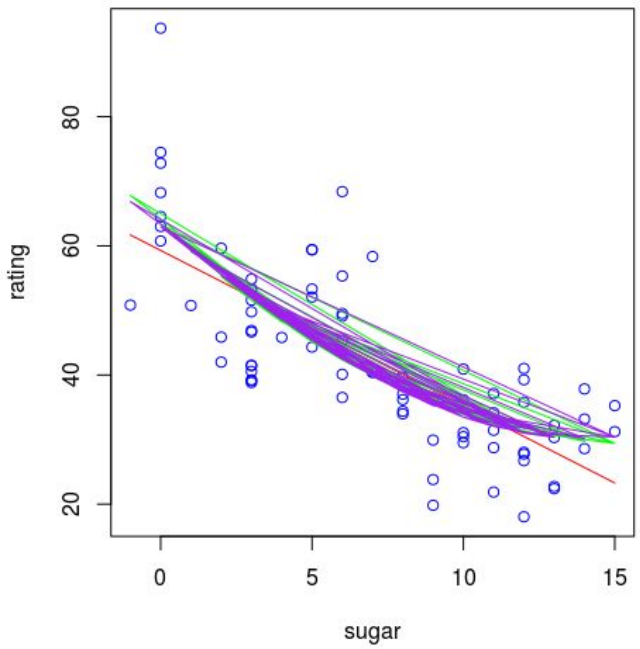
Report 2 on Nutrition

1. Correlation between sugar and rating comes out to be -0.7596747 suggesting a strong negative correlation between the two wherin when one increases the other decreases.



2. Fig. : Plot of rating vs sugar and other linear(red),quadratic(green) and cubic(purple) fits.

3. Correlation is non-linear
    a. Call:
    lm(formula = rating ~ sugar)

    Residuals:
    Min 1Q Median 3Q Max

-17.853 -5.677 -1.439 5.160 34.421

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	59.2844	1.9485	30.43	< 2e-16	***
sugar	-2.4008	0.2373	-10.12	1.15e-15	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.196 on 75 degrees of freedom
Multiple R-squared: 0.5771, Adjusted R-squared: 0.5715
F-statistic: 102.3 on 1 and 75 DF, p-value: 1.153e-15

Summary of Linear Model

b. all:
lm(formula = rating ~ poly(sugar, 2))

Residuals:

Min	1Q	Median	3Q	Max
-16.9733	-4.7005	-0.3318	4.5481	30.3675

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	42.666	1.015	42.040	< 2e-16	***
poly(sugar, 2)1	-93.031	8.906	-10.446	3.27e-16	***
poly(sugar, 2)2	21.750	8.906	2.442	0.017	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.906 on 74 degrees of freedom
Multiple R-squared: 0.6086, Adjusted R-squared: 0.5981

Summary of Quadratic Model

c. all:
lm(formula = rating ~ poly(sugar, 3))

Residuals:

Min	1Q	Median	3Q	Max
-16.0282	-5.0420	-0.3263	4.7912	30.7460

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	42.666	1.021	41.786	< 2e-16	***
poly(sugar, 3)1	-93.031	8.960	-10.383	5.01e-16	***
poly(sugar, 3)2	21.750	8.960	2.427	0.0177	*
poly(sugar, 3)3	2.963	8.960	0.331	0.7419	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.96 on 73 degrees of freedom
Multiple R-squared: 0.6092, Adjusted R-squared: 0.5932
F-statistic: 37.94 on 3 and 73 DF, p-value: 6.903e-15

Summary of Cubic model

3. From the above summary,we see R squared values is max for quadratic regression. Hence Quadratic is the best fit.