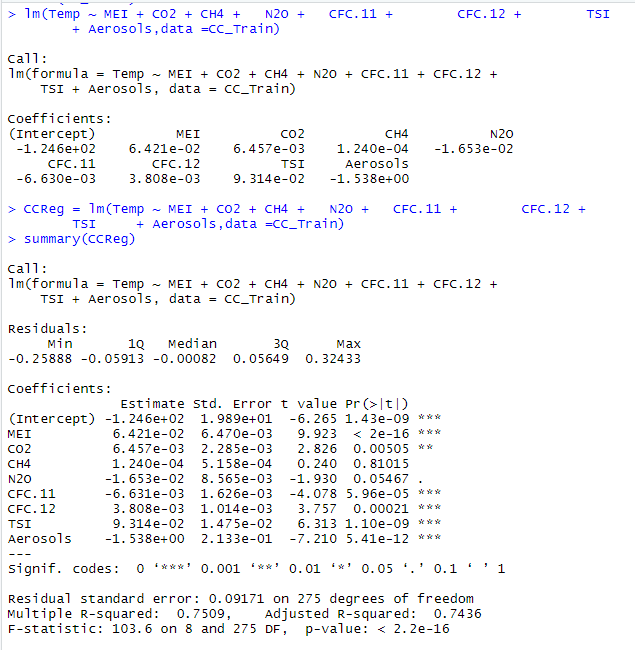
**Climate change**

The linear regression equation given by the model is –

Temp = -124.6 + 0.06421\*MEI + 0.006457\*CO2 + 0.0001240\*CH4 – 0.01653\*N2O - 0.006631\* CFC.11 + 0.003808\*CFC.12 + 0.09314\*TSI -1.538\*Aerosols



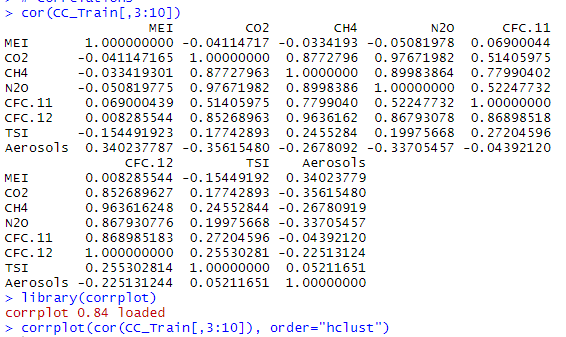
Quality of Model -

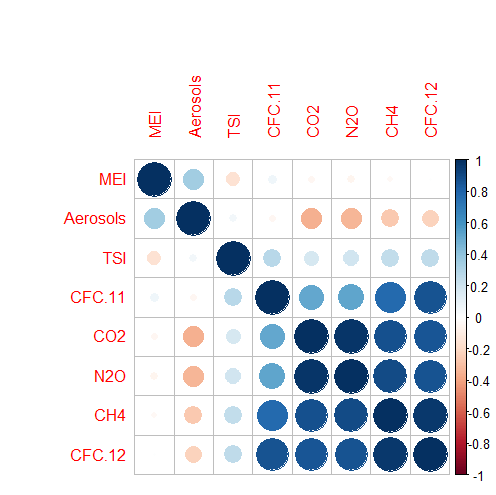
The R2 value is 0.7509. This implies that 75.09% of variation in the global temperatures in the training data can be explained by the linear regression model. This is a good quality level model which can be improved. The independent variables MEI, CO2, CFC.11, CFC.12, TSI and Aerosols are significant at 99% or higher significance level.

Multicollinearity -

The signs of regression coefficients for N2O and CFC-11 in the model are negative or the impact of these variables on global temperature is contradictory to the expected positive impact. This suggests the presence of multicollinearity in the data.

The correlations between independent variables are -





The independent variable N20 is highly correlated with 3 other independent variables –

CO2 having correlation of 0.977

CH4 having correlation of 0.9

CFC.12 having correlation of 0.868

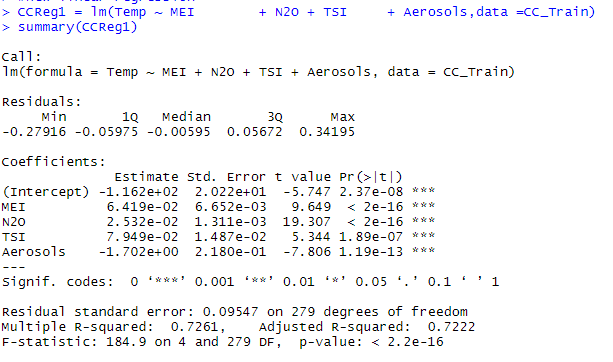
The independent variable CFC.11 is highly correlated with 2 other independent variables –

CH4 having correlation of 0.78

CFC-12 having correlation of 0.869

The updated regression model equation is –

Temp = -116.2 + 0.06419\*MEI + 0.02532\*N2O + 0.07949\*TSI – 1.702\*Aerosols



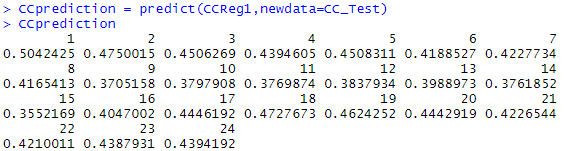
The coefficient for N2O in the updated model is 0.02532. Thus, removing multicollinear variables CO2, CH4, CFC.11 and CFC.12 corrects the impact of N2O on global temperature. The sign of the regression coefficient for N2O becomes positive in the updated model which is true to the expectation that increasing N2O atmospheric concentration increases global temperature.

Quality of new model -

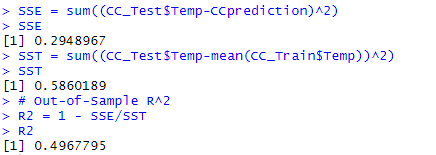
The R2 value reduces to 0.726. The adjusted R2 value also reduces to 0.722 from 0.744 in the previous model.

All the independent variables MEI, N2O, TSI and Aerosols are significant at 99% or higher significance level. Though the model’s explanatory power has reduced but all variables are highly significant and not correlated with each other. This may not be as good quality model as the previous model due to lower R2 . However, this model is still able to explain 72% variation in the global temperature in the training data. Thus, this is a better overall model than the previous model.

The prediction for test data are –



The R2 value for test data reduces to 0.497. This is not a good quality model.



This tells that a good R2  value for training data doesn’t imply good R2 for test data as well and the training model could be overfitting the training data and that is why it may not able to fit well to the new test data.