Nour Shinnawi

STA 2260 Project

Data Set:

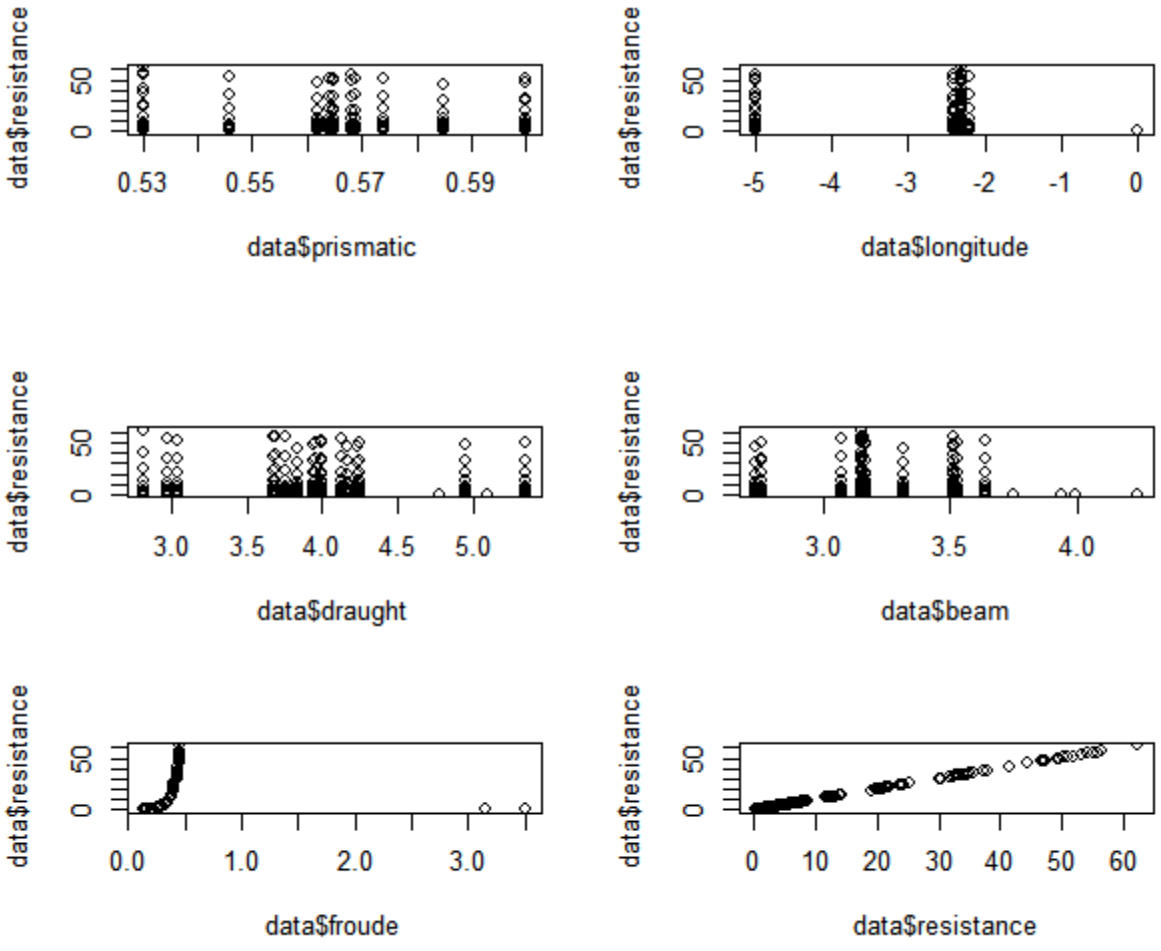
The data set used is Yacht Hydrodynamics. The response variable is the resistance and the predictors are all the other variables. The other variables are:

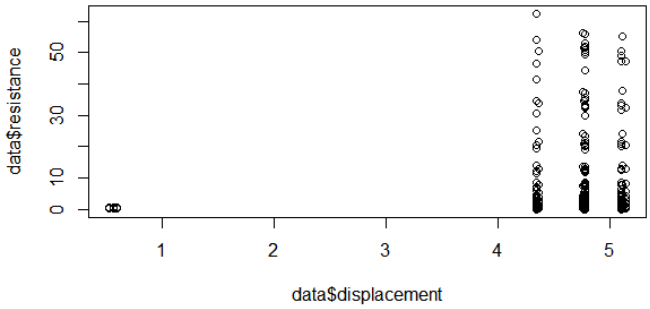
- Longitudinal position of the center of buoyancy, labeled as longitude
- Prismatic coefficient, labeled as prismatic
- Length-displacement ratio, labeled as displacement
- Beam-draught ratio, labeled as draught
- Length-beam ratio, labeled as beam
- Froude number, labeled as froude

First few rows using `head(data)`:

```
  longitude prismatic displacement draught beam froude resistance
1      -2.3     0.568          4.78    3.99 3.17  0.125       0.11
2      -2.3     0.568          4.78    3.99 3.17  0.150       0.27
3      -2.3     0.568          4.78    3.99 3.17  0.175       0.47
4      -2.3     0.568          4.78    3.99 3.17  0.200       0.78
5      -2.3     0.568          4.78    3.99 3.17  0.225       1.18
6      -2.3     0.568          4.78    3.99 3.17  0.250       1.82
```

Scatterplot Analysis and Discussion:





The graphs for prismatic, longitude, draught, displacement, and beam did not provide good scatter plots. Froude created a J shaped curve and resistance created a linear graph.

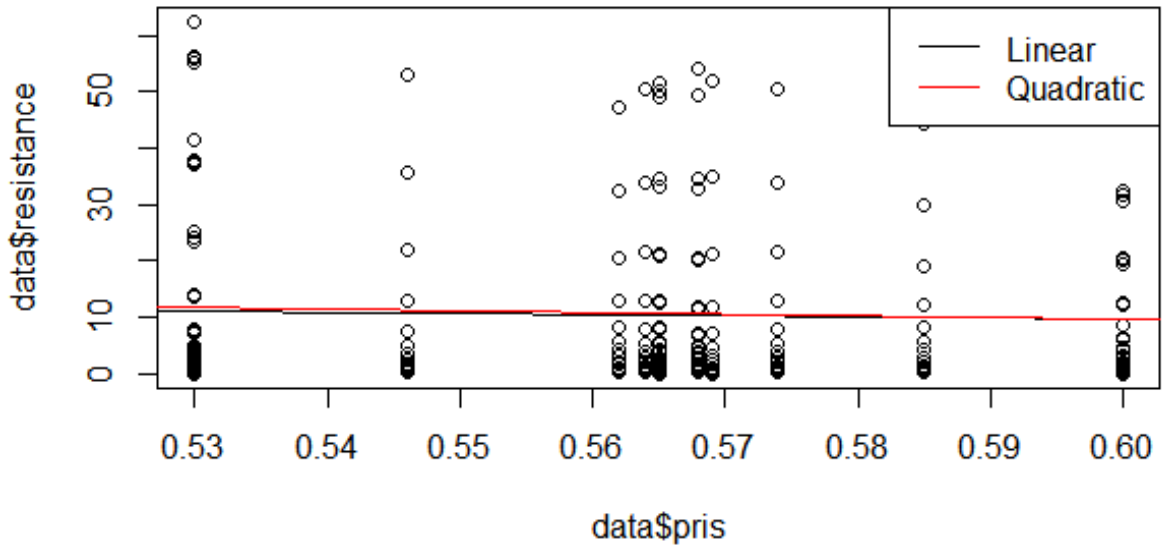
Model #1:

```
fit.disp <- lm(resistance~displacement,data=data)
fit.disp2 <- lm(resistance~displacement+I(displacement^2),data=data)
AIC(fit.disp,fit.disp2)
          df      AIC
fit.disp   3 2483.452
fit.disp2  4 2484.777

fit.froude <- lm(resistance~froude,data=data)
fit.froude2 <- lm(resistance~froude+I(froude^2),data=data)
AIC(fit.froude,fit.froude2)
            df      AIC
fit.froude   3 2492.869
fit.froude2  4 2343.461

fit.pris <- lm(resistance~prismatic,data=data)
fit.pris2 <- lm(resistance~prismatic+I(prismatic^2),data=data)
AIC(fit.pris,fit.pris2)
          df     AIC
fit.pris   3 2082.80
fit.pris2  4 2084.77

fit.beam <- lm(resistance~beam,data=data)
fit.beam2 <- lm(resistance~beam+I(beam^2),data=data)
AIC(fit.beam,fit.beam2)
          df     AIC
fit.beam   3 2492.88
fit.beam2  4 2490.21
```

The lowest AIC is the linear model in prismatic.

```
summary(fit.pris):
Call:
lm(formula = resistance ~ prismatic, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-10.897  -9.440  -7.217   2.590  51.473

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    21.11      23.23   0.909    0.364
prismatic     -19.18      41.16  -0.466    0.642

Residual standard error: 14.96 on 250 degrees of freedom
Multiple R-squared:  0.0008678,       Adjusted R-squared:  -0.003129
F-statistic: 0.2171 on 1 and 250 DF,  p-value: 0.6416
```

```
summary(fit.pris2):
Call:
lm(formula = resistance ~ prismatic + I(prismatic^2), data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-11.066  -9.453  -7.171   2.591  51.304

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      111.9      531.0   0.211    0.833
prismatic       -342.0     1887.0  -0.181    0.856
I(prismatic^2)   286.5     1674.4   0.171    0.864

Residual standard error: 14.99 on 249 degrees of freedom
Multiple R-squared:  0.0009853,       Adjusted R-squared:  -0.007039
F-statistic: 0.1228 on 2 and 249 DF,  p-value: 0.8845
```



The fitted model is: resistance = 21.11 – 19.48prismatic.

R^2 is 0.0008678 and it has an AIC of 2082.821.

<u>Model #2:</u>

Trying other variables:

```
Call:
lm(formula = resistance ~ prismatic + I(prismatic^2) + displacement +
    I(displacement^2) + froude + I(froude^2) + beam + I(beam^2) +
    longitude + draught, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-7.0117 -4.0359  0.3109  3.0490 17.5846

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      170.48581  189.52530   0.900    0.369
prismatic       -362.75059  578.32363  -0.627    0.531
```

```
I(prismatic^2)     299.22365  505.81450   0.592    0.555
displacement       -12.25749   42.47059  -0.289    0.773
I(displacement^2)    1.02446    4.11641   0.249    0.804
froude            -375.84765   16.97197 -22.145   <2e-16 ***
I(froude^2)        861.55678   29.16573  29.540   <2e-16 ***
beam                 3.12880   29.98396   0.104    0.917
I(beam^2)           -0.13853    4.06138  -0.034    0.973
longitude           -0.09003    0.27498  -0.327    0.744
draught              0.86643    3.40874   0.254    0.800
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.173 on 241 degrees of freedom
Multiple R-squared:  0.9251,   Adjusted R-squared:  0.922
F-statistic: 297.6 on 10 and 241 DF,  p-value: < 2.2e-16

AIC(m2.1)
[1] 1447.972
```

After removing some variables from the formula:

```
Call:
lm(formula = resistance ~ prismatic + I(prismatic^2) + froude +
    I(froude^2) + beam + I(beam^2) + longitude, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.7842 -4.0620  0.4997  2.9510 17.9008

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     142.04312  156.06983   0.910    0.364
prismatic      -335.92274  526.27488  -0.638    0.524
I(prismatic^2)  280.55461  466.73539   0.601    0.548
froude         -375.84765   16.87380 -22.274   <2e-16 ***
I(froude^2)     861.55678   28.99702  29.712   <2e-16 ***
beam             -2.18894   23.64554  -0.093    0.926
I(beam^2)         0.31902    3.71157   0.086    0.932
longitude        -0.04488    0.24756  -0.181    0.856
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.149 on 244 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.925,    Adjusted R-squared:  0.9229
F-statistic: 430.1 on 7 and 244 DF,  p-value: < 2.2e-16

> AIC(m2.1)
[1] 1442.166
```

The AIC became even lower. R^2 is 0.925 and the AIC is 1442.166

Model Comparison:

Model #1 had an R^2 of 0.0008678 and AIC of 2082.821. Model #2 had an R^2 of 0.925 and an AIC of 1442.166. Model #2 had more variability for the yacht data and it has greater predictive power based on AIC.