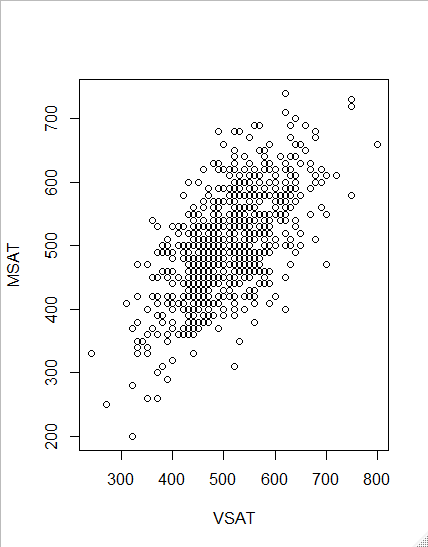
**Final Project**

**Part I: K-Means Clustering**

1. Read in SATData. Remove the missing values from the data frame.

2. a. Create a new data frame SAT\_MV that includes only the Math and Verbal SAT scores.

b. Plot the Math and Verbal SATs.



3. a. Separate the SAT scores into three clusters. Show the output of the Centers.

b. Make a plot of the three clusters. Describe the groups. (Here’s one way to do this, but you

may have a better way.)

plot(SAT\_VM[SAT\_VM.3means$cluster == 1, ], col = "red",

xlim=c(min(SAT\_VM[ ,1]),max(SAT\_VM[ ,1])),

ylim=c(min(SAT\_VM[ ,2]),max(SAT\_VM[ ,2])))

points(SAT\_VM[SAT\_VM.3means$cluster == 2,], col = "blue")

points(SAT\_VM[SAT\_VM.3means$cluster == 3,], col = "seagreen")

points(SAT\_VM.3means$centers,pch=2, col = "black")

Chart, scatter chart

Description automatically generated

c. Change the scaling on the plot so that it goes from the full span of SAT scores: 200 – 800.Chart, scatter chart

Description automatically generated Low = 200-500

Mediocre = 500-650

High = 650-800

This plot of SAT shows 3 different clustering’s.

The green cluster shows that the students that scored relatively low for their VSAT, also scored relatively low for there MSAT.

The Blue shows that the students the scored mediocrely for their VSAT, also scored mediocrely in their math SAT.

The red shows that the students that scored above average or highly on there MSAT, scored above average or also highly in there VSAT.

4. Separate the SAT scores into 4 clusters. Show the output from clusters. Make a plot of the 4 clusters. Is it easier to describe the groups from the four clusters as opposed to the three clusters?

Chart, scatter chart

Description automatically generatedLow = 200-450

Mediocre = 450-550

Above average = 550-700

High = 700-800

This plot of M & V SAT shows 4 different clustering’s.

The red cluster shows students that scored relatively low for their VSAT, also scored relatively low for there MSAT.

The Green shows students the scored mediocrely for their VSAT, also scored mediocrely in their math SAT.

The purple shows students that scored above average on there MSAT, scored above average on VSAT.

The blue shows that most of the students that scored highly on there MSAT, scored highly on their VSAT.