NYSAT

EKLAR

null

# Links to the Datasource

* NYC sat scores for 2012-<https://data.cityofnewyork.us/Education/2012-SAT-Results/f9bf-2cp4>
* school accountability- <https://data.cityofnewyork.us/Education/2006-2012-School-Demographics-and-Accountability-S/ihfw-zy9j>
* NYC general education survey-<https://data.cityofnewyork.us/Education/2012-NYC-General-Education-School-Survey/xiyj-m4sj>

#include summary of dataset   
summary(df)

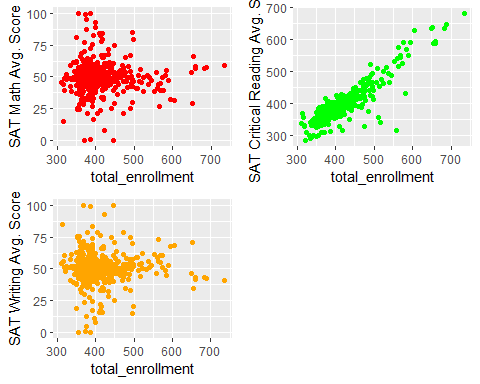
## DBN SCHOOL NAME Num of SAT Test Takers total\_enrollment  
## Length:407 Length:407 Min. :279.0 Min. :312.0   
## Class :character Class :character 1st Qu.:368.0 1st Qu.:371.5   
## Mode :character Mode :character Median :391.0 Median :395.0   
## Mean :400.7 Mean :413.8   
## 3rd Qu.:416.0 3rd Qu.:436.5   
## Max. :679.0 Max. :735.0   
## SAT Critical Reading Avg. Score SAT Math Avg. Score SAT Writing Avg. Score  
## Min. :286.0 Min. : 0.00 Min. : 0.00   
## 1st Qu.:360.0 1st Qu.: 44.05 1st Qu.: 45.10   
## Median :381.0 Median : 49.80 Median : 50.20   
## Mean :394.3 Mean : 49.82 Mean : 50.18   
## 3rd Qu.:411.0 3rd Qu.: 54.90 3rd Qu.: 55.95   
## Max. :682.0 Max. :100.00 Max. :100.00   
## male\_per female\_per black\_per white\_per   
## Min. : 0.00 Min. : 0.000 Min. : 2.40 Min. : 0.000   
## 1st Qu.:20.15 1st Qu.: 0.900 1st Qu.: 18.75 1st Qu.: 1.000   
## Median :33.20 Median : 1.800 Median : 43.50 Median : 2.900   
## Mean :39.31 Mean : 7.748 Mean : 42.73 Mean : 9.398   
## 3rd Qu.:57.05 3rd Qu.: 7.450 3rd Qu.: 62.45 3rd Qu.: 9.250   
## Max. :95.70 Max. :82.100 Max. :100.00 Max. :89.500   
## hispanic\_per asian\_per ell\_percent frl\_percent   
## Min. : 0.00 Min. : 15.80 Min. :279.0 Min. :312.0   
## 1st Qu.: 2.70 1st Qu.: 58.90 1st Qu.:368.0 1st Qu.:371.5   
## Median : 6.70 Median : 69.80 Median :391.0 Median :395.0   
## Mean :12.98 Mean : 66.16 Mean :400.7 Mean :413.8   
## 3rd Qu.:13.75 3rd Qu.: 76.80 3rd Qu.:416.0 3rd Qu.:436.5   
## Max. :94.60 Max. :100.00 Max. :679.0 Max. :735.0

# Explorotary data analysis

# How SAT is scored-(Ehtesham)  
  
# check the code in the Joining the three datasets section to make sure that   
 # we have the nessesary variables + any accuracy issues (Ehtesham)  
  
# add the survey response columns from all\_joined to df dataframe (laknath)  
  
# make characrter columns numeric that have SAT scores and percentage (Abinav)  
  
# Based on the real data deternime how we combine scores/or whether to keep seperate (Abhinav)  
 # should include histogram also boxplot (abhinav)  
  
# see if there is difference in aveage scores between just high scool vs those other types  
 # see if tere is difference between male and female (Ehtesham)  
  
  
# correlation plot (multi coleniarity) -Raj  
 # also include the numbers -Raj  
  
# plot histogram of (distribution) of SAT scores -Raj/Kushboo

# Determine if there is a relationship between enrollment and SAT scores-scatterplot (Laknath)

pl\_math<-ggplot(data = df)+  
 geom\_point(mapping = aes(x = total\_enrollment,y = `SAT Math Avg. Score`),  
 color="red",position = "jitter")  
  
pl\_reading<-ggplot(data = df)+  
 geom\_point(mapping = aes(x = total\_enrollment,y = `SAT Critical Reading Avg. Score`),  
 color="green",position = "jitter")  
  
pl\_writing<-ggplot(data = df)+  
 geom\_point(mapping = aes(x = total\_enrollment,y = `SAT Writing Avg. Score`),  
 color="orange",position = "jitter")  
  
ggarrange(pl\_math,pl\_reading,pl\_writing)

 It looks like there is a positve linear association between enrollment and SAT reading scores