Capstone Initial Results and Code

Connell O'Leary

March 20, 2017

```
setwd("C:/Users/Connell/Documents/R")
df1 <- read.csv("Independent.Variables.Total.csv")
```

Correlation for Language Test Scores Grades 4&6, 6&8, 4&8

```
cor(df1[,c(2,3)], use="complete.obs", method="pearson")

##                   Grade.4.Language Grade.6.Language
## Grade.4.Language         1.0000000        0.9641289
## Grade.6.Language         0.9641289        1.0000000

cor(df1[,c(3,4)], use="complete.obs", method="pearson")

##                   Grade.6.Language Grade.8.Language
## Grade.6.Language         1.0000000        0.9818791
## Grade.8.Language         0.9818791        1.0000000

cor(df1[,c(2,4)], use="complete.obs", method="pearson")

##                   Grade.4.Language Grade.8.Language
## Grade.4.Language         1.0000000        0.9929347
## Grade.8.Language         0.9929347        1.0000000
```

Correlation for Math Test Scores Grades 4&6, 6&8, 4&8

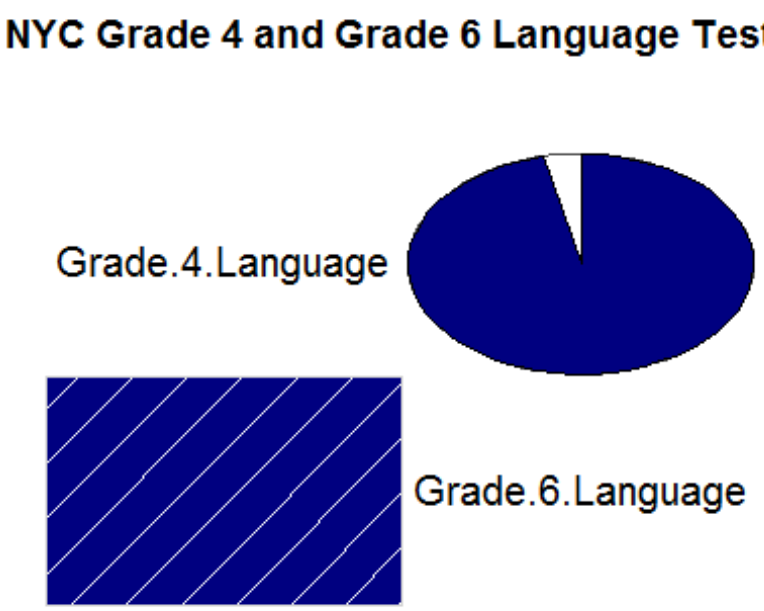
```
cor(df1[,c(5,6)], use="complete.obs", method="pearson")

##               Grade.4.Math Grade.6.Math
## Grade.4.Math     1.000000     0.975553
## Grade.6.Math     0.975553     1.000000

cor(df1[,c(6,7)], use="complete.obs", method="pearson")

##               Grade.6.Math Grade.8.Math
## Grade.6.Math    1.0000000    0.9806946
## Grade.8.Math    0.9806946    1.0000000

cor(df1[,c(5,7)], use="complete.obs", method="pearson")

##               Grade.4.Math Grade.8.Math
## Grade.4.Math    1.0000000    0.9485096
## Grade.8.Math    0.9485096    1.0000000
```

Assigning Colours for each Borough

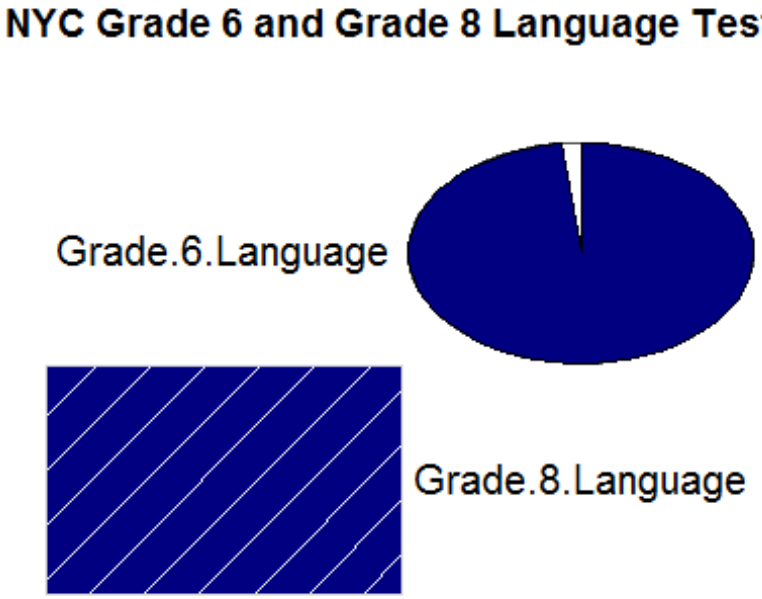
```
df1$color<- factor(df1$Borough, levels=c("Manhattan", "Bronx", "Brooklin",
"Queens", "Staten Island"), labels=rainbow(5))
```

Visulaizing Correlations - corrgram

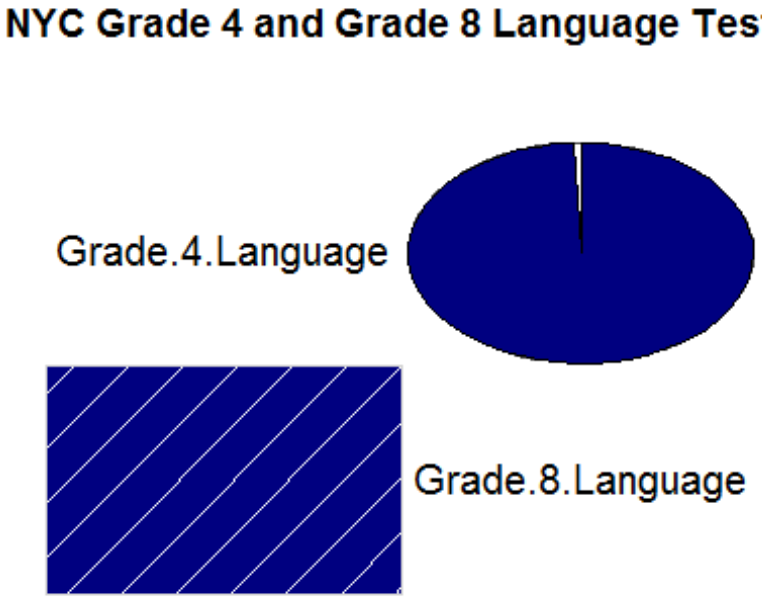
```
library(corrgram)

## Warning: package 'corrgram' was built under R version 3.2.5

## Warning: replacing previous import by 'magrittr::%>%' when loading
## 'dendextend'

corrgram(df1[,2:3], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grade 4 and Grade 6 Language Tests")
```



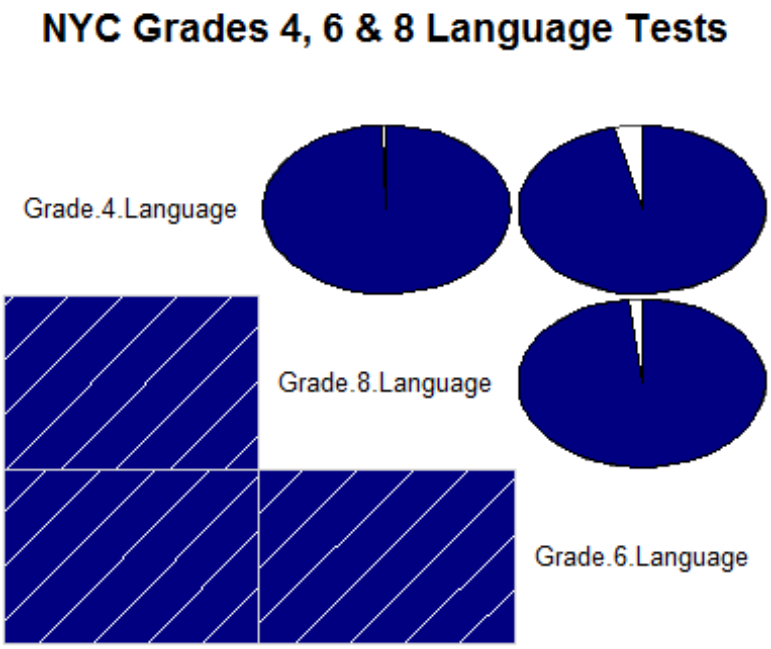
```
corrgram(df1[,3:4], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grade 6 and Grade 8 Language Tests")
```



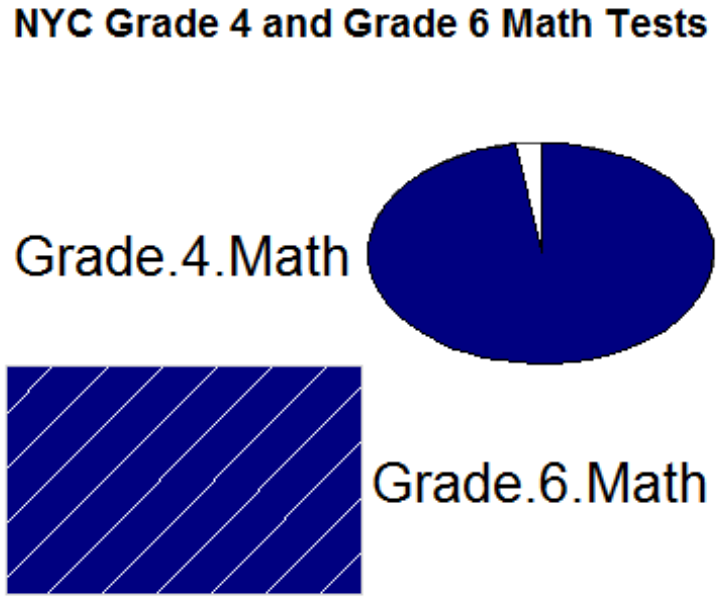
```
corrgram(df1[, c(2,4)], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grade 4 and Grade 8 Language Tests")
```



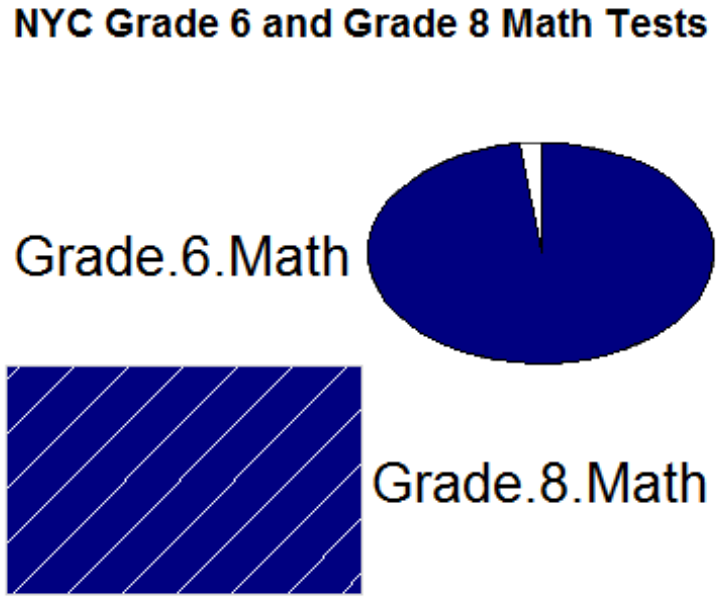
```
corrgram(df1[, 2:4], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grades 4, 6 & 8 Language Tests")
```



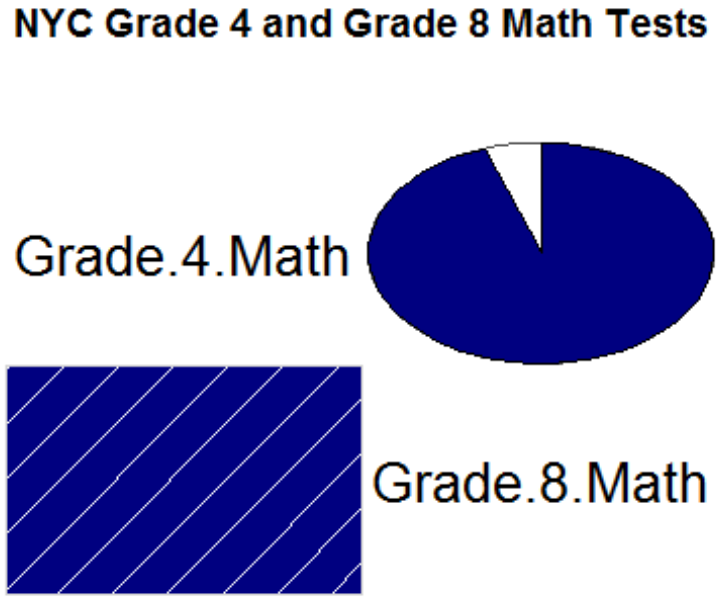
```
corrgram(df1[, 5:6], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grade 4 and Grade 6 Math Tests")
```



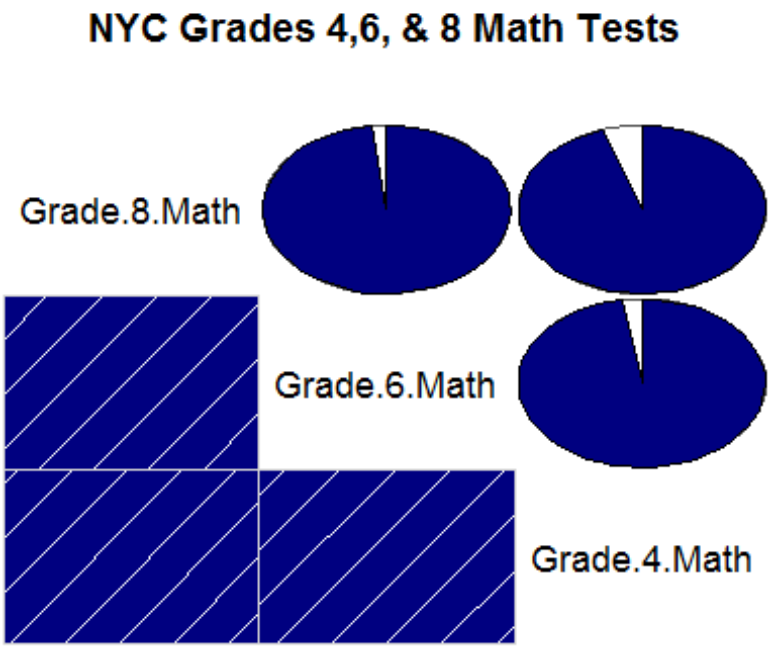
```
corrgram(df1[, 6:7], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grade 6 and Grade 8 Math Tests")
```



```
corrgram(df1[, c(5,7)], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grade 4 and Grade 8 Math Tests")
```

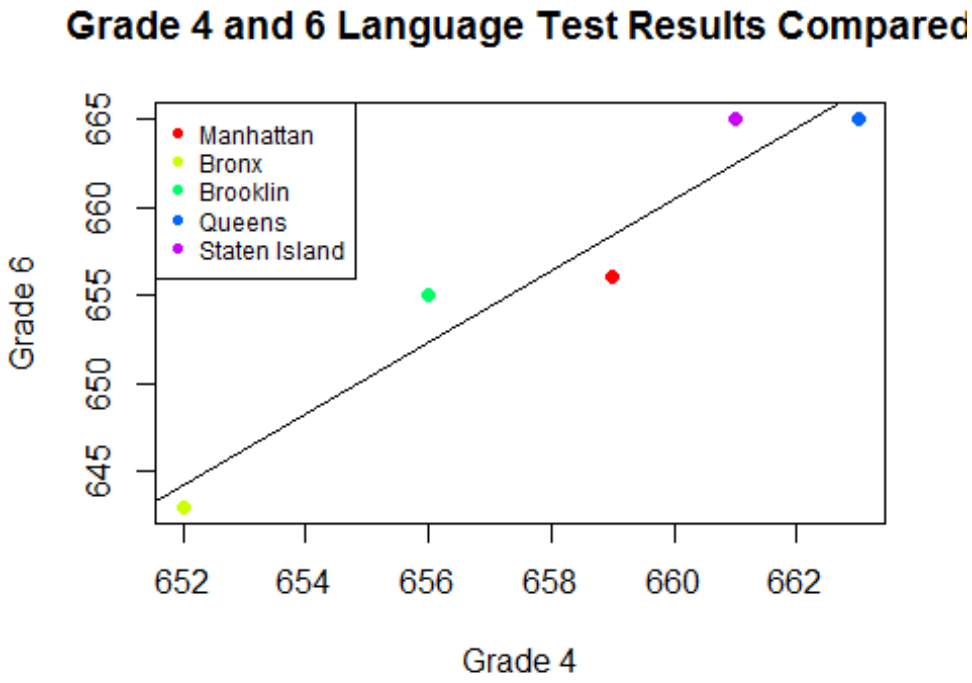


```
corrgram(df1[, 5:7], order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
         main="NYC Grades 4,6, & 8 Math Tests")
```

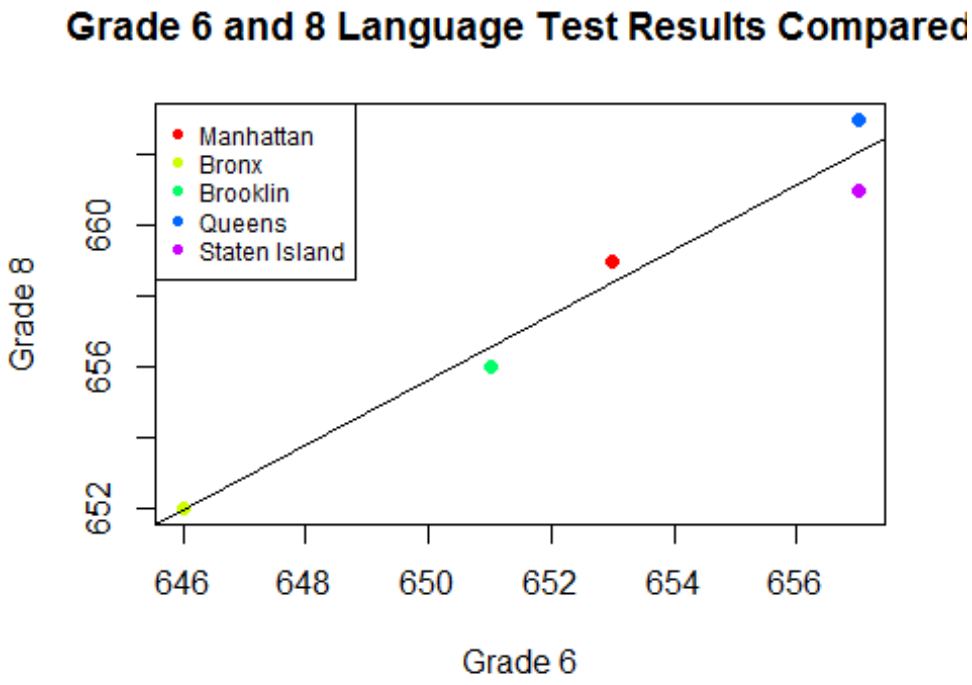


Scatter Plots for Language Correlation

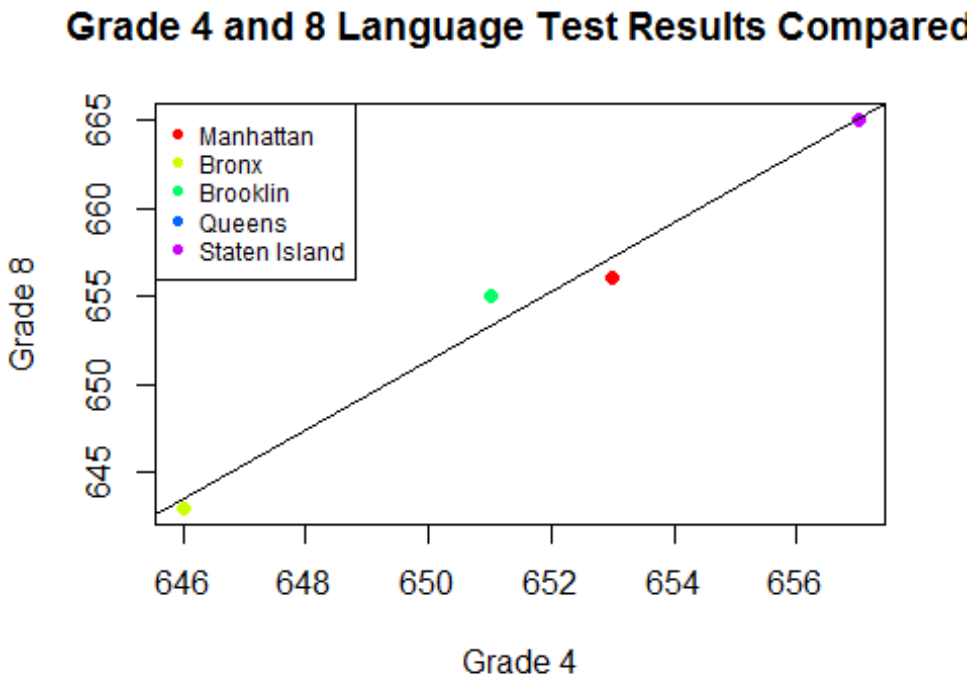
```
plot(df1$Grade.4.Language~df1$Grade.6.Language, main="Grade 4 and 6 Language
Test Results Compared",
     xlab="Grade 4", ylab = "Grade 6", pch=16, col=as.character(df1$color))
legend("topleft", pch=16, col=as.character(df1$color),
legend=unique(df1$Borough), text.font = 1, cex = 0.75)
abline(lm(df1$Grade.4.Language~df1$Grade.6.Language), col="black")
```



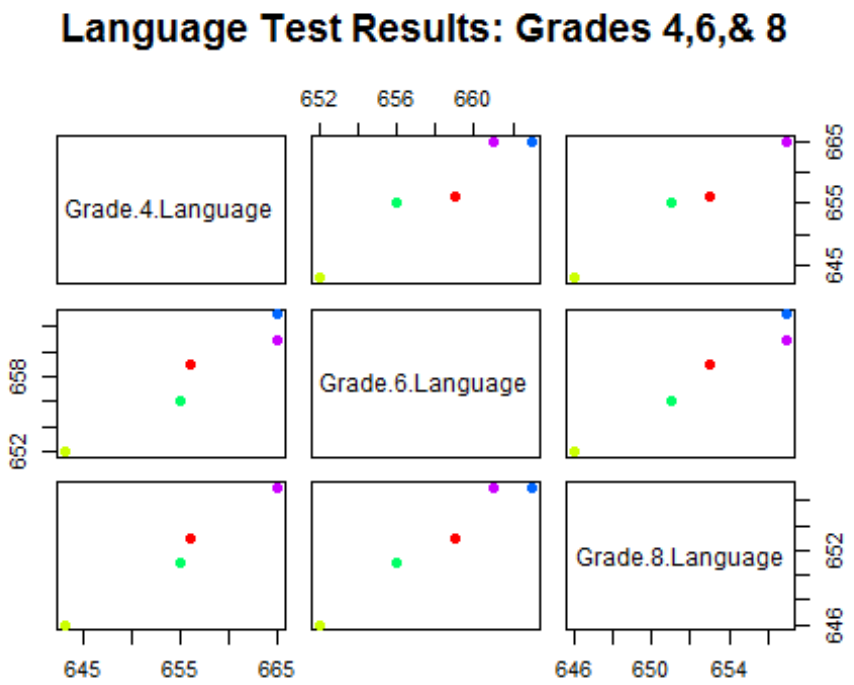
```r
plot(df1$Grade.6.Language~df1$Grade.8.Language, main="Grade 6 and 8 Language
Test Results Compared",
     xlab="Grade 6", ylab = "Grade 8", pch=16, col=as.character(df1$color))
legend("topleft", pch=16, col=as.character(df1$color),
legend=unique(df1$Borough), text.font = 1, cex = 0.75)
abline(lm(df1$Grade.6.Language~df1$Grade.8.Language), col="black")
```



```r
plot(df1$Grade.4.Language~df1$Grade.8.Language, main="Grade 4 and 8 Language
Test Results Compared",
     xlab="Grade 4", ylab = "Grade 8", pch=16, col=as.character(df1$color))
legend("topleft", pch=16, col=as.character(df1$color),
legend=unique(df1$Borough), text.font = 1, cex = 0.75)
abline(lm(df1$Grade.4.Language~df1$Grade.8.Language), col="black")
```

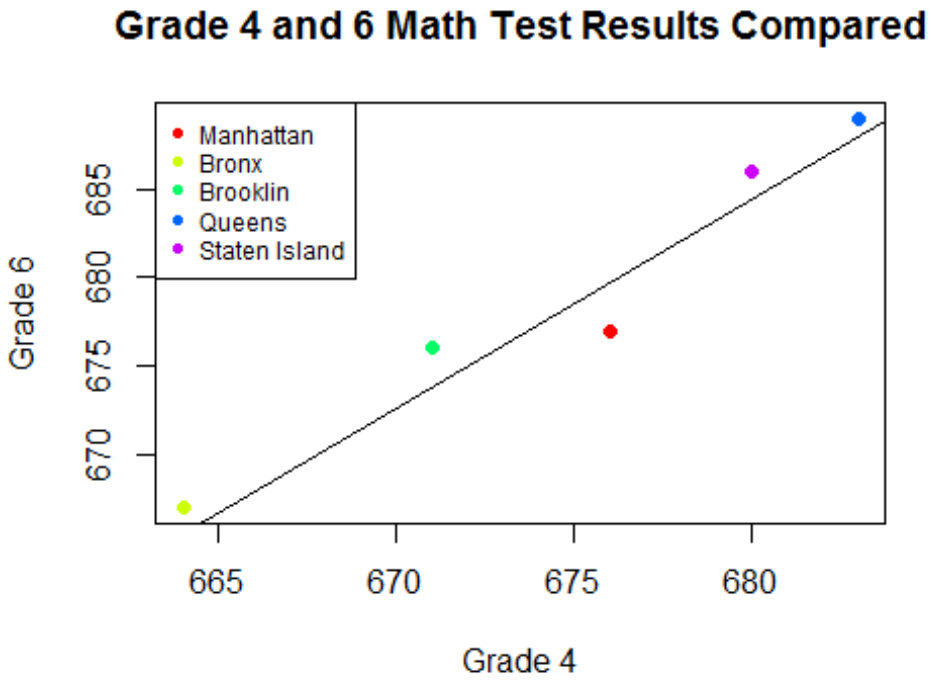


```
pairs(df1[,c(2:4)], main="Language Test Results: Grades 4,6,& 8",
col=as.character(df1$color), pch=16)
```

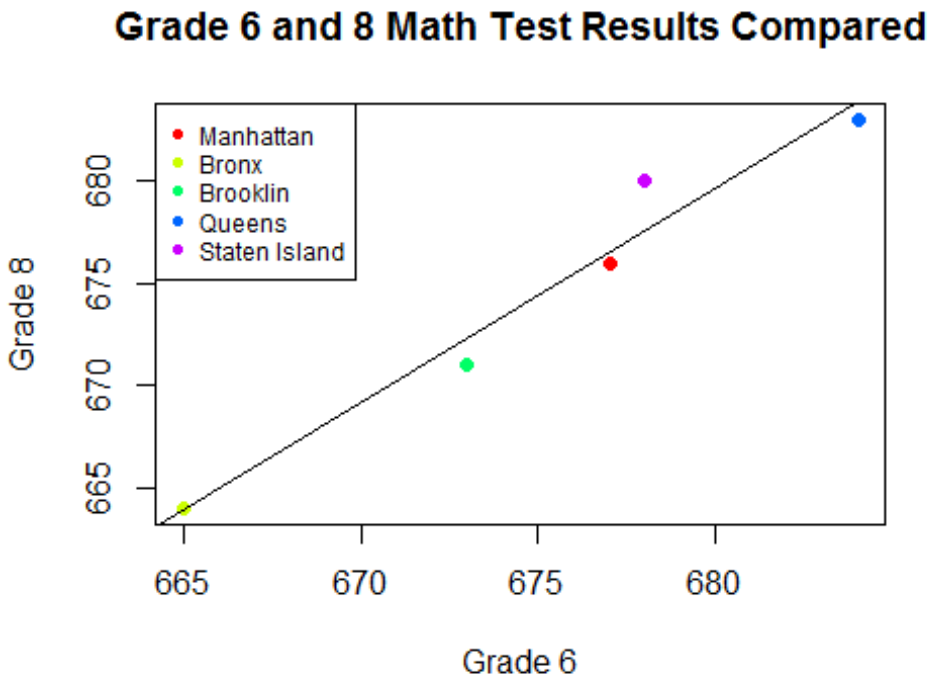


Scatter Plots for Math Correlation

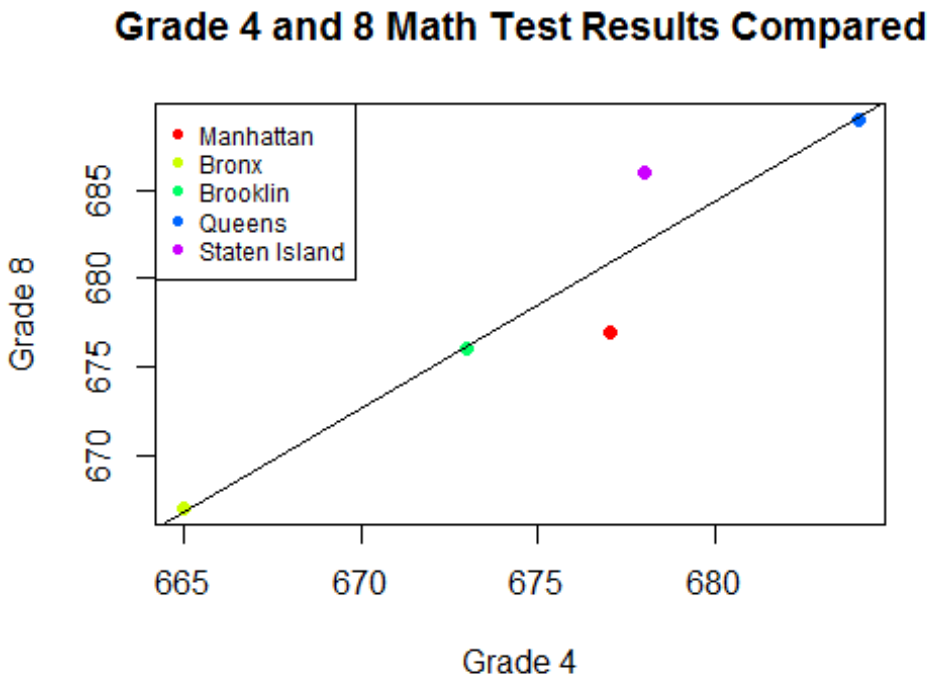
```
plot(df1$Grade.4.Math~df1$Grade.6.Math, main="Grade 4 and 6 Math Test Results
Compared",
     xlab="Grade 4", ylab = "Grade 6", pch=16, col=as.character(df1$color))
legend("topleft", pch=16, col=as.character(df1$color),
legend=unique(df1$Borough), text.font = 1, cex = 0.75)
abline(lm(df1$Grade.4.Math~df1$Grade.8.Math), col="black")
```



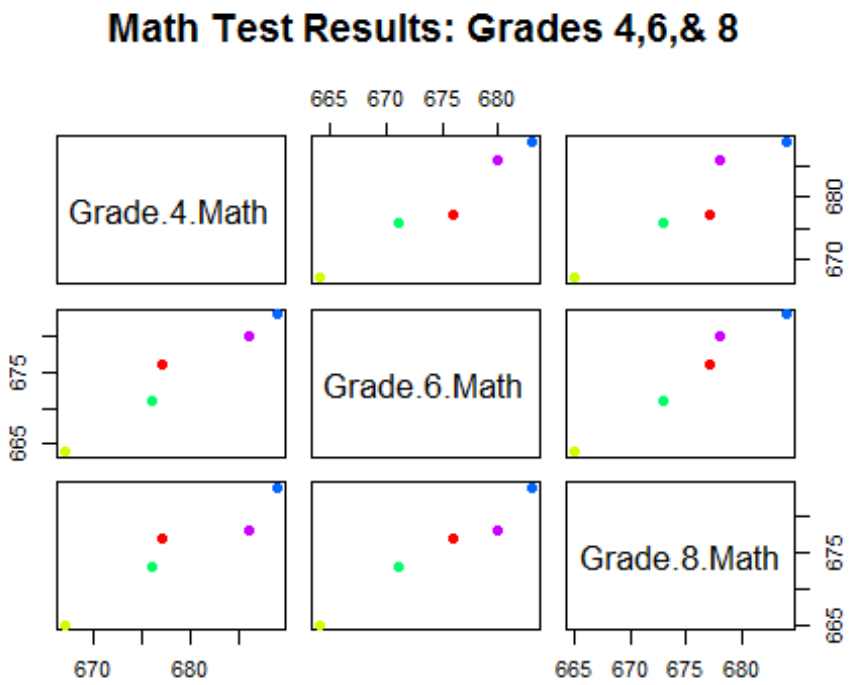
```r
plot(df1$Grade.6.Math~df1$Grade.8.Math, main="Grade 6 and 8 Math Test Results
Compared",
     xlab="Grade 6", ylab = "Grade 8", pch=16, col=as.character(df1$color))
legend("topleft", pch=16, col=as.character(df1$color),
legend=unique(df1$Borough), text.font = 1, cex = 0.75)
abline(lm(df1$Grade.6.Math~df1$Grade.8.Math), col="black")
```



```
plot(df1$Grade.4.Math~df1$Grade.8.Math, main="Grade 4 and 8 Math Test Results
Compared",
     xlab="Grade 4", ylab = "Grade 8", pch=16, col=as.character(df1$color))
legend("topleft", pch=16, col=as.character(df1$color),
legend=unique(df1$Borough), text.font = 1, cex = 0.75)
abline(lm(df1$Grade.4.Math~df1$Grade.8.Math), col="black")
```



```
pairs(df1[,c(5:7)], main="Math Test Results: Grades 4,6,& 8",
col=as.character(df1$color), pch=16)
```



Multivariate Regression - Grade 4 and 2008 Class Size

```
mod.grade4 <- lm(cbind(Grade.4.Language, Grade.4.Math)~Class.Size.2008, data
= df1)
mod.grade4

##
## Call:
## lm(formula = cbind(Grade.4.Language, Grade.4.Math) ~ Class.Size.2008,
##     data = df1)
##
## Coefficients:
##                   Grade.4.Language  Grade.4.Math
## (Intercept)       542.500           562.000
## Class.Size.2008     4.885             5.000
```

```r
summary(mod.grade4)
```

```
## Response Grade.4.Language :
##
## Call:
## lm(formula = Grade.4.Language ~ Class.Size.2008, data = df1)
##
## Residuals:
##       1        2        3        4        5
##  6.0385 -11.8462   0.1538   5.2692   0.3846
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      542.500     84.851   6.394  0.00775 **
## Class.Size.2008    4.885      3.623   1.348  0.27031
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.261 on 3 degrees of freedom
## Multiple R-squared:  0.3773, Adjusted R-squared:  0.1698
## F-statistic: 1.818 on 1 and 3 DF,  p-value: 0.2703
##
##
## Response Grade.4.Math :
##
## Call:
## lm(formula = Grade.4.Math ~ Class.Size.2008, data = df1)
##
## Residuals:
##   1   2   3   4   5
##   5 -10  -1   7  -1
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      562.000     78.672   7.144  0.00565 **
## Class.Size.2008    5.000      3.359   1.489  0.23334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.659 on 3 degrees of freedom
## Multiple R-squared:  0.4248, Adjusted R-squared:  0.2331
## F-statistic: 2.216 on 1 and 3 DF,  p-value: 0.2333
```

Multivariate Regression - Grade 6 and 2010 Class Size

```
mod.grade6 <- lm(cbind(Grade.6.Language, Grade.6.Math)~Class.Size.2010, data
= df1)
mod.grade6
```

```
##
## Call:
## lm(formula = cbind(Grade.6.Language, Grade.6.Math) ~ Class.Size.2010,
##     data = df1)
##
## Coefficients:
##                  Grade.6.Language  Grade.6.Math
## (Intercept)      633.4615          629.9615
## Class.Size.2010    0.9231            1.6731
```

```
summary(mod.grade6)
```

```
## Response Grade.6.Language :
##
## Call:
## lm(formula = Grade.6.Language ~ Class.Size.2010, data = df1)
##
## Residuals:
##       1       2       3       4       5
##  3.3846 -5.4615 -1.4615  3.6923 -0.1538
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     633.4615    25.7055  24.643 0.000146 ***
## Class.Size.2010   0.9231     0.9564   0.965 0.405660
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.362 on 3 degrees of freedom
## Multiple R-squared:  0.2369, Adjusted R-squared:  -0.01741
## F-statistic: 0.9315 on 1 and 3 DF,  p-value: 0.4057
##
##
## Response Grade.6.Math :
##
## Call:
## lm(formula = Grade.6.Math ~ Class.Size.2010, data = df1)
##
## Residuals:
##       1       2       3       4       5
##  5.8846 -9.4615 -2.4615  6.1923 -0.1538
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      629.962     44.177  14.260 0.000747 ***
```

```
## Class.Size.2010    1.673      1.644   1.018 0.383663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.496 on 3 degrees of freedom
## Multiple R-squared:  0.2567, Adjusted R-squared:  0.008954
## F-statistic: 1.036 on 1 and 3 DF,  p-value: 0.3837
```

Multivariate Regression - Grade 8 and 2012 Class Size

```
mod.grade8 <- lm(cbind(Grade.8.Language, Grade.8.Math)~Class.Size.2012, data
= df1)
mod.grade8

##
## Call:
## lm(formula = cbind(Grade.8.Language, Grade.8.Math) ~ Class.Size.2012,
##     data = df1)
##
## Coefficients:
##                  Grade.8.Language  Grade.8.Math
## (Intercept)      599.259           615.167
## Class.Size.2012    1.926             2.167

summary(mod.grade8)

## Response Grade.8.Language :
##
## Call:
## lm(formula = Grade.8.Language ~ Class.Size.2012, data = df1)
##
## Residuals:
##        1        2        3        4        5
##  3.66667 -5.25926 -0.25926  1.88889 -0.03704
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      599.259     32.714  18.318 0.000355 ***
## Class.Size.2012    1.926      1.175   1.639 0.199757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.862 on 3 degrees of freedom
## Multiple R-squared:  0.4724, Adjusted R-squared:  0.2965
## F-statistic: 2.686 on 1 and 3 DF,  p-value: 0.1998
##
##
## Response Grade.8.Math :
##
## Call:
## lm(formula = Grade.8.Math ~ Class.Size.2012, data = df1)
```

```
##
## Residuals:
##       1       2       3       4       5
##  5.5000 -8.6667 -0.6667  6.0000 -2.1667
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      615.167     59.197  10.392   0.0019 **
## Class.Size.2012    2.167      2.126   1.019   0.3832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.988 on 3 degrees of freedom
## Multiple R-squared:  0.2571, Adjusted R-squared:  0.009466
## F-statistic: 1.038 on 1 and 3 DF,  p-value: 0.3832
```

Stepwise Regression Grade 4

```
lm(Grade.4.Language ~ Class.Size.2008 + Farmers.Markets.Days +
Total.Wifi.Hotposts + Total.Libraries + Violent.Crimes+ Property.Crimes +
Housing.Starts +
Total.Businesses + Vocational.Courses, data=df1, direction="both")

## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.

##
## Call:
## lm(formula = Grade.4.Language ~ Class.Size.2008 + Farmers.Markets.Days +
##     Total.Wifi.Hotposts + Total.Libraries + Violent.Crimes +
##     Property.Crimes + Housing.Starts + Total.Businesses +
Vocational.Courses,
##     data = df1, direction = "both")
##
## Coefficients:
##          (Intercept)       Class.Size.2008  Farmers.Markets.Days
##            264.16544              15.85314               0.02528
##  Total.Wifi.Hotposts       Total.Libraries        Violent.Crimes
##              0.03437               0.15765                   NA
##      Property.Crimes        Housing.Starts      Total.Businesses
##                   NA                   NA                   NA
##   Vocational.Courses
##                   NA

step(lm(Grade.4.Language ~ Class.Size.2008 + Total.Wifi.Hotposts +
Total.Libraries, data=df1, direction="both"))

## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
## Start:  AIC=-15.18
## Grade.4.Language ~ Class.Size.2008 + Total.Wifi.Hotposts + Total.Libraries
##
##                       Df Sum of Sq    RSS     AIC
## <none>                                0.05 -15.180
## - Total.Libraries      1     33.39  33.44  15.501
## - Total.Wifi.Hotposts  1    176.14 176.19  23.811
## - Class.Size.2008      1    321.21 321.26  26.814

##
## Call:
## lm(formula = Grade.4.Language ~ Class.Size.2008 + Total.Wifi.Hotposts +
##     Total.Libraries, data = df1, direction = "both")
##
## Coefficients:
##         (Intercept)      Class.Size.2008  Total.Wifi.Hotposts
##           276.93951             15.34311              0.03433
##     Total.Libraries
##             0.15743
```

```
step(lm(Grade.4.Language + Grade.4.Math ~ Class.Size.2008 +
Total.Wifi.Hotposts + Total.Libraries, data=df1, direction="both"))
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
## Start:  AIC=6.25
## Grade.4.Language + Grade.4.Math ~ Class.Size.2008 + Total.Wifi.Hotposts +
##     Total.Libraries
##
##                       Df Sum of Sq     RSS    AIC
## <none>                                3.52  6.250
## - Total.Libraries      1    162.28  165.80 23.507
## - Total.Wifi.Hotposts  1    604.83  608.35 30.007
## - Class.Size.2008      1   1201.27 1204.79 33.423

##
## Call:
## lm(formula = Grade.4.Language + Grade.4.Math ~ Class.Size.2008 +
##     Total.Wifi.Hotposts + Total.Libraries, data = df1, direction = "both")
##
## Coefficients:
##         (Intercept)      Class.Size.2008  Total.Wifi.Hotposts
##           600.53259             29.67126              0.06362
##     Total.Libraries
##             0.34708
```

Stepwise Regression Grade 6

```
lm(Grade.6.Language ~ Class.Size.2010 + Farmers.Markets.Days +
Total.Wifi.Hotposts + Total.Libraries + Violent.Crimes+ Property.Crimes +
Housing.Starts +
Total.Businesses + Vocational.Courses, data=df1, direction="both")
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
##
## Call:
## lm(formula = Grade.6.Language ~ Class.Size.2010 + Farmers.Markets.Days +
##     Total.Wifi.Hotposts + Total.Libraries + Violent.Crimes +
##     Property.Crimes + Housing.Starts + Total.Businesses +
Vocational.Courses,
##     data = df1, direction = "both")
##
## Coefficients:
##          (Intercept)       Class.Size.2010  Farmers.Markets.Days
##            616.25303               1.43066               -0.21449
##  Total.Wifi.Hotposts       Total.Libraries        Violent.Crimes
##              0.01713               0.08130                    NA
##      Property.Crimes        Housing.Starts      Total.Businesses
##                   NA                    NA                    NA
##   Vocational.Courses
##                   NA
```

```
step(lm(Grade.6.Language ~ Class.Size.2010 + Total.Wifi.Hotposts +
Total.Libraries, data=df1, direction="both"))
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
## Start:  AIC=6.2
## Grade.6.Language ~ Class.Size.2010 + Total.Wifi.Hotposts + Total.Libraries
##
##                       Df Sum of Sq    RSS     AIC
## <none>                               3.490  6.2023
## - Total.Libraries      1     9.312 12.802 10.7007
## - Total.Wifi.Hotposts  1    45.578 49.068 17.4189
## - Class.Size.2010      1    70.513 74.003 19.4733
```

```
##
## Call:
## lm(formula = Grade.6.Language ~ Class.Size.2010 + Total.Wifi.Hotposts +
##     Total.Libraries, data = df1, direction = "both")
##
## Coefficients:
##         (Intercept)      Class.Size.2010  Total.Wifi.Hotposts
##           551.14416              3.59436              0.01747
```

```
##     Total.Libraries
##             0.08314
```

```
step(lm(Grade.6.Language + Grade.6.Math ~ Class.Size.2010 +
Total.Wifi.Hotposts + Total.Libraries, data=df1, direction="both"))
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
## Start:  AIC=15.92
## Grade.6.Language + Grade.6.Math ~ Class.Size.2010 + Total.Wifi.Hotposts +
##     Total.Libraries
##
##                       Df Sum of Sq    RSS    AIC
## <none>                               24.39 15.925
## - Total.Libraries      1     66.55  90.95 20.504
## - Total.Wifi.Hotposts  1    340.37 364.77 27.449
## - Class.Size.2010      1    532.55 556.95 29.565
```

```
##
## Call:
## lm(formula = Grade.6.Language + Grade.6.Math ~ Class.Size.2010 +
##     Total.Wifi.Hotposts + Total.Libraries, data = df1, direction = "both")
##
## Coefficients:
##         (Intercept)      Class.Size.2010  Total.Wifi.Hotposts
##           1.039e+03            9.878e+00            4.773e-02
##     Total.Libraries
##           2.223e-01
```

Stepwise Regression Grade 8

```
lm(Grade.8.Language ~ Class.Size.2012 + Farmers.Markets.Days +
Total.Wifi.Hotposts + Total.Libraries + Violent.Crimes+ Property.Crimes +
Housing.Starts +
Total.Businesses + Vocational.Courses, data=df1, direction="both")
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
##
## Call:
## lm(formula = Grade.8.Language ~ Class.Size.2012 + Farmers.Markets.Days +
##     Total.Wifi.Hotposts + Total.Libraries + Violent.Crimes +
##     Property.Crimes + Housing.Starts + Total.Businesses +
Vocational.Courses,
##     data = df1, direction = "both")
##
## Coefficients:
##          (Intercept)       Class.Size.2012  Farmers.Markets.Days
##            467.81264              6.25424               0.14259
##  Total.Wifi.Hotposts       Total.Libraries        Violent.Crimes
```

```
##              0.01291               0.02855                     NA
##      Property.Crimes        Housing.Starts      Total.Businesses
##                   NA                    NA                     NA
##   Vocational.Courses
##                   NA
```

```
step(lm(Grade.8.Language ~ Class.Size.2012 + Total.Wifi.Hotposts +
Total.Libraries, data=df1, direction="both"))
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
## Start:  AIC=-3.75
## Grade.8.Language ~ Class.Size.2012 + Total.Wifi.Hotposts + Total.Libraries
##
##                       Df Sum of Sq    RSS     AIC
## <none>                               0.477 -3.7454
## - Total.Libraries      1     1.971  2.449  2.4307
## - Total.Wifi.Hotposts  1    39.903 40.380 16.4445
## - Class.Size.2012      1    83.501 83.979 20.1056
```

```
##
## Call:
## lm(formula = Grade.8.Language ~ Class.Size.2012 + Total.Wifi.Hotposts +
##     Total.Libraries, data = df1, direction = "both")
##
## Coefficients:
##         (Intercept)      Class.Size.2012  Total.Wifi.Hotposts
##           519.88589              4.51666              0.01400
##     Total.Libraries
##             0.03734
```

```
step(lm(Grade.8.Language + Grade.8.Math ~ Class.Size.2012 +
Total.Wifi.Hotposts + Total.Libraries, data=df1, direction="both"))
```

```
## Warning in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
## extra argument 'direction' is disregarded.
```

```
## Start:  AIC=-21.44
## Grade.8.Language + Grade.8.Math ~ Class.Size.2012 + Total.Wifi.Hotposts +
##     Total.Libraries
##
##                       Df Sum of Sq    RSS     AIC
## <none>                               0.01 -21.439
## - Total.Libraries      1     47.75  47.76  17.284
## - Total.Wifi.Hotposts  1    270.06 270.07  25.946
## - Class.Size.2012      1    513.87 513.88  29.163
```

```
##
## Call:
## lm(formula = Grade.8.Language + Grade.8.Math ~ Class.Size.2012 +
##     Total.Wifi.Hotposts + Total.Libraries, data = df1, direction = "both")
```

```
##
## Coefficients:
##         (Intercept)      Class.Size.2012  Total.Wifi.Hotposts
##           993.89737             11.20461              0.03642
##     Total.Libraries
##             0.18379
```