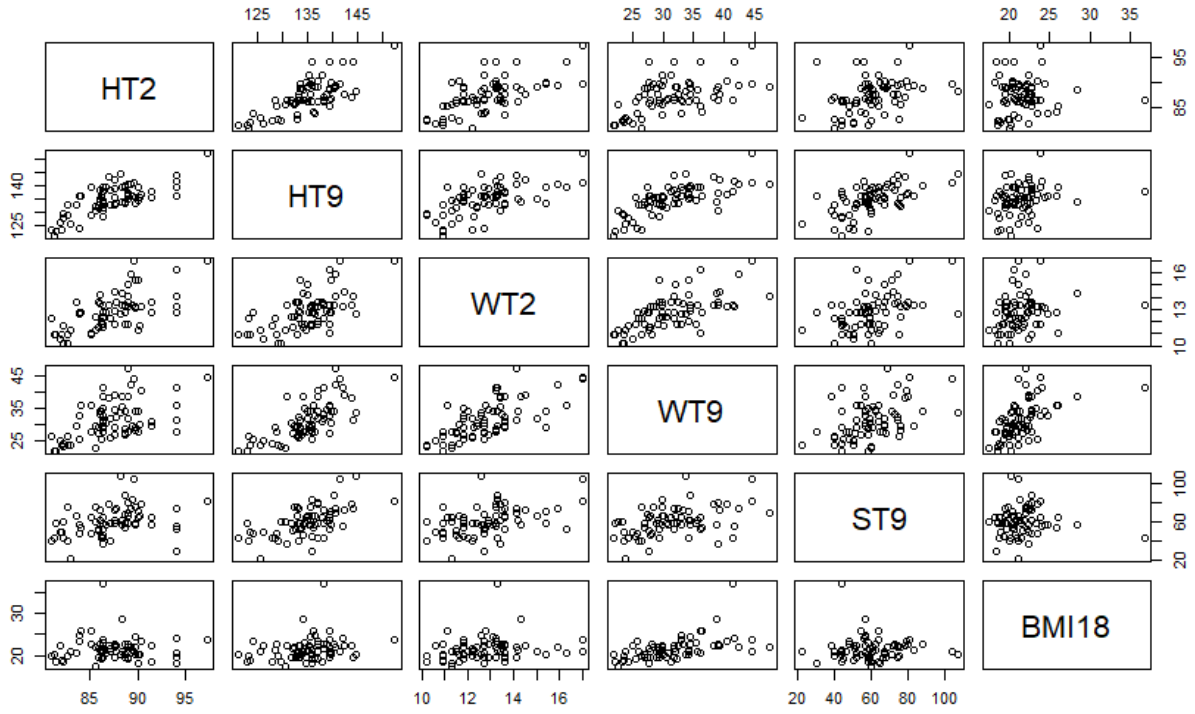
STAT 8030 HW 2

3.3

3.3.1

```
> ## 3.3
> # 3.3.1
>
> data(BGSgirls )
> head(BGSgirls)
    WT2  HT2  WT9   HT9  LG9 ST9 WT18  HT18 LG18 ST18 BMI18 Soma
67 13.6 87.7 32.5 133.4 28.4  74 56.9 158.9 34.6  143  22.5  5.0
68 11.3 90.0 27.8 134.8 26.9  65 49.9 166.0 33.8  117  18.1  4.0
69 17.0 89.6 44.4 141.5 31.9 104 55.3 162.2 35.1  143  21.0  5.5
70 13.2 90.3 40.5 137.1 31.8  79 65.9 167.8 39.3  148  23.4  5.5
71 13.3 89.4 29.9 136.1 27.7  83 62.3 170.9 36.3  152  21.3  4.5
72 11.3 85.5 22.8 130.6 23.4  60 47.4 164.9 31.8  126  17.4  3.0
> D = BGSgirls[,c(2,4,1,3,6,11)]
>
```

- ScatterPlot



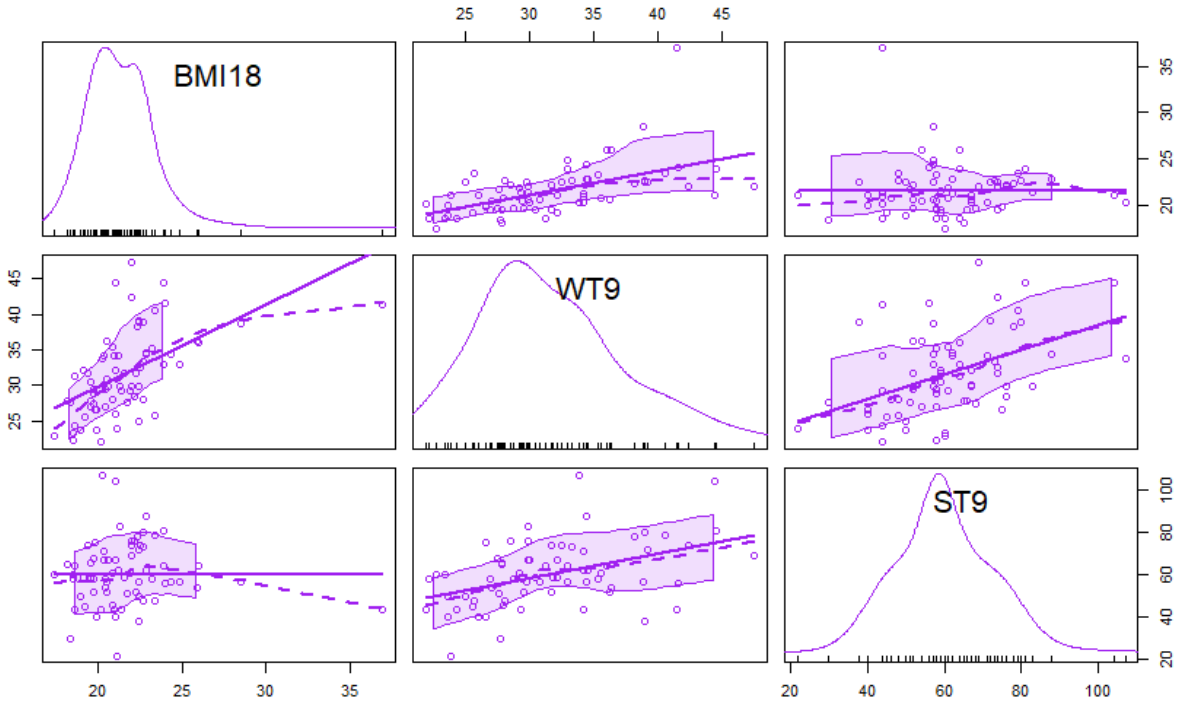
- Summary

  The regressions are all linear and It looks like an ideal case for multiple linear regression. The correlation gives out the same information as the scatterplot matrix.

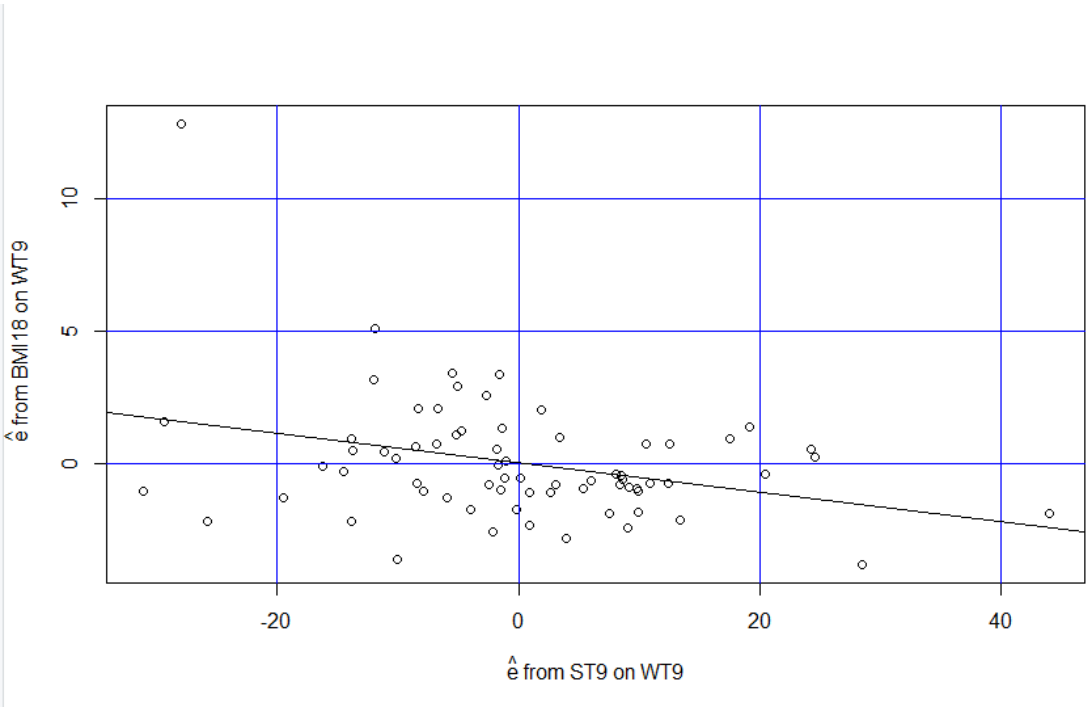
- Sample Correlation

```
> # Correlation
> cor(D)
              HT2       HT9       WT2       WT9          ST9       BMI18
HT2   1.00000000 0.7383562 0.6445495 0.5229277 0.361724146 0.042573733
HT9   0.73835617 1.0000000 0.6071247 0.7276123 0.603368147 0.236907969
WT2   0.64454954 0.6071247 1.0000000 0.6925390 0.451581158 0.190947873
WT9   0.52292768 0.7276123 0.6925390 1.0000000 0.453004062 0.545925753
ST9   0.36172415 0.6033681 0.4515812 0.4530041 1.000000000 0.005603061
BMI18 0.04257373 0.2369080 0.1909479 0.5459258 0.005603061 1.000000000
> |
```

3.3.2



The above image is the Marginal Plots of BMI18 vs WT9 vs ST9. We see there is no correlation between BMI18 and ST9. Now, we generate Added-Variable Plot.



From the above plot, if we see the fitted line, we can see that BMI18 and ST9 are negatively correlated on WT9. Also, going through the Y-axis on which BMI9 on WT9 is assigned, we can see there is a point with a very large value of BMI18. Even if that point is removed, there is still going to be a bad relationship between the, because a lot of point are negatively correlated.

3.3.3

```
> # 3.3.3
>
> M2=lm(BMI18~HT2+HT9+WT2+WT9+ST9,data=D)
> summary(M2)

Call:
lm(formula = BMI18 ~ HT2 + HT9 + WT2 + WT9 + ST9, data = D)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0948 -1.2186 -0.2533  1.0090 10.4951

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.855335   8.781156   3.514 0.000817 ***
HT2         -0.193997   0.130819  -1.483 0.142996
HT9          0.008057   0.096344   0.084 0.933613
WT2         -0.317779   0.278736  -1.140 0.258505
WT9          0.419762   0.075211   5.581  5.2e-07 ***
ST9         -0.044416   0.022219  -1.999 0.049853 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.14 on 64 degrees of freedom
Multiple R-squared:  0.4431,    Adjusted R-squared:  0.3996
F-statistic: 10.19 on 5 and 64 DF,  p-value: 3.294e-07
```

Sigma cap value – 2.14

R-square value – 0.4431

We can say that only these variables are significantly associated with BMI18 since the p-values of WT9, ST9 are less than 0.05 for the test of beta to be 0

4.2

4.2.1

```
> # 4.2.1
>
> data("Transact")
> p <- (Transact$t1+Transact$t2)/2
> q <- (Transact$t1-Transact$t2)
>
> M_1 <-lm(time~t1+t2,data = Transact)
> M_2 <-lm(time~a+d,data = Transact)
> M_3 <-lm(time~t2+d,data = Transact)
> M_4 <-lm(time~t1+t2+a+d,data = Transact)
>
```

We can see that p and q are exact linear combinations of t1 and t2, so only 2 items added after the intercept can be estimated and hence the other items are missing

4.2.2

Term	M1	M2	M3	M4
Intercept	144.369	144.369	144.369	144.369
t1	5.46			5.462
t2	2.035		7.497	2.035
Average trans time		7.497		

D		1.714	5.462	
Standard error	170.5	170.5	170.5	170.5
R square	0.9	0.9	0.9	0.9

As we can see above the t1 and t2 estimate is same for M1 and M2 and the standard error and R square are same for each fit. The t1 and t2 estimate for M4 are different.

4.2.3

In M1, with fixed t1, estimate is the change in response for a unit change in t2 whereas in M3, estimate is the change in Y for change in t2 when t1-t2 is fixed.

12.6

a.

12.6

a.) Mean score of the four groups are as follows:

1) Old program, English not spoken at home
$= \beta_0$

2.) Old program, English spoken at home
$= \beta_0 + \beta_2$

3.) New program, English not spoken at home
$= \beta_0 + \beta_1$

4.) New program, English spoken at home
$= \beta_0 + \beta_1 + \beta_2 + \beta_3$

b.

b.) For given students who dont speak English at home, it is the difference in mean score of the students who where in the new program and those who where in old program.

$$= [\beta_0 + \beta_1(1)] - [\beta_0 + \beta_1(0)]$$

$$= \beta_0 + \beta_1 - [\beta_0 + 0]$$

$$= \beta_0 + \beta_1 - \beta_0$$

$$= \beta_1$$

c.

c.) For given students who speak English at home, it is the mean score between the students who where in the new program and those who where in the old program.

$$= [\beta_0 + \beta_1(1) + \beta_2(1) + \beta_3(1)] - [\beta_0 + \beta_2]$$

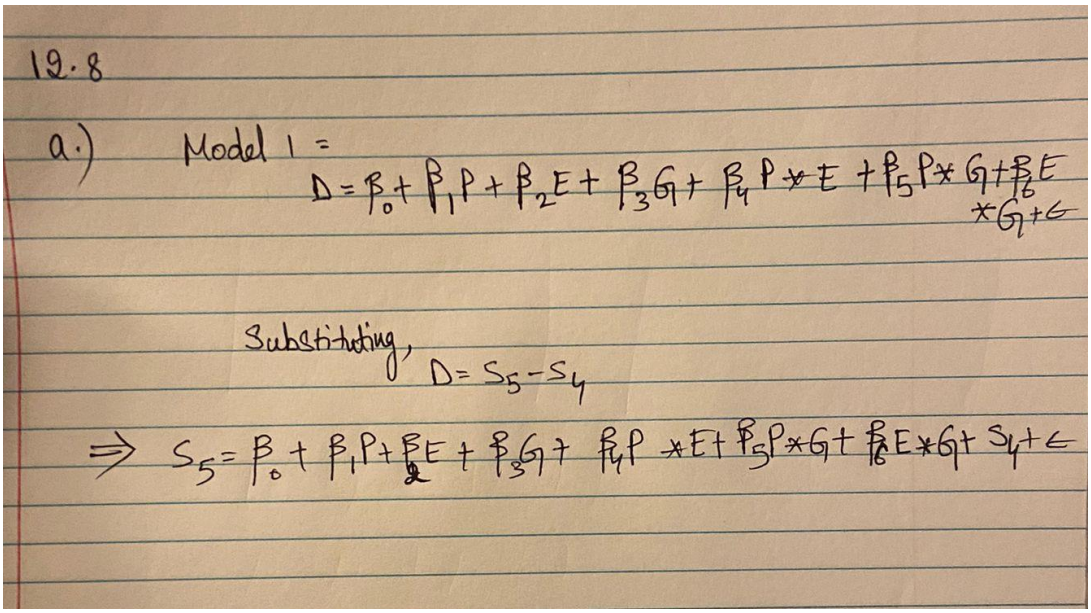
$$= \beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 - \beta_2$$

$$= \beta_1 + \beta_3$$

d. Term is very important in order to help the program difference to be different for English speaking homes and non-English speaking homes.

12.8

a.



b. It is because the scores of 4th grade influences the scores of 5th grade in model 2.