Massey University
228.371 Assignment 1

1. **Question**
   Suppose that a standard gas turbine has, on average, a heat rate of 9750kJ/kWh. Perform a t-test to see if the mean heat rate for the turbines in your data file exceeds 9750kJ/kWh. What do you conclude about your set of gas turbines compared to the average gas turbine?

   Solution

```
        One Sample t-test

data:  HeatRate
t = 5.0875, df = 31, p-value = 8.357e-06
alternative hypothesis: true mean is greater than 9750
95 percent confidence interval:
 10750.53      Inf
sample estimates:
mean of x
 11250.66
```

   Conclusion: The sample is not representative and the conclusions drawn from my data should not be applied to gas turbines in general.

2. **Question**
   The temperature is measured at two points for each turbine. Compare these two measurements and make a statement (backed with evidence) about the mean reduction in temperature from the inlet to the exhaust points of the turbines.

   Solution
   You must have used a paired t-test here.

```
        Paired t-test

data:  YourData$InletTemp and YourData$ExhTemp
t = 29.6404, df = 31, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 584.2062 670.5438
sample estimates:
mean of the differences
              627.375
```

   You could interpret this using the confidence interval but it would be better to use the hypothesis test. The reduction is significantly different to zero.

3. **Question**
   Do turbines with the advanced engine type have higher power than those with traditional engines? What assumptions are required if your findings are to be generalised to all advanced

and traditional turbines?

Solution

```
        Welch Two Sample t-test

data:  YourData$Power[YourData$Engine == "Advanced"] and YourData$Power[YourData$Engine == "
t = 3.1219, df = 9.323, p-value = 0.005886
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 49163.72      Inf
sample estimates:
mean of x mean of y
 160779.4   42352.0
```

This must be a one-sided hypothesis test. You should not use a confidence interval to answer this question.

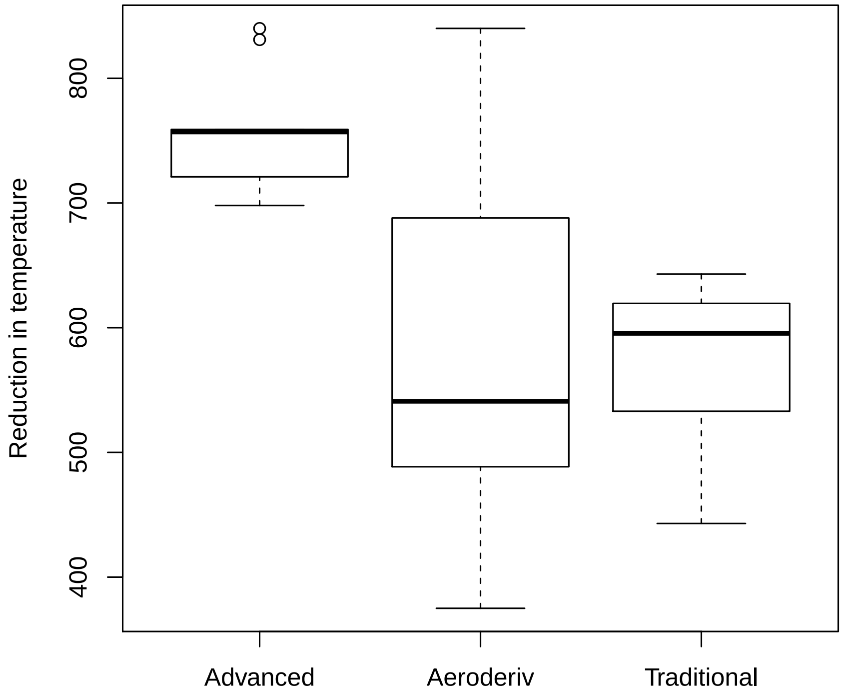
The advanced engines are more powerful than the traditional ones as the mean is significantly greater than the traditional mean.

If the outcome from this hypothesis test is to be used to generalise to a wider population of turbines, the sample used must have come from a random sample of each type of turbine.

4. **Question**
   Determine if the reduction in temperature from the inlet to the exhaust is different for the three types of turbine. (This should include a graphical representation as well as a model.)

   Solution

```
            Df Sum Sq Mean Sq F value Pr(>F)
Engine       2 208896  104448   12.86  1e-04 ***
Residuals   29 235529    8122
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1
```



There are at least one pair of engine types that have different power.

5. **Question**
   Conduct a hypothesis test to ascertain if there is potentially a linear relationship between air flow and power.

   Solution
   This should be done using a formal hypothesis test on correlation.

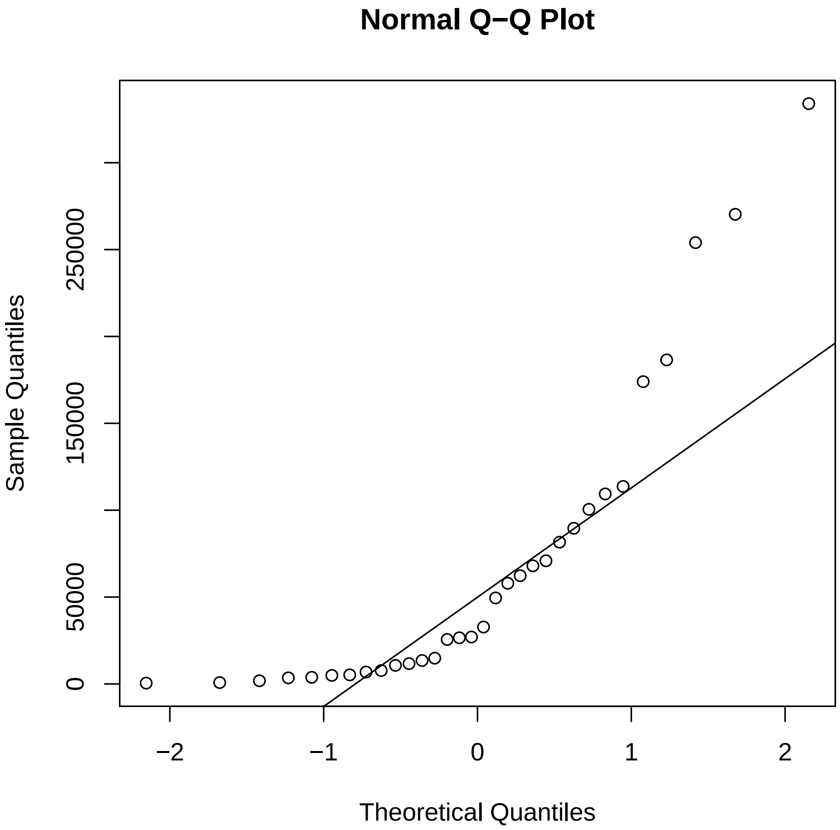
```
        Pearson's product-moment correlation

data:  YourData$Power and YourData$Airflow
t = 24.4559, df = 30, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9505811 0.9882525
sample estimates:
     cor
0.975826
```

   There is significant correlation between these two variables. It could mean there is a linear relationship.

6. **Question**
   Use both a graphical method and a formal hypothesis test to make a judgement about the

normality of the power of the turbines in your sample.

Solution
The graphical method comes from 228.271 - yes last year!!!



and the formal test:

The data are NOT normally distributed.

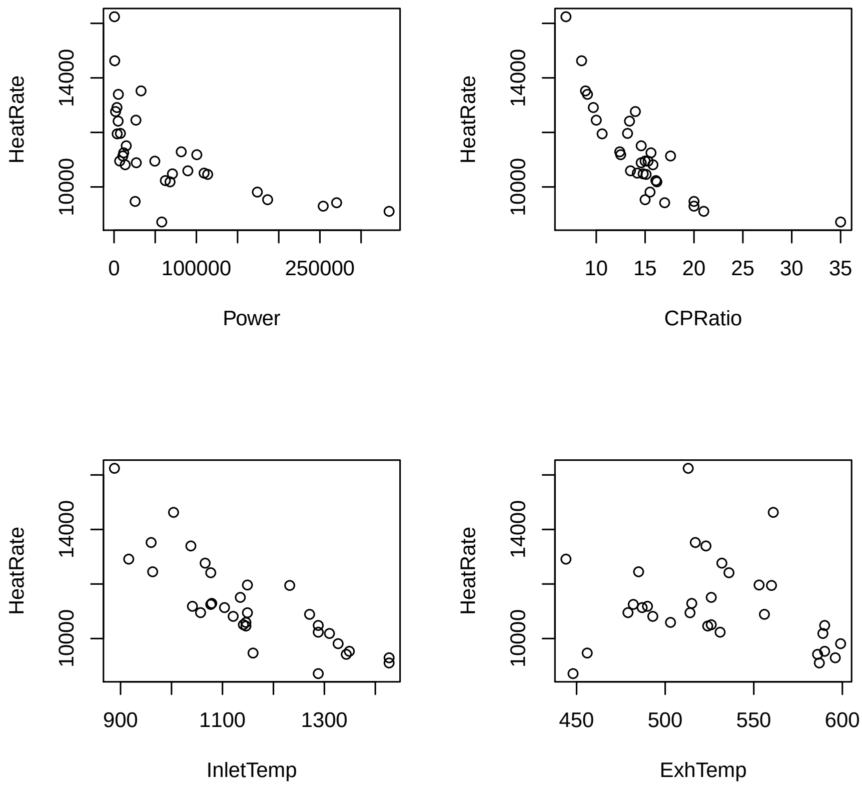
7. **Question**
Using the heat rate as the response variable, construct suitable graphs to identify which of the numeric variables might prove useful as predictors in a simple linear regression of the form $E(y) = \beta_0 + \beta_1 x$.

Solution

```
        Engine         Shafts           RPM             CPRatio
 Advanced   : 9   Min.   :1.000   Min.   : 3000   Min.   : 6.90
 Aeroderiv  : 7   1st Qu.:1.000   1st Qu.: 3600   1st Qu.:12.47
 Traditional:16   Median :1.000   Median : 5250   Median :14.70
                  Mean   :1.344   Mean   : 8542   Mean   :14.72
                  3rd Qu.:2.000   3rd Qu.:11160   3rd Qu.:15.88
                  Max.   :3.000   Max.   :33000   Max.   :35.00
   InletTemp        ExhTemp          Airflow           Power
 Min.   : 888   Min.   :444.0   Min.   :  3.00   Min.   :   486
 1st Qu.:1064   1st Qu.:492.2   1st Qu.: 28.25   1st Qu.:  7506
```

```
 Median :1144   Median :525.0   Median :137.50   Median : 29893
 Mean   :1155   Mean   :527.8   Mean   :197.31   Mean   : 69377
 3rd Qu.:1288   3rd Qu.:560.2   3rd Qu.:326.75   3rd Qu.: 92325
 Max.   :1427   Max.   :599.0   Max.   :737.00   Max.   :334000
    HeatRate       DiffTemps
 Min.   : 8714   Min.   :375.0
 1st Qu.:10221   1st Qu.:548.5
 Median :10948   Median :619.5
 Mean   :11251   Mean   :627.4
 3rd Qu.:12076   3rd Qu.:716.5
 Max.   :16243   Max.   :840.0
```

It doesn't make sense to plot graphs against categorical variables if a simple linear regression model is to be used. Also note that some variables have near categorical status as they have a small number of possible values.



You must have commented on which of the plots shows the greatest linearity and therefore suitability for use in a linear regression model.

8. **Question**
Use speed as a candidate for the predictor in a simple linear regression model to explain the heat rate of the turbines.

Solution

```
Call:
lm(formula = HeatRate ~ RPM)

Residuals:
    Min      1Q  Median      3Q     Max
-1646.9  -636.9  -142.7   322.9  3004.9

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.713e+03  2.661e+02  36.498  < 2e-16 ***
RPM         1.800e-01  2.352e-02   7.654 1.55e-08 ***
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 987.1 on 30 degrees of freedom
Multiple R-squared:  0.6613,       Adjusted R-squared:   0.65
F-statistic: 58.58 on 1 and 30 DF,  p-value: 1.547e-08
```

You should comment about the suitability of this model by investigating the p-value of the slope parameter in your model.

9. **Question**
   Choose another variable as the predictor in a simple linear regression model. Is this model better or worse than using speed as the sole predictor of heat rate?

   Solution
   Choosing which of the variables to use in this second model is up to you and must correspond to your working in Question 7. In my graph of HeatRate vs CPRatio, there is a nasty outlier which may well influence the model. You should have looked for this sort of thing in your data. I've chosen the InletTemp for illustration here.

```
Call:
lm(formula = HeatRate ~ InletTemp)

Residuals:
     Min       1Q   Median       3Q      Max
-1736.10  -744.62    97.27   530.23  2446.05

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 22258.585   1410.918   15.78 4.55e-16 ***
InletTemp      -9.529      1.212   -7.86 9.01e-09 ***
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 969.7 on 30 degrees of freedom
Multiple R-squared:  0.6731,       Adjusted R-squared:  0.6622
F-statistic: 61.78 on 1 and 30 DF,  p-value: 9.01e-09
```

In my case, using the InletTemp variable gave a reasonable model but on the basis of both R-squared and the standard deviation of the residuals, the Speed variable is a better predictor. For your data the inlet temperature would have been a better predictor.

N.B. your answer will have been checked against the plots given earlier.

10. **Question**
    Now fit a model that has both speed and the other variable you chose. Determine if the interaction of the two variables is required. You must therefore choose between the models $E(y) = \beta_0 + \beta_1 Speed + \beta_2 x_2$ and $E(y) = \beta_0 + \beta_1 Speed + \beta_2 x_2 + \beta_3 Speed \times x_2$.

    Solution
    You can use the anova() command:

```
Analysis of Variance Table

Model 1: HeatRate ~ RPM + InletTemp
Model 2: HeatRate ~ RPM * InletTemp
  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1     29 11092145
2     28 11063434  1     28711 0.0727 0.7895
```

or, look at the p-value of the interaction term in the print out of the model:

```
Call:
lm(formula = HeatRate ~ RPM * InletTemp)

Residuals:
     Min       1Q   Median       3Q      Max
-1184.60  -228.96    39.11   318.04  1522.42

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.733e+04  1.361e+03  12.730 3.65e-13 ***
RPM            1.510e-01  1.296e-01   1.165    0.254
InletTemp     -6.096e+00  1.162e+00  -5.248 1.41e-05 ***
RPM:InletTemp -3.471e-05  1.288e-04  -0.270    0.789
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 628.6 on 28 degrees of freedom
Multiple R-squared:  0.8718,       Adjusted R-squared:  0.8581
F-statistic: 63.48 on 3 and 28 DF,  p-value: 1.314e-12
```

In my case, the interaction term was not required. For completeness, the model without the interaction is:

```
Call:
lm(formula = HeatRate ~ RPM + InletTemp)

Residuals:
     Min       1Q   Median       3Q      Max
-1176.45  -253.22    27.88   315.86  1518.99

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 17516.8905  1145.4277  15.293 2.04e-15 ***
RPM             0.1164     0.0174   6.690 2.44e-07 ***
InletTemp      -6.2848     0.9126  -6.887 1.45e-07 ***
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1
```

```
Residual standard error: 618.5 on 29 degrees of freedom
Multiple R-squared:  0.8715,        Adjusted R-squared:  0.8626
F-statistic: 98.33 on 2 and 29 DF,  p-value: 1.201e-13
```

You must have shown the value of the interaction somehow.

11. **Question**
    Create a multiple regression model that uses all numeric variables in your data set as predictors of the heat rate. Do not include any interactions or polynomial terms.

    Solution

```
Call:
lm(formula = HeatRate ~ Shafts + RPM + CPRatio + InletTemp +
    ExhTemp + Airflow + Power)

Residuals:
    Min      1Q  Median      3Q     Max
-805.61 -267.49  -94.66  260.23  732.86

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.516e+04  1.503e+03  10.088 4.14e-10 ***
Shafts      -6.759e+01  2.093e+02  -0.323  0.74955
RPM          6.759e-02  2.173e-02   3.111  0.00476 **
CPRatio     -1.175e+01  3.775e+01  -0.311  0.75838
InletTemp   -1.006e+01  1.901e+00  -5.289 2.00e-05 ***
ExhTemp      1.433e+01  4.623e+00   3.101  0.00488 **
Airflow     -2.740e+00  2.443e+00  -1.122  0.27301
Power        5.262e-03  5.897e-03   0.892  0.38112
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 420.5 on 24 degrees of freedom
Multiple R-squared:  0.9508,        Adjusted R-squared:  0.9365
F-statistic:  66.3 on 7 and 24 DF,  p-value: 3.738e-14
```

    Your comments here should have focused on the value of this model. Look at the p-values for the various parameters. Are they all significant?

12. **Question**
    Give a practical interpretation of your estimates of the β's from this model.

    Solution
    For a unit increase:

```
- in Shafts there is a change in the HeatRate of -67.587714
- in RPM there is a change in the HeatRate of 0.067593
- in CPRatio there is a change in the HeatRate of -11.745688
- in InletTemp there is a change in the HeatRate of -10.056488
- in ExhTemp there is a change in the HeatRate of 14.334562
- in Airflow there is a change in the HeatRate of -2.740353
- in Power there is a change in the HeatRate of 0.005262
```

This assumes that all the other terms are in the model.

13. **Question**
Consider the standard deviation of the residuals from the models created thus far. Use this as a means of describing the value of the last model compared to the model that included only two predictors.

Solution
From my models above, the two values I need are:

```
[1] 618.456
```

```
[1] 420.5036
```

For the models used in this set of solutions, there appears to be a noticeable difference between these values. Your answer will depend on what you presented above.

14. **Question**
Interpret the R^2 value from the last model.

Solution
The model with all numeric variables included explains 95.08% of the variation in the HeatRate of the turbines.

15. **Question**
Is this model useful for predicting heat rate? Justify your answer using a hypothesis test for the utility of the entire model at a significance level of $\alpha = 0.01$.

Solution

Yes. The p-value from my model is less than 0.01, It was (rounded to 4 decimal places) 0

This means that some part of the model is useful. It does not say that this model is the best model.

16. **Question**
Investigate the leverages of your multiple regression model. Are there any points that are having undue influence on the model?

Solution
A summary of the leverages from the full multiple regression model is:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0735  0.1447  0.1873  0.2500  0.3422  0.7628
```

For my data, there were 1 observations that had leverages greater than twice the average leverage. We may need to investigate the addition of another variable in situations where there are a substantial number of observations with high leverage. If there was only one observation with high leverage then we would need to check that observation's suitability for inclusion in the sample that is used to create the model. (You could count the number of these points off the residual analysis plot showing Cook's distances.)

17. **Question**
Now create a reduced model by appropriately removing terms from your multiple regression model.

Solution
Using the stepwise algorithm in R, I get:

```
Start:  AIC=393.45
HeatRate ~ Shafts + RPM + CPRatio + InletTemp + ExhTemp + Airflow +
    Power

            Df Sum of Sq     RSS    AIC
- CPRatio    1     17118 4260877 391.58
- Shafts     1     18439 4262197 391.59
- Power      1    140770 4384528 392.49
- Airflow    1    222558 4466317 393.08
<none>                   4243758 393.45
- ExhTemp    1   1699823 5943582 402.23
- RPM        1   1711662 5955421 402.29
- InletTemp  1   4945835 9189593 416.17

Step:  AIC=391.58
HeatRate ~ Shafts + RPM + InletTemp + ExhTemp + Airflow + Power

            Df Sum of Sq      RSS    AIC
- Shafts     1     40971  4301848 389.88
- Power      1    131656  4392533 390.55
- Airflow    1    214261  4475138 391.15
<none>                    4260877 391.58
- RPM        1   1740116  6000993 400.53
- ExhTemp    1   3044486  7305363 406.83
- InletTemp  1   9873864 14134740 427.95

Step:  AIC=389.88
HeatRate ~ RPM + InletTemp + ExhTemp + Airflow + Power

            Df Sum of Sq      RSS    AIC
- Power      1    132612  4434460 388.85
- Airflow    1    210425  4512273 389.41
<none>                    4301848 389.88
- RPM        1   1960427  6262275 399.90
- ExhTemp    1   6493465 10795312 417.32
- InletTemp  1  14710137 19011985 435.44

Step:  AIC=388.85
HeatRate ~ RPM + InletTemp + ExhTemp + Airflow

            Df Sum of Sq      RSS    AIC
- Airflow    1    203236  4637696 388.29
<none>                    4434460 388.85
- RPM        1   4010400  8444860 407.47
- ExhTemp    1   6371857 10806317 415.36
- InletTemp  1  21857728 26292188 443.81

Step:  AIC=388.29
```

```
HeatRate ~ RPM + InletTemp + ExhTemp

            Df Sum of Sq      RSS    AIC
<none>                    4637696 388.29
- ExhTemp    1   6454449 11092145 414.19
- RPM        1   6583628 11221324 414.56
- InletTemp  1  22722741 27360437 443.08

Call:
lm(formula = HeatRate ~ RPM + InletTemp + ExhTemp)

Residuals:
    Min      1Q  Median      3Q     Max
-731.47 -272.46  -66.24  267.65  802.80

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.457e+04  8.892e+02  16.389 6.99e-16 ***
RPM          8.068e-02  1.280e-02   6.305 8.09e-07 ***
InletTemp   -1.028e+01  8.778e-01 -11.713 2.63e-12 ***
ExhTemp      1.490e+01  2.387e+00   6.242 9.55e-07 ***
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 407 on 28 degrees of freedom
Multiple R-squared:  0.9463,       Adjusted R-squared:  0.9405
F-statistic: 164.4 on 3 and 28 DF,  p-value: < 2.2e-16
```

You really should describe how you removed variables or at least which ones were removed to get full credit. (There isn't really a perfect answer here!)

18. **Question**
Compare this reduced model with the complete multiple regression model that included all predictors, using a single hypothesis test.

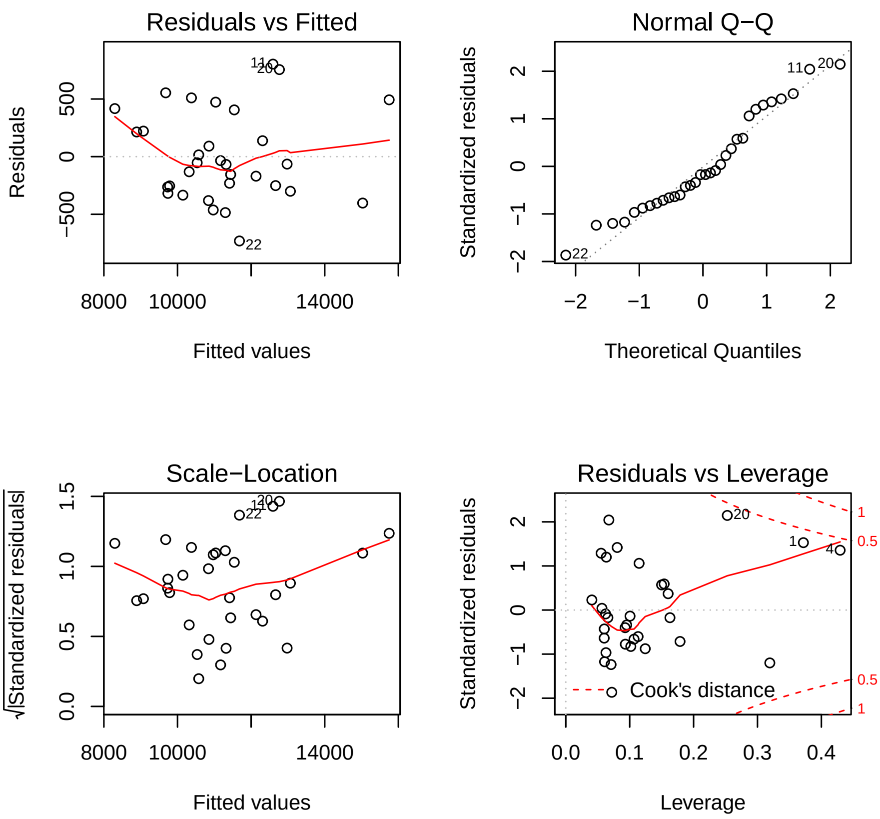
Solution
Using the anova() command with this last model:

```
Analysis of Variance Table

Model 1: HeatRate ~ Shafts + RPM + CPRatio + InletTemp + ExhTemp + Airflow +
    Power
Model 2: HeatRate ~ RPM + InletTemp + ExhTemp
  Res.Df     RSS Df Sum of Sq     F Pr(>F)
1     24 4243758
2     28 4637696 -4   -393937 0.557  0.696
```

The added complexity of the full multiple regression model is not a useful addition as it does not improve the fit significantly.

19. **Question**
Generate the residual analysis for your reduced model. Is there anything to be concerned about?

Solution



Make sure you pass comments about the constant variance and the normality of residuals at the very least. Comments about the leverages and Cook's distances can be made also.

20. **Question**
Obtain the Cook's distances and variance inflation factors for this model. Is there anything to worry about?

Solution
The variance inflation factors for my model are:

```
      RPM InletTemp   ExhTemp
 1.741145  2.976502  2.142442
```

The maximum variance inflation factor was 2.977, so no variables warrant removal on these grounds.

(A VIF greater than 10, is a clear indication that the model is flawed, but a VIF greater than 5 is an indication that some change to the model should be investigated.)

A summary of the Cook's distances is:

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.0000231 0.0038290 0.0158100 0.0550000 0.0309700 0.3897000
```

There were 0 Cook's distances greater than one.

21. **Question**
For any single predictor, demonstrate your knowledge about how to build a polynomial regression model to explain heat rate. That is, a model of the form $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 \ldots$ etc.

Solution
If you use the poly() command, remember to force the values generated by the command to be raw not centered ones.

```
Analysis of Variance Table

Model 1: HeatRate ~ InletTemp
Model 2: HeatRate ~ poly(InletTemp, 2, raw = T)
Model 3: HeatRate ~ poly(InletTemp, 3, raw = T)
Model 4: HeatRate ~ poly(InletTemp, 4, raw = T)
  Res.Df      RSS Df Sum of Sq      F  Pr(>F)
1     30 28211849
2     29 23562506  1   4649344 5.5410 0.02611 *
3     28 22665235  1    897270 1.0693 0.31027
4     27 22655257  1      9979 0.0119 0.91397
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1
```

We should use the last model that has any significance attached to it in the table here. If the last one had been significant, then a higher degree polynomial would be required. In my case, only the second order model for the HeatRate as estimated by the InletTemp was justified but your comments must link to your results.

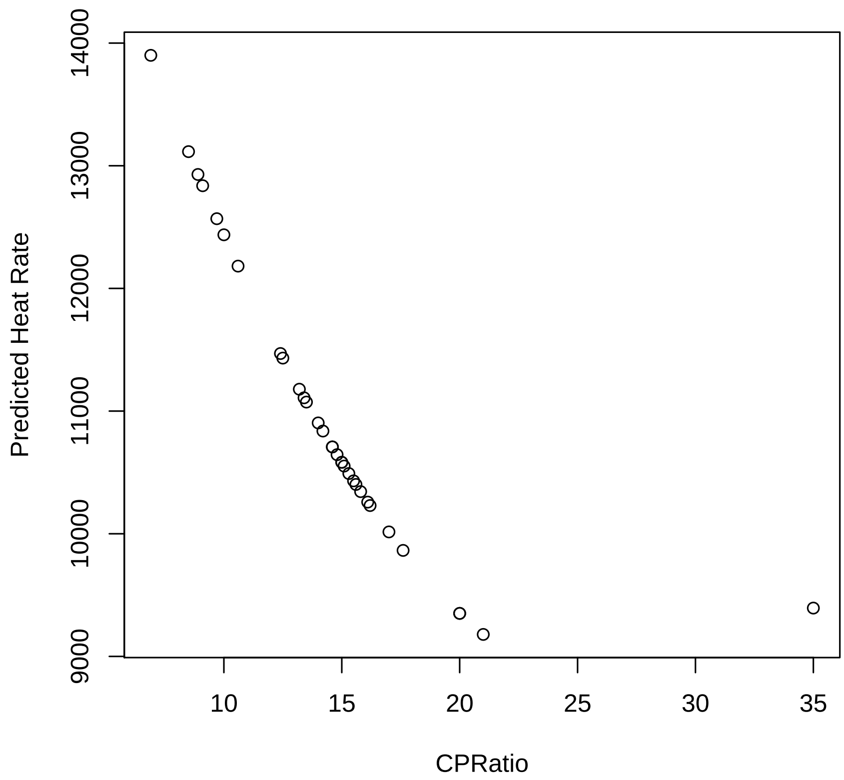
22. **Question**
Consider a model for heat rate of a gas turbine that is a function of cycle speed and cycle pressure ratio. Fit a second-order model using just these two variables. Construct a graph that compares the cycle pressure ratio with the heat rate that is predicted under this model for a cycle speed of 4900 rpm.

Solution
After creating new variables for the quadratic and interaction terms, I get the following model:

```
Call:
lm(formula = HeatRate ~ RPM + CPRatio + RPM2 + CPRatio2 + CPRatioRPM)

Residuals:
    Min      1Q  Median      3Q     Max
-946.48 -232.73  -52.36  335.34  893.16

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.895e+04  1.407e+03  13.462 3.15e-13 ***
RPM         -2.102e-01  1.584e-01  -1.327   0.1960
CPRatio     -7.656e+02  1.198e+02  -6.393 9.00e-07 ***
RPM2         4.203e-06  2.626e-06   1.600   0.1216
CPRatio2     1.246e+01  2.218e+00   5.617 6.65e-06 ***
CPRatioRPM   1.697e-02  8.261e-03   2.055   0.0501 .
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1
```

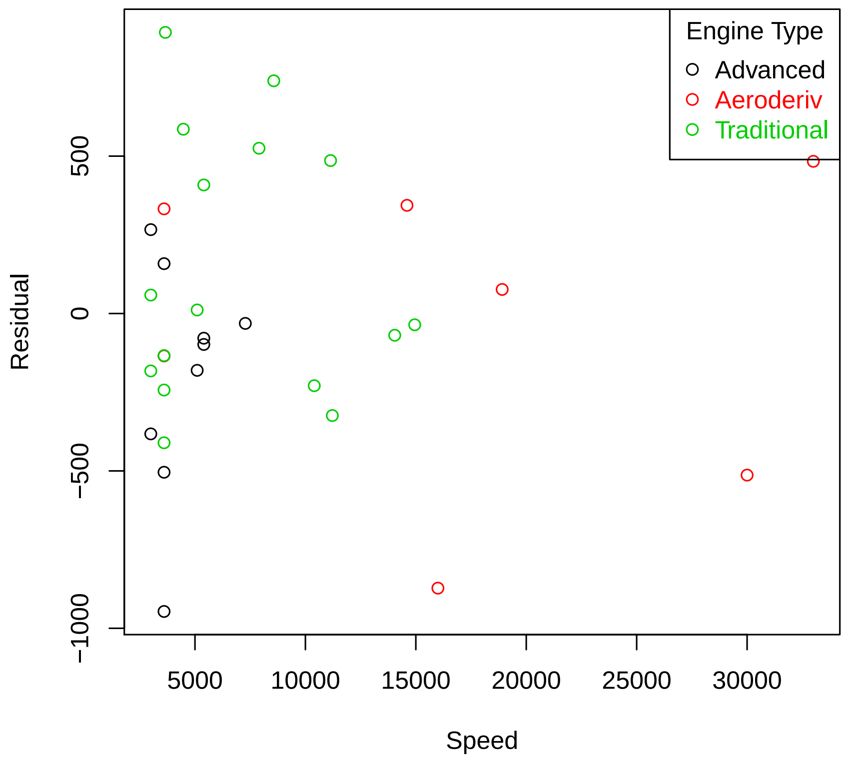
```
Residual standard error: 470.9 on 26 degrees of freedom
Multiple R-squared:  0.9332,       Adjusted R-squared:  0.9204
F-statistic: 72.66 on 5 and 26 DF,  p-value: 1.934e-14
```

I then created a new data frame of values for the five variables in the model and make a prediction for all CPRatios, holding the speed constant.



23. **Question**
Up to this point we have ignored the engine type variable in your data. Create a graph that plots the residuals from the last model against the Speed, which indicates the differences between the three engine types.

Solution
I've chosen to use colour to identify the different Engine Types. You could use a different symbol (which would help the colour-blind) if you preferred.



Do you believe the engine type is important? Your answer must link to your graph.

24. **Question**
Fit the model that adds the engine type to the second-order model using only cycle pressure ratio and cycle speed. Does allowing for differences in the mean response for each engine type improve the ability of the model to predict heat rate?

Solution
You only needed to add in the additive term for the engine type in this instance.

```
Call:
lm(formula = HeatRate ~ Engine + RPM + CPRatio + RPM2 + CPRatio2 +
    CPRatioRPM)

Residuals:
    Min      1Q  Median      3Q     Max
-729.72 -271.31   68.11  272.37  702.78

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.725e+04  1.569e+03  10.994 7.48e-11 ***
EngineAeroderiv    2.270e+02  3.424e+02   0.663 0.513723
EngineTraditional  4.827e+02  2.277e+02   2.120 0.044518 *
RPM               -1.403e-01  1.552e-01  -0.904 0.374830
CPRatio           -6.354e+02  1.316e+02  -4.826 6.46e-05 ***
```

```
RPM2               3.765e-06  2.546e-06   1.479 0.152210
CPRatio2           1.023e+01  2.460e+00   4.158 0.000353 ***
CPRatioRPM         1.220e-02  8.328e-03   1.465 0.155829
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 448.4 on 24 degrees of freedom
Multiple R-squared:  0.9441,       Adjusted R-squared:  0.9278
F-statistic:  57.9 on 7 and 24 DF,  p-value: 1.716e-13

Analysis of Variance Table

Model 1: HeatRate ~ Engine + RPM + CPRatio + RPM2 + CPRatio2 + CPRatioRPM
Model 2: HeatRate ~ RPM + CPRatio + RPM2 + CPRatio2 + CPRatioRPM
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     24 4825094
2     26 5764747 -2   -939653 2.3369 0.1182
```

There is insufficient evidence to justify the addition of engine type to our second order model.

25. **Question**
The second-order model using cycle pressure ratio and cycle speed is called a response surface model. Do you believe there is a different response surface for each engine type?

Solution
To get a different response surface, you needed to allow for the interaction between the engine type and the other variables.

```
Call:
lm(formula = HeatRate ~ Engine * (RPM + CPRatio + RPM2 + CPRatio2 +
    CPRatioRPM))

Residuals:
    Min      1Q  Median      3Q     Max
-378.53  -62.78  -23.59   20.36  758.24

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -1.516e+04  3.123e+04  -0.485   0.6349
EngineAeroderiv              -1.726e+05  7.549e+04  -2.287   0.0383 *
EngineTraditional             3.880e+04  3.140e+04   1.236   0.2369
RPM                           4.394e+00  4.365e+00   1.007   0.3312
CPRatio                       1.957e+03  2.791e+03   0.701   0.4948
RPM2                         -1.210e-04  1.286e-04  -0.941   0.3626
CPRatio2                     -3.887e+01  6.064e+01  -0.641   0.5319
CPRatioRPM                   -1.865e-01  2.081e-01  -0.896   0.3853
EngineAeroderiv:RPM           8.667e+00  6.268e+00   1.383   0.1884
EngineTraditional:RPM        -4.351e+00  4.373e+00  -0.995   0.3367
EngineAeroderiv:CPRatio       1.185e+04  5.565e+03   2.129   0.0514 .
EngineTraditional:CPRatio    -3.607e+03  2.826e+03  -1.277   0.2225
EngineAeroderiv:RPM2         -6.150e-05  1.436e-04  -0.428   0.6749
EngineTraditional:RPM2        1.184e-04  1.289e-04   0.918   0.3742
EngineAeroderiv:CPRatio2     -1.806e+02  9.733e+01  -1.855   0.0848 .
```

```
EngineTraditional:CPRatio2    8.987e+01  6.288e+01   1.429   0.1749
EngineAeroderiv:CPRatioRPM   -3.101e-01  2.710e-01  -1.144   0.2717
EngineTraditional:CPRatioRPM  1.905e-01  2.084e-01   0.914   0.3762
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 314.7 on 14 degrees of freedom
Multiple R-squared:  0.9839,       Adjusted R-squared:  0.9644
F-statistic: 50.43 on 17 and 14 DF,  p-value: 1.177e-09
```

The anova() command is useful for determining the value of the additional variables as a block.

```
Analysis of Variance Table

Model 1: HeatRate ~ Engine + RPM + CPRatio + RPM2 + CPRatio2 + CPRatioRPM
Model 2: HeatRate ~ Engine * (RPM + CPRatio + RPM2 + CPRatio2 + CPRatioRPM)
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     24 4825094
2     14 1386824 10   3438270 3.4709 0.01697 *
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1
```

The interaction is significant so there is evidence that a different response surface for each engine type is warranted.