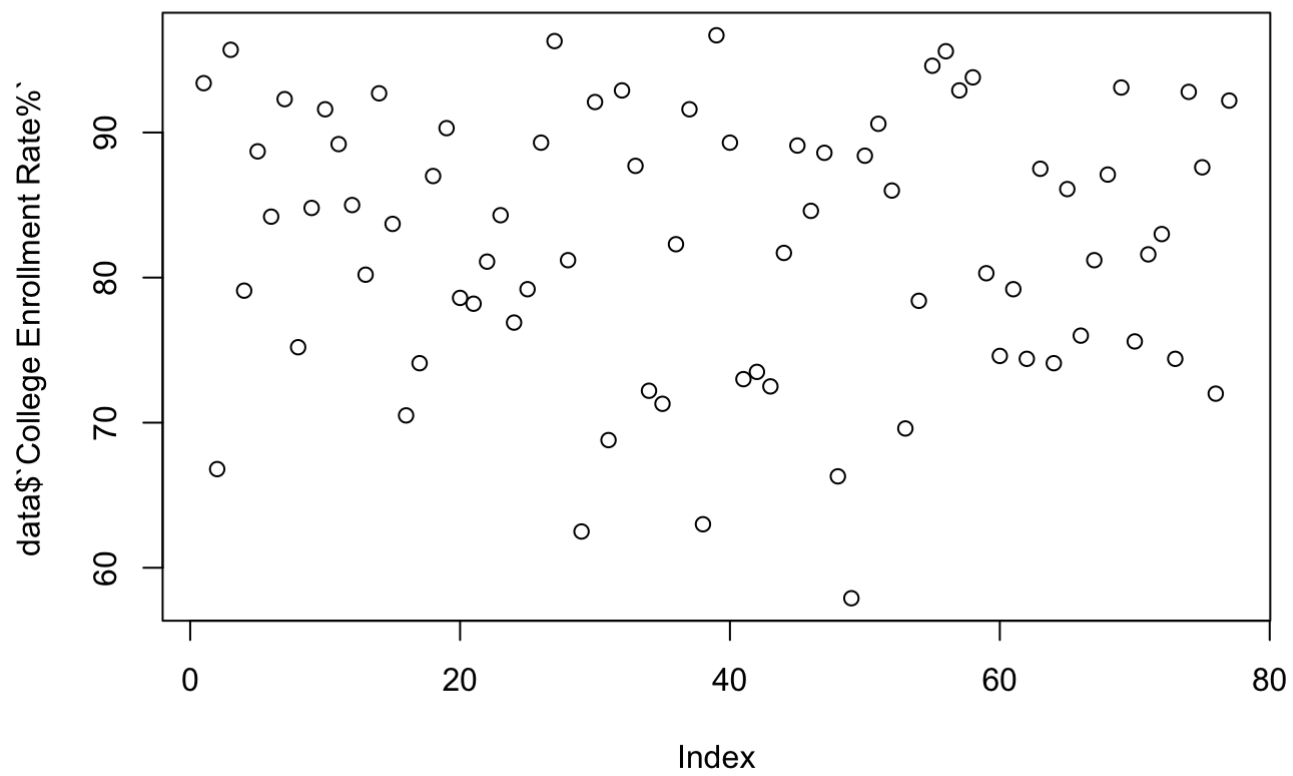
```
plot(data$`9th Grade EXPLORE (2010)`,data$`College Enrollment Rate %`)
```
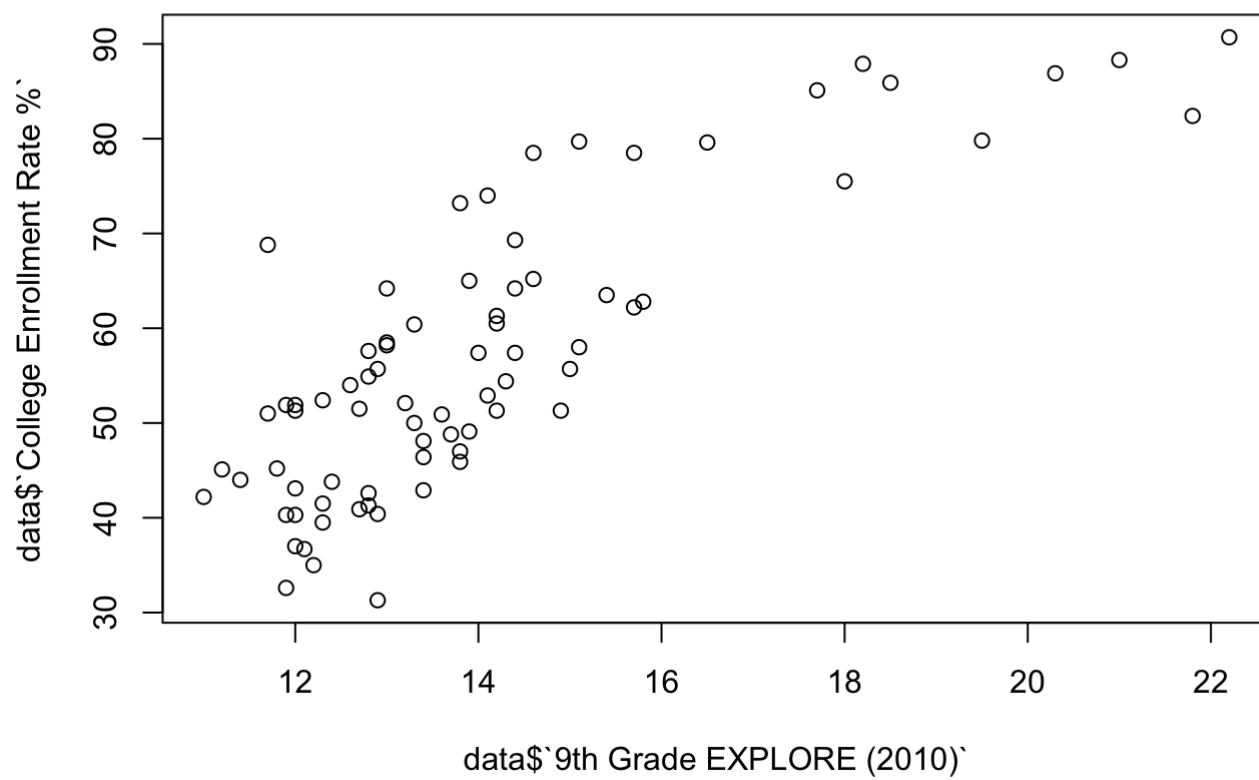
```
plot(data$`Safety Score`,data$`College Enrollment Rate %`)
```
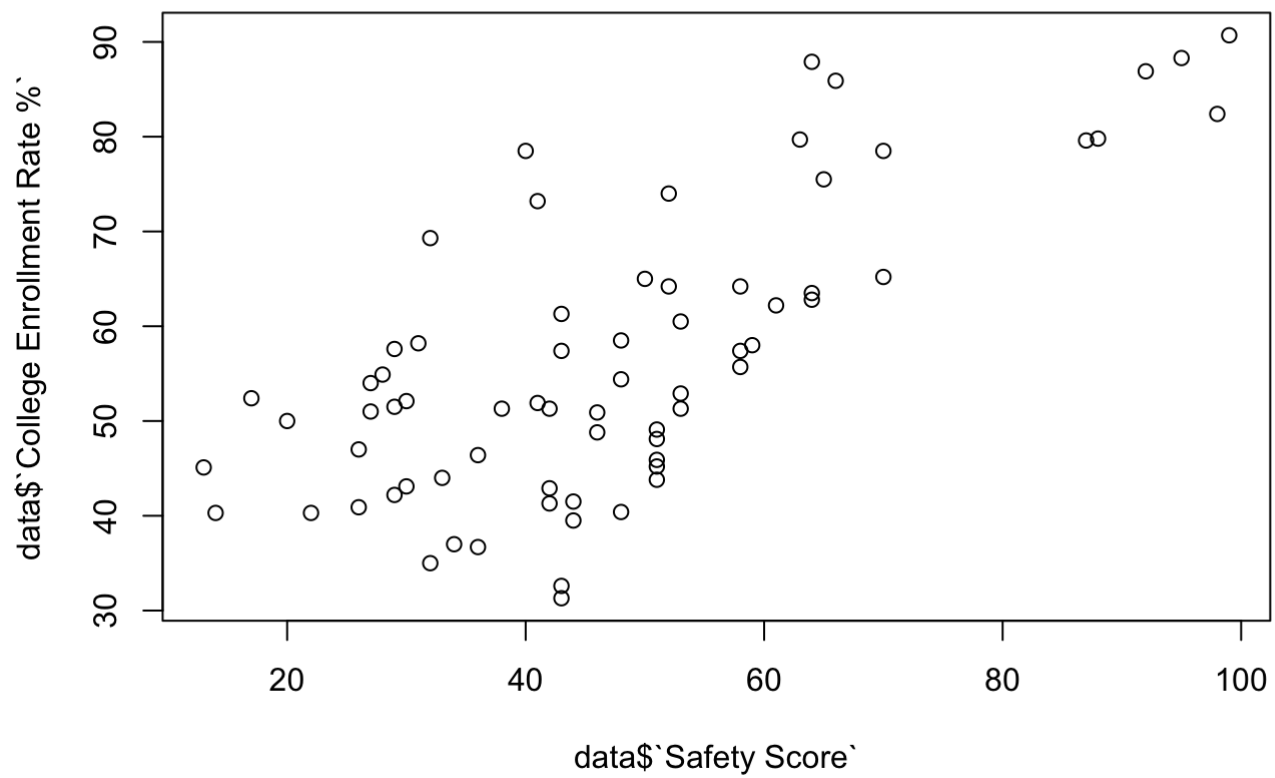
```
plot(data$`Environment Score`,data$`College Enrollment Rate %`)
```
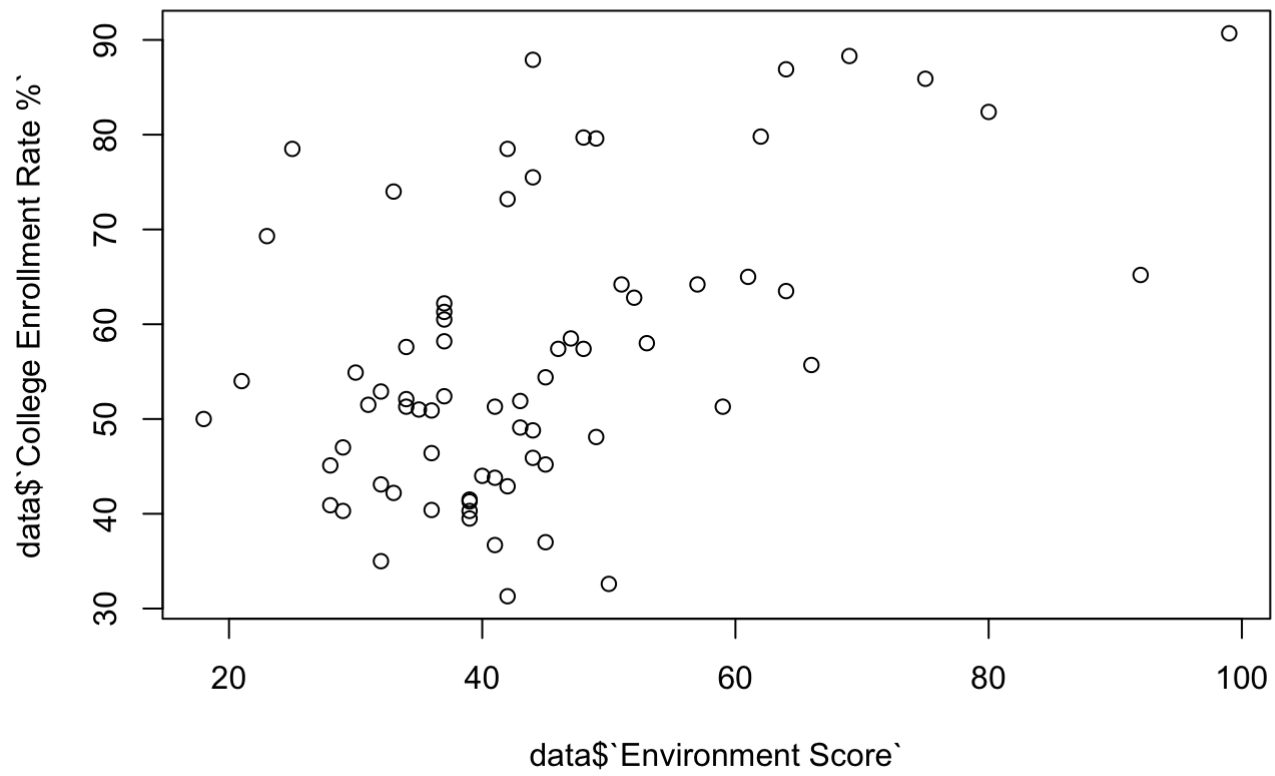
```
plot(data$`11th Grade Average ACT (2011)`,data$`College Enrollment Rate %`)
```
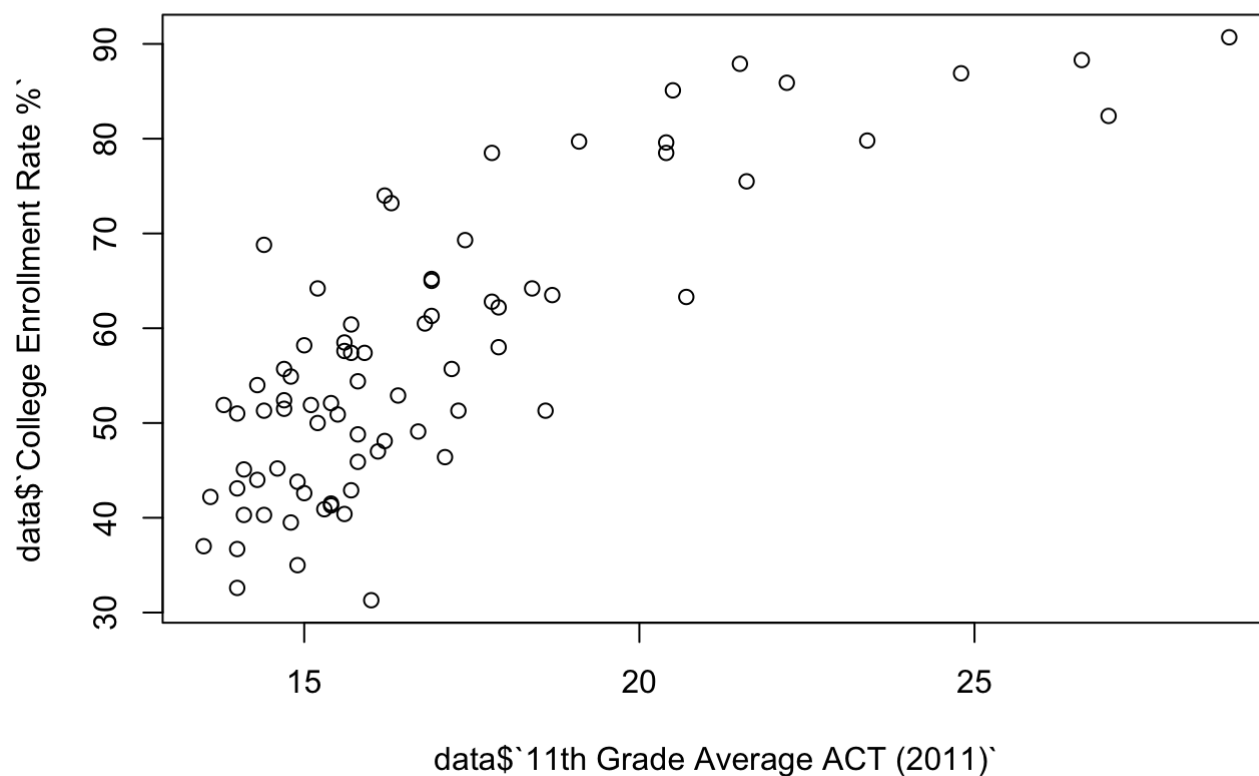
Identify Correlation

```
cor(data$`Average Student Attendance`,data$`College Enrollment Rate %`, use = "pairwise.
complete.obs")
```

```
## [1] 0.5702698
```

```
cor(data$`9th Grade EXPLORE (2010)`,data$`College Enrollment Rate %`, use = "pairwise.co
mplete.obs")
```

```
## [1] 0.8171213
```

```
cor(data$`Safety Score`,data$`College Enrollment Rate %`,use = "pairwise.complete.obs")
```

```
## [1] 0.7215355
```

```
cor(data$`Environment Score`,data$`College Enrollment Rate %`, use = "pairwise.complete.
obs")
```

```
## [1] 0.5070781
```

```
cor(data$`11th Grade Average ACT (2011)`,data$`College Enrollment Rate %`, use = "pairwi
se.complete.obs")
```

```
## [1] 0.8028903
```

Write Linear Models

```
(attendance.model<-lm(`Average Student Attendance`~`College Enrollment Rate %`, na.actio
n=na.omit, data=data))
```

```
##
## Call:
## lm(formula = `Average Student Attendance` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Coefficients:
##                (Intercept)  `College Enrollment Rate %`
##                    62.7255                       0.3464
```

```
(explore.model<-lm(`9th Grade EXPLORE (2010)`~`College Enrollment Rate %`, na.action=na.
omit, data=data))
```

```
##
## Call:
## lm(formula = `9th Grade EXPLORE (2010)` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Coefficients:
##                (Intercept)  `College Enrollment Rate %`
##                     6.5396                       0.1324
```

```
(safety.model<-lm(`Safety Score`~`College Enrollment Rate %`, na.action=na.omit, data=da
ta))
```

```
##
## Call:
## lm(formula = `Safety Score` ~ `College Enrollment Rate %`, data = data,
##     na.action = na.omit)
##
## Coefficients:
##                (Intercept)  `College Enrollment Rate %`
##                    -5.1079                       0.9312
```

```
(environment.model<-lm(`Environment Score`~`College Enrollment Rate %`, na.action=na.omi
t, data=data))
```

```
##
## Call:
## lm(formula = `Environment Score` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Coefficients:
##               (Intercept)  `College Enrollment Rate %`
##                   15.3714                       0.5058
```

```
(act.model<-lm(`11th Grade Average ACT (2011)`~`College Enrollment Rate %`, na.action=n
a.omit, data=data))
```

```
##
## Call:
## lm(formula = `11th Grade Average ACT (2011)` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Coefficients:
##               (Intercept)  `College Enrollment Rate %`
##                    7.1217                       0.1711
```

View Summaries of Linear Models

```
summary(attendance.model)
```

```
##
## Call:
## lm(formula = `Average Student Attendance` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8923  -4.4962   0.8267   4.8279  12.9443
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  62.72547    3.39409  18.481  < 2e-16 ***
## `College Enrollment Rate %`   0.34641    0.05762   6.012 6.16e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.54 on 75 degrees of freedom
## Multiple R-squared:  0.3252, Adjusted R-squared:  0.3162
## F-statistic: 36.15 on 1 and 75 DF,  p-value: 6.162e-08
```

```
summary(explore.model) #explore model has the highest adj r^2
```

```
##
## Call:
## lm(formula = `9th Grade EXPLORE (2010)` ~ `College Enrollment Rate %`,
##      data = data, na.action = na.omit)
##
## Residuals:
##     Min       1Q  Median       3Q      Max
## -3.9455  -1.1374   0.0949   0.8640   4.3545
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  6.53961    0.63874   10.24 7.94e-16 ***
## `College Enrollment Rate %`  0.13235    0.01085   12.19  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.419 on 74 degrees of freedom
##    (1 observation deleted due to missingness)
## Multiple R-squared:  0.6677, Adjusted R-squared:  0.6632
## F-statistic: 148.7 on 1 and 74 DF,  p-value: < 2.2e-16
```

```
summary(safety.model)
```

```
##
## Call:
## lm(formula = `Safety Score` ~ `College Enrollment Rate %`, data = data,
##      na.action = na.omit)
##
## Residuals:
##     Min       1Q  Median       3Q      Max
## -27.988 -10.056    1.893   10.446   26.381
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -5.1079     6.3439  -0.805    0.424
## `College Enrollment Rate %`   0.9312     0.1084   8.593 1.83e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.66 on 68 degrees of freedom
##    (7 observations deleted due to missingness)
## Multiple R-squared:  0.5206, Adjusted R-squared:  0.5136
## F-statistic: 73.85 on 1 and 68 DF,  p-value: 1.835e-12
```

```
summary(environment.model)
```

```
##
## Call:
## lm(formula = `Environment Score` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.075  -8.743   1.488   6.054  43.652
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 15.3714     6.1038   2.518   0.0141 *
## `College Enrollment Rate %`  0.5058     0.1043   4.851 7.48e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.15 on 68 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.2571, Adjusted R-squared:  0.2462
## F-statistic: 23.54 on 1 and 68 DF,  p-value: 7.476e-06
```

```
summary(act.model)
```

```
##
## Call:
## lm(formula = `11th Grade Average ACT (2011)` ~ `College Enrollment Rate %`,
##     data = data, na.action = na.omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4940 -1.3438 -0.1538  0.9361  6.1587
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 7.12174    0.86414   8.241 4.14e-12 ***
## `College Enrollment Rate %` 0.17111    0.01467  11.664  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.92 on 75 degrees of freedom
## Multiple R-squared:  0.6446, Adjusted R-squared:  0.6399
## F-statistic:   136 on 1 and 75 DF,  p-value: < 2.2e-16
```

Creating New Data Frame with these 5 Factors

```
data.numeric <- cbind(data$`College Enrollment Rate %`, data$`Average Student Attendance
`, data$`9th Grade EXPLORE (2010)`, data$`Safety Score`, data$`Environment Score`, data
$`11th Grade Average ACT (2011)`)

data.numeric <- as.data.frame(data.numeric)

colnames(data.numeric) <- c("College Enrollment Rate %", "Average Student Attendance",
"9th Grade EXPLORE (2010)", "Safety Score", "Environment Score", "11th Grade Average ACT
(2011)")

rownames(data.numeric) <- data$`Name of School`

data.numeric <- na.omit(data.numeric)

data.numeric[,1] <- as.numeric(data.numeric[,1])
data.numeric[,2] <- as.numeric(data.numeric[,2])
data.numeric[,3] <- as.numeric(data.numeric[,3])
data.numeric[,4] <- as.numeric(data.numeric[,4])
data.numeric[,5] <- as.numeric(data.numeric[,5])
data.numeric[,6] <- as.numeric(data.numeric[,6])

str(data.numeric)
```

```
## 'data.frame':    70 obs. of  6 variables:
##  $ College Enrollment Rate %    : num  82.4 51.9 90.7 65 78.5 73.2 50 45.1 48.1 58.5
...
##  $ Average Student Attendance   : num  93.4 66.8 95.7 88.7 84.2 92.3 75.2 84.8 91.6 8
9.2 ...
##  $ 9th Grade EXPLORE (2010)     : num  21.8 11.9 22.2 13.9 14.6 13.8 13.3 11.2 13.4 1
3 ...
##  $ Safety Score                 : num  98 41 99 50 40 41 20 13 51 48 ...
##  $ Environment Score            : num  80 43 99 61 25 42 18 28 49 47 ...
##  $ 11th Grade Average ACT (2011): num  27 13.8 28.8 16.9 17.8 16.3 15.2 14.1 16.2 15.
6 ...
##  - attr(*, "na.action")= 'omit' Named int  4 38 39 48 57 67 76
##   ..- attr(*, "names")= chr  "Michele Clark Academic Prep Magnet High School" "Wendel
l Phillips Academy High School" "DeVry University Advantage Academy High School" "Orr Ac
ademy High School" ...
```

```
head(data.numeric)
```

```
##                                                College Enrollment Rate %
## Walter Payton College Preparatory High School                      82.4
## Manley Career Academy High School                                  51.9
## Northside College Preparatory High School                          90.7
## Uplift Community High School                                        65.0
## Morgan Park High School                                             78.5
## Bronzeville Scholastic Academy High School                         73.2
##                                                Average Student Attendance
## Walter Payton College Preparatory High School                      93.4
## Manley Career Academy High School                                  66.8
## Northside College Preparatory High School                          95.7
## Uplift Community High School                                        88.7
## Morgan Park High School                                             84.2
## Bronzeville Scholastic Academy High School                         92.3
##                                                9th Grade EXPLORE (2010)
## Walter Payton College Preparatory High School                      21.8
## Manley Career Academy High School                                  11.9
## Northside College Preparatory High School                          22.2
## Uplift Community High School                                        13.9
## Morgan Park High School                                             14.6
## Bronzeville Scholastic Academy High School                         13.8
##                                                Safety Score
## Walter Payton College Preparatory High School        98
## Manley Career Academy High School                    41
## Northside College Preparatory High School            99
## Uplift Community High School                         50
## Morgan Park High School                              40
## Bronzeville Scholastic Academy High School          41
##                                                Environment Score
## Walter Payton College Preparatory High School        80
## Manley Career Academy High School                    43
## Northside College Preparatory High School            99
## Uplift Community High School                         61
## Morgan Park High School                              25
## Bronzeville Scholastic Academy High School          42
##                                                11th Grade Average ACT (2011)
## Walter Payton College Preparatory High School                      27.0
## Manley Career Academy High School                                  13.8
## Northside College Preparatory High School                          28.8
## Uplift Community High School                                        16.9
## Morgan Park High School                                             17.8
## Bronzeville Scholastic Academy High School                         16.3
```

Correlation Matrix

```
cor(data.numeric)
```

```
##                                 College Enrollment Rate %
## College Enrollment Rate %                      1.0000000
## Average Student Attendance                     0.5843717
## 9th Grade EXPLORE (2010)                       0.8350476
## Safety Score                                   0.7215355
## Environment Score                              0.5070781
## 11th Grade Average ACT (2011)                  0.8211524
##                                 Average Student Attendance
## College Enrollment Rate %                       0.5843717
## Average Student Attendance                      1.0000000
## 9th Grade EXPLORE (2010)                        0.6602512
## Safety Score                                    0.6335752
## Environment Score                               0.5363853
## 11th Grade Average ACT (2011)                   0.6549657
##                                 9th Grade EXPLORE (2010) Safety Score
## College Enrollment Rate %                        0.8350476    0.7215355
## Average Student Attendance                       0.6602512    0.6335752
## 9th Grade EXPLORE (2010)                         1.0000000    0.8683360
## Safety Score                                     0.8683360    1.0000000
## Environment Score                                0.6749397    0.7851015
## 11th Grade Average ACT (2011)                    0.9805297    0.8620837
##                                 Environment Score
## College Enrollment Rate %              0.5070781
## Average Student Attendance             0.5363853
## 9th Grade EXPLORE (2010)               0.6749397
## Safety Score                           0.7851015
## Environment Score                      1.0000000
## 11th Grade Average ACT (2011)          0.6800312
##                                 11th Grade Average ACT (2011)
## College Enrollment Rate %                          0.8211524
## Average Student Attendance                         0.6549657
## 9th Grade EXPLORE (2010)                           0.9805297
## Safety Score                                       0.8620837
## Environment Score                                  0.6800312
## 11th Grade Average ACT (2011)                      1.0000000
```

Multiple Regression

```
View(data.numeric)
summary(lm(`College Enrollment Rate %` ~., data=data.numeric))
```

```
##
## Call:
## lm(formula = `College Enrollment Rate %` ~ ., data = data.numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.837  -5.510  -1.950   6.667  17.290
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -20.49384   12.04670  -1.701   0.0938 .
## `Average Student Attendance`     0.12608    0.15951   0.790   0.4322
## `9th Grade EXPLORE (2010)`       4.43451    2.18781   2.027   0.0468 *
## `Safety Score`                   0.06603    0.12636   0.523   0.6031
## `Environment Score`             -0.15059    0.11013  -1.367   0.1763
## `11th Grade Average ACT (2011)`  0.44096    1.61912   0.272   0.7862
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.515 on 64 degrees of freedom
## Multiple R-squared:  0.7081, Adjusted R-squared:  0.6853
## F-statistic: 31.05 on 5 and 64 DF,  p-value: 6.754e-16
```

Adusted R-squared - .6503

Next Attempt- grab all numeric values from the data set

All Numeric Columns Method

```
all.numeric <- cbind(data[,c(18,22,24,33,34,35,36,64,65,69,80)])
all.numeric<- na.omit(all.numeric)

head(all.numeric)
```

```
##    Safety Score Environment Score Instruction Score
## 1           98                80                77
## 2           41                43                31
## 3           99                99                88
## 5           50                61                62
## 6           40                25                38
## 7           41                42                43
##    Average Student Attendance Rate of Misconducts (per 100 students)
## 1                      93.4                                      0.7
## 2                      66.8                                     19.7
## 3                      95.7                                      2.8
## 5                      88.7                                      8.1
## 6                      84.2                                     11.1
## 7                      92.3                                     16.2
##    Average Teacher Attendance
## 1                       96.1
## 2                       95.4
## 3                       96.8
## 5                       95.9
## 6                       94.6
## 7                       95.6
##    Individualized Education Program Compliance Rate College Eligibility %
## 1                                             100.0                 96.4
## 2                                              98.4                  6.7
## 3                                              98.4                   98
## 5                                              98.3   19.600000000000001
## 6                                              99.5                 33.6
## 7                                              97.6                 14.7
##    Graduation Rate % Freshman on Track Rate % College Enrollment.Rate %
## 1               96.9                    90.7                       82.4
## 2                 49                    59.3                       51.9
## 3               97.6                    95.9                       90.7
## 5               62.4                    91.4                       65.0
## 6                 80                    73.3                       78.5
## 7               85.3       71.099999999999994                     73.2
```

Converting Columns to Numeric

```
all.numeric[,8] <- as.numeric(all.numeric[,8])
all.numeric[,9] <- as.numeric(all.numeric[,9])
all.numeric[,10] <- as.numeric(all.numeric[,10])
```

```
## Warning: NAs introduced by coercion
```

Correlation Matrix

```
pairs(all.numeric)
```
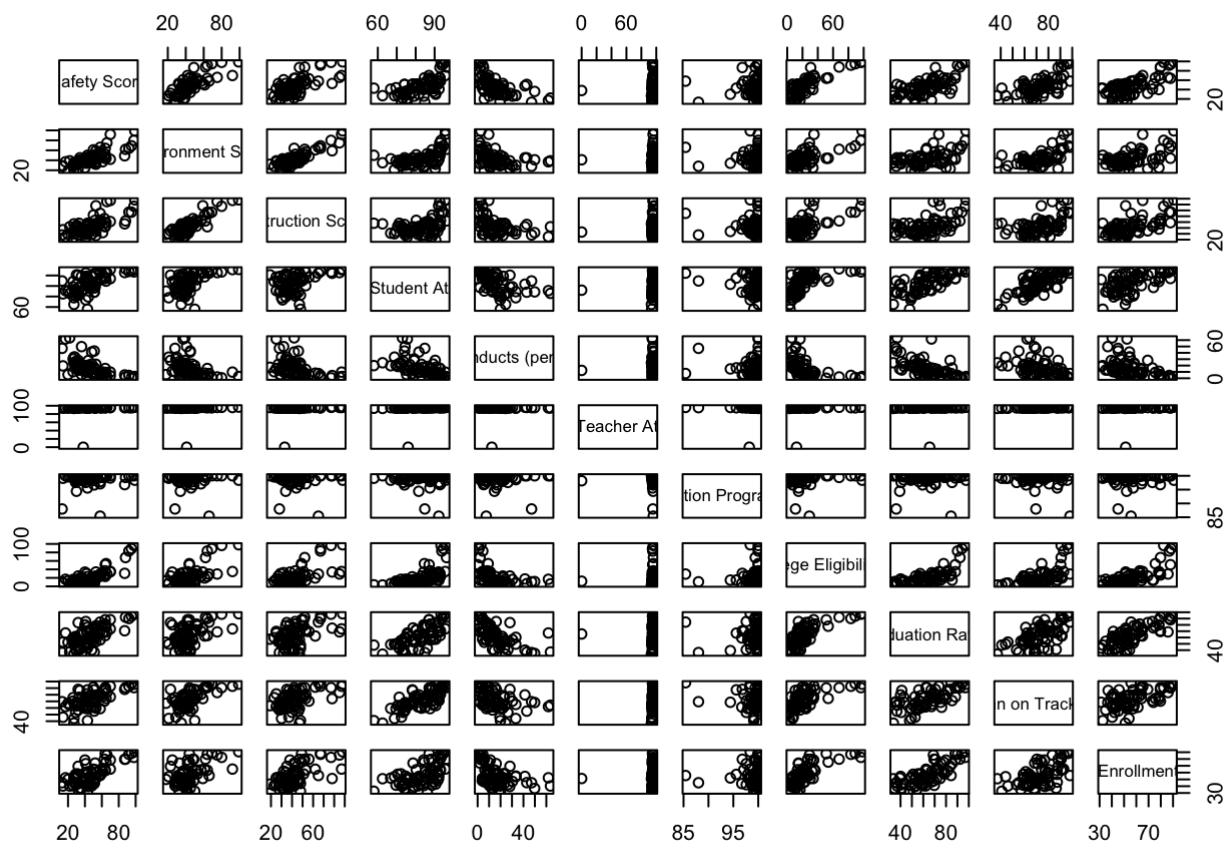
Running Multiple Regression with all variables

```
model.all.variables <- lm(`College Enrollment.Rate %`~., data=all.numeric)
summary(model.all.variables)
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ ., data = all.numeric)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -19.0296  -4.4795  -0.1388   4.8983  17.3797
##
## Coefficients:
##                                                     Estimate Std. Error
## (Intercept)                                        252.146658 150.205393
## `Safety Score`                                       0.364282   0.115596
## `Environment Score`                                 -0.707361   0.166076
## `Instruction Score`                                  0.648437   0.138998
## `Average Student Attendance`                         0.354205   0.216280
## `Rate of Misconducts (per 100 students)`             0.178420   0.095412
## `Average Teacher Attendance`                        -2.813760   1.582791
## `Individualized Education Program Compliance Rate`   0.004121   0.406547
## `College Eligibility %`                              0.129061   0.089650
## `Graduation Rate %`                                  0.276224   0.107720
## `Freshman on Track Rate %`                           0.070907   0.115689
##                                                     t value Pr(>|t|)
## (Intercept)                                           1.679  0.09860 .
## `Safety Score`                                        3.151  0.00257 **
## `Environment Score`                                  -4.259 7.61e-05 ***
## `Instruction Score`                                   4.665 1.86e-05 ***
## `Average Student Attendance`                          1.638  0.10690
## `Rate of Misconducts (per 100 students)`              1.870  0.06653 .
## `Average Teacher Attendance`                         -1.778  0.08069 .
## `Individualized Education Program Compliance Rate`    0.010  0.99195
## `College Eligibility %`                               1.440  0.15535
## `Graduation Rate %`                                   2.564  0.01295 *
## `Freshman on Track Rate %`                            0.613  0.54233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.681 on 58 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7844, Adjusted R-squared:  0.7472
## F-statistic:  21.1 on 10 and 58 DF,  p-value: 7.627e-16
```

Drop1 Technique

Drop1 first attempt

```
drop1(model.all.variables)
```

```
## Single term deletions
##
## Model:
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `Individualized Education Program Compliance Rate`
+
##     `College Eligibility %` + `Graduation Rate %` + `Freshman on Track Rate %`
##                                                     Df Sum of Sq    RSS
## <none>                                                           3421.7
## `Safety Score`                                       1    585.87 4007.5
## `Environment Score`                                  1   1070.23 4491.9
## `Instruction Score`                                  1   1283.88 4705.5
## `Average Student Attendance`                         1    158.23 3579.9
## `Rate of Misconducts (per 100 students)`             1    206.30 3628.0
## `Average Teacher Attendance`                         1    186.44 3608.1
## `Individualized Education Program Compliance Rate`   1      0.01 3421.7
## `College Eligibility %`                              1    122.26 3543.9
## `Graduation Rate %`                                  1    387.91 3809.6
## `Freshman on Track Rate %`                           1     22.16 3443.8
##                                                       AIC
## <none>                                              291.36
## `Safety Score`                                      300.27
## `Environment Score`                                 308.14
## `Instruction Score`                                 311.35
## `Average Student Attendance`                        292.48
## `Rate of Misconducts (per 100 students)`            293.40
## `Average Teacher Attendance`                        293.02
## `Individualized Education Program Compliance Rate`  289.36
## `College Eligibility %`                             291.78
## `Graduation Rate %`                                 296.77
## `Freshman on Track Rate %`                          289.81
```

Dropping Freshman on Track Rate %

```
m2 <- lm(`College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
    `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 100 s
tudents)` +
    `Average Teacher Attendance` + `Individualized Education Program Compliance Rate` +
    `College Eligibility %` + `Graduation Rate %`, data=all.numeric)

summary(m2)
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `Individualized Education Program Compliance Rate`
+
##     `College Eligibility %` + `Graduation Rate %`, data = all.numeric)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -23.8838  -4.9579  -0.5031   4.4371  15.8005
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                         5.66102   44.56527
## `Safety Score`                                      0.34834    0.11644
## `Environment Score`                                -0.71737    0.16388
## `Instruction Score`                                 0.65480    0.13993
## `Average Student Attendance`                        0.23261    0.16959
## `Rate of Misconducts (per 100 students)`            0.17783    0.09619
## `Average Teacher Attendance`                       -0.07119    0.08585
## `Individualized Education Program Compliance Rate` -0.01655    0.40790
## `College Eligibility %`                             0.14520    0.09027
## `Graduation Rate %`                                 0.31806    0.10631
##                                                    t value Pr(>|t|)
## (Intercept)                                          0.127  0.89934
## `Safety Score`                                       2.992  0.00402 **
## `Environment Score`                                 -4.377 4.88e-05 ***
## `Instruction Score`                                  4.679 1.68e-05 ***
## `Average Student Attendance`                         1.372  0.17530
## `Rate of Misconducts (per 100 students)`             1.849  0.06943 .
## `Average Teacher Attendance`                        -0.829  0.41025
## `Individualized Education Program Compliance Rate`  -0.041  0.96776
## `College Eligibility %`                              1.608  0.11298
## `Graduation Rate %`                                  2.992  0.00402 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.773 on 60 degrees of freedom
## Multiple R-squared:  0.772,  Adjusted R-squared:  0.7378
## F-statistic: 22.57 on 9 and 60 DF,  p-value: 3.756e-16
```

Drop1 Second Attempt

```
drop1(m2)
```

```
## Single term deletions
##
## Model:
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `Individualized Education Program Compliance Rate`
+
##     `College Eligibility %` + `Graduation Rate %`
##                                                    Df Sum of Sq   RSS
## <none>                                                          3625.3
## `Safety Score`                                      1    540.73 4166.0
## `Environment Score`                                 1   1157.82 4783.1
## `Instruction Score`                                 1   1323.07 4948.4
## `Average Student Attendance`                        1    113.67 3739.0
## `Rate of Misconducts (per 100 students)`            1    206.50 3831.8
## `Average Teacher Attendance`                        1     41.55 3666.9
## `Individualized Education Program Compliance Rate`  1      0.10 3625.4
## `College Eligibility %`                             1    156.33 3781.6
## `Graduation Rate %`                                 1    540.88 4166.2
##                                                       AIC
## <none>                                              296.30
## `Safety Score`                                      304.04
## `Environment Score`                                 313.70
## `Instruction Score`                                 316.08
## `Average Student Attendance`                        296.46
## `Rate of Misconducts (per 100 students)`            298.18
## `Average Teacher Attendance`                        295.10
## `Individualized Education Program Compliance Rate`  294.31
## `College Eligibility %`                             297.26
## `Graduation Rate %`                                 304.04
```

Removing Individualized Education Program Compliance Rate

```
m3 <- lm(`College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
    `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 100 s
tudents)` +
    `Average Teacher Attendance` + `College Eligibility %` + `Graduation Rate %`, data=a
ll.numeric)


summary(m3)
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `College Eligibility %` +
##     `Graduation Rate %`, data = all.numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.9218  -4.9684  -0.5033   4.4503  15.7608
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                 3.94162   13.71154   0.287
## `Safety Score`                              0.34769    0.11438   3.040
## `Environment Score`                        -0.71703    0.16232  -4.417
## `Instruction Score`                         0.65514    0.13853   4.729
## `Average Student Attendance`                0.23395    0.16494   1.418
## `Rate of Misconducts (per 100 students)`    0.17799    0.09532   1.867
## `Average Teacher Attendance`               -0.07138    0.08502  -0.840
## `College Eligibility %`                     0.14502    0.08942   1.622
## `Graduation Rate %`                         0.31792    0.10538   3.017
##                                            Pr(>|t|)
## (Intercept)                                 0.77473
## `Safety Score`                              0.00349 **
## `Environment Score`                        4.16e-05 ***
## `Instruction Score`                        1.38e-05 ***
## `Average Student Attendance`                0.16116
## `Rate of Misconducts (per 100 students)`    0.06665 .
## `Average Teacher Attendance`                0.40443
## `College Eligibility %`                     0.11001
## `Graduation Rate %`                         0.00372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.709 on 61 degrees of freedom
## Multiple R-squared:  0.772,  Adjusted R-squared:  0.7421
## F-statistic: 25.81 on 8 and 61 DF,  p-value: < 2.2e-16
```

Drop1 third attempt

```
drop1(m3)
```

```
## Single term deletions
##
## Model:
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `College Eligibility %` +
##     `Graduation Rate %`
##                                            Df Sum of Sq    RSS    AIC
## <none>                                                   3625.4 294.31
## `Safety Score`                             1    549.16 4174.6 302.18
## `Environment Score`                        1   1159.68 4785.1 311.73
## `Instruction Score`                        1   1329.25 4954.6 314.17
## `Average Student Attendance`               1    119.57 3745.0 294.58
## `Rate of Misconducts (per 100 students)`   1    207.25 3832.6 296.20
## `Average Teacher Attendance`               1     41.89 3667.3 293.11
## `College Eligibility %`                    1    156.32 3781.7 295.26
## `Graduation Rate %`                        1    540.96 4166.4 302.04
```

Removing Average Teacher Attendance

```
m4 <- lm(`College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
    `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 100 s
tudents)` +
    + `College Eligibility %` +
    `Graduation Rate %`, data=all.numeric)

summary(m4)
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     +`College Eligibility %` + `Graduation Rate %`, data = all.numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.7116  -4.9772  -0.3718   4.3848  16.0132
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                  -0.75611   12.48804  -0.061
## `Safety Score`                                0.33732    0.11344   2.974
## `Environment Score`                          -0.68988    0.15869  -4.347
## `Instruction Score`                           0.63253    0.13556   4.666
## `Average Student Attendance`                  0.20407    0.16067   1.270
## `Rate of Misconducts (per 100 students)`      0.16923    0.09452   1.790
## `College Eligibility %`                       0.14327    0.08918   1.606
## `Graduation Rate %`                           0.33256    0.10368   3.208
##                                              Pr(>|t|)
## (Intercept)                                   0.95192
## `Safety Score`                                0.00419 **
## `Environment Score`                          5.22e-05 ***
## `Instruction Score`                          1.69e-05 ***
## `Average Student Attendance`                  0.20879
## `Rate of Misconducts (per 100 students)`      0.07826 .
## `College Eligibility %`                       0.11326
## `Graduation Rate %`                           0.00212 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.691 on 62 degrees of freedom
## Multiple R-squared:  0.7693, Adjusted R-squared:  0.7433
## F-statistic: 29.54 on 7 and 62 DF,  p-value: < 2.2e-16
```

Drop 1 forth attempt

```
drop1(m4)
```

```
## Single term deletions
##
## Model:
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     +`College Eligibility %` + `Graduation Rate %`
##                                      Df Sum of Sq    RSS    AIC
## <none>                                            3667.3 293.11
## `Safety Score`                        1    523.00 4190.3 300.44
## `Environment Score`                   1   1117.89 4785.2 309.74
## `Instruction Score`                   1   1287.75 4955.0 312.18
## `Average Student Attendance`          1     95.42 3762.7 292.91
## `Rate of Misconducts (per 100 students)`  1    189.62 3856.9 294.64
## `College Eligibility %`               1    152.65 3819.9 293.96
## `Graduation Rate %`                   1    608.56 4275.9 301.86
```

Per drop1 method, this model is best Variables included

```
colnames(m4$model)
```

```
## [1] "College Enrollment.Rate %"
## [2] "Safety Score"
## [3] "Environment Score"
## [4] "Instruction Score"
## [5] "Average Student Attendance"
## [6] "Rate of Misconducts (per 100 students)"
## [7] "College Eligibility %"
## [8] "Graduation Rate %"
```

Step Method

```
all.numeric <- na.omit(all.numeric)
head(all.numeric)
```

```
##    Safety Score Environment Score Instruction Score
## 1           98                80                77
## 2           41                43                31
## 3           99                99                88
## 5           50                61                62
## 6           40                25                38
## 7           41                42                43
##    Average Student Attendance Rate of Misconducts (per 100 students)
## 1                        93.4                                    0.7
## 2                        66.8                                   19.7
## 3                        95.7                                    2.8
## 5                        88.7                                    8.1
## 6                        84.2                                   11.1
## 7                        92.3                                   16.2
##    Average Teacher Attendance
## 1                        96.1
## 2                        95.4
## 3                        96.8
## 5                        95.9
## 6                        94.6
## 7                        95.6
##    Individualized Education Program Compliance Rate College Eligibility %
## 1                                             100.0                  96.4
## 2                                              98.4                   6.7
## 3                                              98.4                  98.0
## 5                                              98.3                  19.6
## 6                                              99.5                  33.6
## 7                                              97.6                  14.7
##    Graduation Rate % Freshman on Track Rate % College Enrollment.Rate %
## 1              96.9                    90.7                        82.4
## 2              49.0                    59.3                        51.9
## 3              97.6                    95.9                        90.7
## 5              62.4                    91.4                        65.0
## 6              80.0                    73.3                        78.5
## 7              85.3                    71.1                        73.2
```

```
step(model.all.variables, direction="both")
```

```
## Start:  AIC=291.36
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##      `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##      `Average Teacher Attendance` + `Individualized Education Program Compliance Rate`
+
##      `College Eligibility %` + `Graduation Rate %` + `Freshman on Track Rate %`
##
##                                                        Df Sum of Sq    RSS
## - `Individualized Education Program Compliance Rate`    1      0.01 3421.7
## - `Freshman on Track Rate %`                            1     22.16 3443.8
## <none>                                                              3421.7
## - `College Eligibility %`                               1    122.26 3543.9
## - `Average Student Attendance`                          1    158.23 3579.9
## - `Average Teacher Attendance`                          1    186.44 3608.1
## - `Rate of Misconducts (per 100 students)`              1    206.30 3628.0
## - `Graduation Rate %`                                   1    387.91 3809.6
## - `Safety Score`                                        1    585.87 4007.5
## - `Environment Score`                                   1   1070.23 4491.9
## - `Instruction Score`                                   1   1283.88 4705.5
##                                                            AIC
## - `Individualized Education Program Compliance Rate`   289.36
## - `Freshman on Track Rate %`                           289.81
## <none>                                                 291.36
## - `College Eligibility %`                              291.78
## - `Average Student Attendance`                         292.48
## - `Average Teacher Attendance`                         293.02
## - `Rate of Misconducts (per 100 students)`             293.40
## - `Graduation Rate %`                                  296.77
## - `Safety Score`                                       300.27
## - `Environment Score`                                  308.14
## - `Instruction Score`                                  311.35
##
## Step:  AIC=289.36
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##      `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##      `Average Teacher Attendance` + `College Eligibility %` +
##      `Graduation Rate %` + `Freshman on Track Rate %`
##
##                                                        Df Sum of Sq    RSS
## - `Freshman on Track Rate %`                            1     22.44 3444.1
## <none>                                                              3421.7
## - `College Eligibility %`                               1    122.72 3544.4
## - `Average Student Attendance`                          1    159.28 3580.9
## - `Average Teacher Attendance`                          1    186.53 3608.2
## + `Individualized Education Program Compliance Rate`    1      0.01 3421.7
## - `Rate of Misconducts (per 100 students)`              1    206.42 3628.1
## - `Graduation Rate %`                                   1    388.87 3810.5
## - `Safety Score`                                        1    598.47 4020.1
## - `Environment Score`                                   1   1070.83 4492.5
## - `Instruction Score`                                   1   1290.05 4711.7
##                                                            AIC
```

```
## - `Freshman on Track Rate %`                              287.81
## <none>                                                    289.36
## - `College Eligibility %`                                 289.79
## - `Average Student Attendance`                            290.50
## - `Average Teacher Attendance`                            291.02
## + `Individualized Education Program Compliance Rate`       291.36
## - `Rate of Misconducts (per 100 students)`                291.40
## - `Graduation Rate %`                                      294.79
## - `Safety Score`                                           298.48
## - `Environment Score`                                      306.15
## - `Instruction Score`                                      309.44
##
## Step:  AIC=287.81
## `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##      `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##      `Average Teacher Attendance` + `College Eligibility %` +
##      `Graduation Rate %`
##
##                                                          Df Sum of Sq     RSS
## <none>                                                                  3444.1
## - `College Eligibility %`                                 1    128.32  3572.4
## + `Freshman on Track Rate %`                              1     22.44  3421.7
## - `Average Teacher Attendance`                            1    190.21  3634.3
## + `Individualized Education Program Compliance Rate`      1      0.29  3443.8
## - `Rate of Misconducts (per 100 students)`               1    219.91  3664.0
## - `Average Student Attendance`                            1    268.61  3712.7
## - `Graduation Rate %`                                     1    421.79  3865.9
## - `Safety Score`                                          1    602.14  4046.2
## - `Environment Score`                                     1   1051.94  4496.0
## - `Instruction Score`                                     1   1271.46  4715.6
##                                                                AIC
## <none>                                                      287.81
## - `College Eligibility %`                                  288.34
## + `Freshman on Track Rate %`                               289.36
## - `Average Teacher Attendance`                             289.52
## + `Individualized Education Program Compliance Rate`       289.81
## - `Rate of Misconducts (per 100 students)`                290.08
## - `Average Student Attendance`                             290.99
## - `Graduation Rate %`                                      293.78
## - `Safety Score`                                           296.93
## - `Environment Score`                                      304.20
## - `Instruction Score`                                      307.49
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `College Eligibility %` +
##     `Graduation Rate %`, data = all.numeric)
##
## Coefficients:
##                              (Intercept)
##                                 254.0380
##                            `Safety Score`
##                                   0.3655
##                       `Environment Score`
##                                  -0.6868
##                       `Instruction Score`
##                                   0.6417
##              `Average Student Attendance`
##                                   0.4126
## `Rate of Misconducts (per 100 students)`
##                                   0.1834
##              `Average Teacher Attendance`
##                                  -2.8406
##                   `College Eligibility %`
##                                   0.1319
##                       `Graduation Rate %`
##                                   0.2851
```

Summary of output of Step method

```
stepModel <- lm(formula = `College Enrollment.Rate %` ~ `Safety Score` + `Environment Sc
ore` +
    `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 100 s
tudents)` +
    `Average Teacher Attendance` + `College Eligibility %` +
    `Graduation Rate %`, data = all.numeric)

summary(stepModel)
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ `Safety Score` + `Environment Score` +
##     `Instruction Score` + `Average Student Attendance` + `Rate of Misconducts (per 10
0 students)` +
##     `Average Teacher Attendance` + `College Eligibility %` +
##     `Graduation Rate %`, data = all.numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.5567  -4.2581   0.2134   4.4530  17.4781
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             254.03803  141.37108   1.797
## `Safety Score`                            0.36552    0.11286   3.239
## `Environment Score`                      -0.68679    0.16043  -4.281
## `Instruction Score`                       0.64172    0.13635   4.706
## `Average Student Attendance`              0.41261    0.19074   2.163
## `Rate of Misconducts (per 100 students)`  0.18345    0.09372   1.957
## `Average Teacher Attendance`             -2.84063    1.56047  -1.820
## `College Eligibility %`                   0.13186    0.08819   1.495
## `Graduation Rate %`                       0.28515    0.10519   2.711
##                                          Pr(>|t|)
## (Intercept)                               0.07738 .
## `Safety Score`                            0.00196 **
## `Environment Score`                      6.81e-05 ***
## `Instruction Score`                      1.53e-05 ***
## `Average Student Attendance`              0.03452 *
## `Rate of Misconducts (per 100 students)`  0.05497 .
## `Average Teacher Attendance`              0.07369 .
## `College Eligibility %`                   0.14011
## `Graduation Rate %`                       0.00874 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.576 on 60 degrees of freedom
## Multiple R-squared:  0.783,  Adjusted R-squared:  0.754
## F-statistic: 27.06 on 8 and 60 DF,  p-value: < 2.2e-16
```

Variables used in output of step method

```
colnames(stepModel$model)
```

```
## [1] "College Enrollment.Rate %"
## [2] "Safety Score"
## [3] "Environment Score"
## [4] "Instruction Score"
## [5] "Average Student Attendance"
## [6] "Rate of Misconducts (per 100 students)"
## [7] "Average Teacher Attendance"
## [8] "College Eligibility %"
## [9] "Graduation Rate %"
```

Calculate Relative Importance of Attributes

```
suppressMessages(library(relaimpo))

metrics.attributes <- calc.relimp(stepModel, type = c("lmg", "first", "last","betasq",
"pratt"))
```

```
## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in v
ector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.
```

```
metrics.attributes$lmg.rank
```

```
##                         Safety Score
##                                    3
##                    Environment Score
##                                    6
##                    Instruction Score
##                                    4
##           Average Student Attendance
##                                    5
## Rate of Misconducts (per 100 students)
##                                    7
##           Average Teacher Attendance
##                                    8
##                 College Eligibility %
##                                    2
##                    Graduation Rate %
##                                    1
```
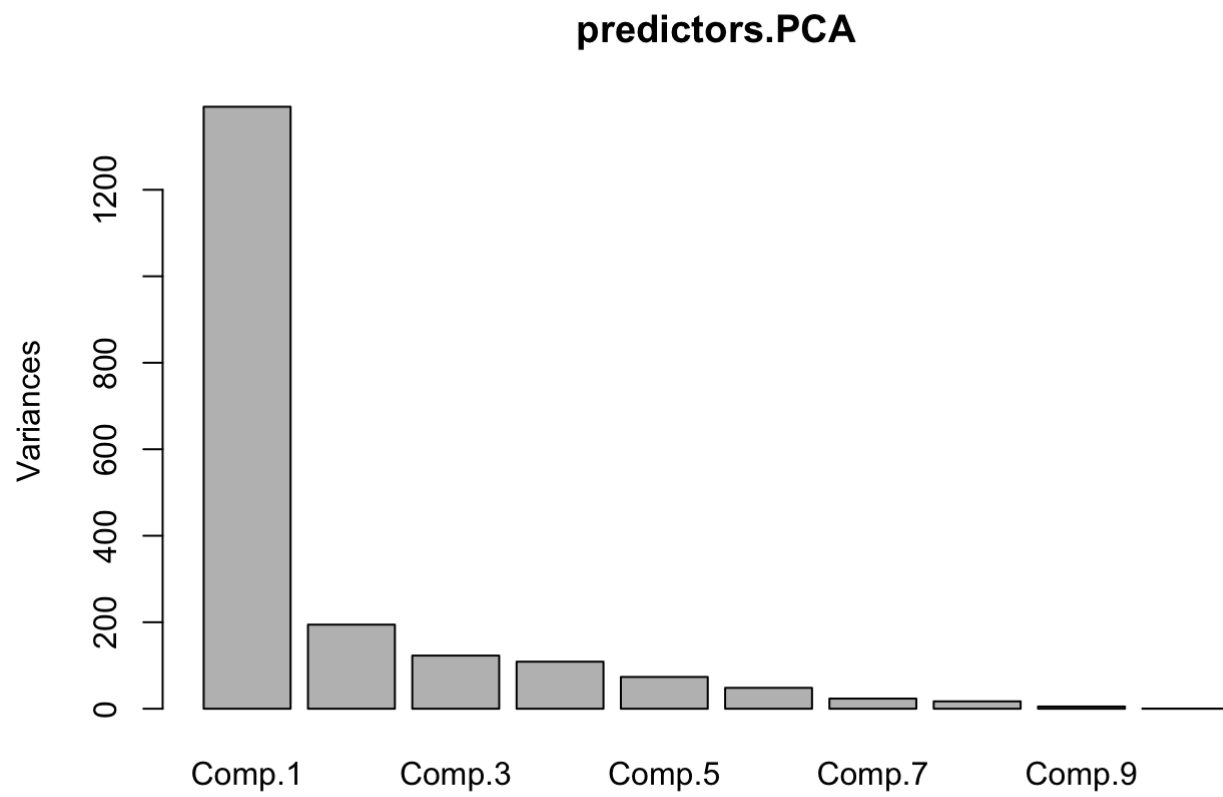
Manual Calculation of R squared

```
sum(metrics.attributes$lmg)
```

```
## [1] 0.7829778
```

PCA Analysis

```
predictors.PCA <- princomp(all.numeric[,-11])
plot(predictors.PCA)
```

predictors.PCA


```
predictors.PCA$loadings
```

```
##
## Loadings:
##                                               Comp.1 Comp.2 Comp.3
## Safety Score                                   0.490
## Environment Score                              0.336  0.539  0.137
## Instruction Score                              0.313  0.520  0.238
## Average Student Attendance                     0.171 -0.155
## Rate of Misconducts (per 100 students)        -0.236  0.377 -0.732
## Average Teacher Attendance
## Individualized Education Program Compliance Rate
## College Eligibility %                          0.508 -0.124 -0.616
## Graduation Rate %                              0.360 -0.494
## Freshman on Track Rate %                       0.275
##                                               Comp.4 Comp.5 Comp.6
## Safety Score                                   0.229  0.602  0.503
## Environment Score                             -0.135
## Instruction Score                                   -0.506 -0.125
## Average Student Attendance                    -0.330
## Rate of Misconducts (per 100 students)        -0.229 -0.102  0.433
## Average Teacher Attendance
## Individualized Education Program Compliance Rate
## College Eligibility %                          0.347        -0.468
## Graduation Rate %                                   -0.569  0.511
## Freshman on Track Rate %                      -0.796  0.201 -0.242
##                                               Comp.7 Comp.8 Comp.9
## Safety Score                                   0.177  0.223
## Environment Score                             -0.399 -0.629
## Instruction Score                              0.252  0.488
## Average Student Attendance                    -0.750  0.509
## Rate of Misconducts (per 100 students)               0.130
## Average Teacher Attendance
## Individualized Education Program Compliance Rate            -0.996
## College Eligibility %
## Graduation Rate %                                    -0.175
## Freshman on Track Rate %                       0.418
##                                               Comp.10
## Safety Score
## Environment Score
## Instruction Score
## Average Student Attendance
## Rate of Misconducts (per 100 students)
## Average Teacher Attendance                    -0.998
## Individualized Education Program Compliance Rate
## College Eligibility %
## Graduation Rate %
## Freshman on Track Rate %
##
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
## SS loadings     1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0
## Proportion Var  0.1    0.1    0.1    0.1    0.1    0.1    0.1    0.1
## Cumulative Var  0.1    0.2    0.3    0.4    0.5    0.6    0.7    0.8
##               Comp.9 Comp.10
## SS loadings     1.0    1.0
```

```
## Proportion Var     0.1      0.1
## Cumulative Var     0.9      1.0
```

Cumulative Sum of R Squared

```
cumsum(predictors.PCA$sdev^2/sum(predictors.PCA$sdev^2))
```

```
##     Comp.1     Comp.2     Comp.3     Comp.4     Comp.5     Comp.6     Comp.7
## 0.7009629 0.7988772 0.8607500 0.9155171 0.9525200 0.9768313 0.9886399
##     Comp.8     Comp.9    Comp.10
## 0.9972424 0.9998290 1.0000000
```

Create "Rotated" Data frame

```
schoolPCAFactors<- predictors.PCA$scores
schoolRotated<-as.data.frame(cbind("College Enrollment.Rate %"=all.numeric$`College Enro
llment.Rate %`,schoolPCAFactors))
```

Fit Linear Model with PCA factors

```
linModPCA<-lm(`College Enrollment.Rate %`~.,data=schoolRotated)
summary(linModPCA)
```

```
##
## Call:
## lm(formula = `College Enrollment.Rate %` ~ ., data = schoolRotated)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -19.0296  -4.4795  -0.1388   4.8983  17.3797
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 56.65072    0.92465  61.267  < 2e-16 ***
## Comp.1       0.32066    0.02478  12.938  < 2e-16 ***
## Comp.2      -0.16749    0.06631  -2.526 0.014298 *
## Comp.3      -0.13331    0.08342  -1.598 0.115455
## Comp.4       0.06807    0.08867   0.768 0.445795
## Comp.5      -0.29736    0.10787  -2.757 0.007794 **
## Comp.6       0.18760    0.13308   1.410 0.163968
## Comp.7       0.43231    0.19095   2.264 0.027329 *
## Comp.8       0.90063    0.22372   4.026 0.000167 ***
## Comp.9      -0.02430    0.40799  -0.060 0.952709
## Comp.10      2.81967    1.58667   1.777 0.080796 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.681 on 58 degrees of freedom
## Multiple R-squared:  0.7844, Adjusted R-squared:  0.7472
## F-statistic:  21.1 on 10 and 58 DF,  p-value: 7.627e-16
```