Lesson 7 Lecture Example

Rick Davila

5/11/2020

Lesson 7 - Install packages

Install necessary packages using library()

Perform data housekeeping - upload, name columns, display to make sure it reads properly, etc.

```
knitr::opts_chunk$set(echo = TRUE)

library(e1071)
library(xtable)
library("xlsx") # Needed to read data
```

```
## Warning: package 'xlsx' was built under R version 4.0.3
```

```
library(psych) # For geometric mean in Example 5.3
```

```
## Warning: package 'psych' was built under R version 4.0.3
```

```
rm(list = ls())
```

Read data file (data-ex-5-1.xlsx)

```
exL <- read.xlsx("data-table-B09.xlsx",
                   sheetIndex = 1,
                   colIndex = c(1,2,3,4,5),
                   as.data.frame = TRUE,
                   header = TRUE)
```

Assign labels to data columns using names() and attach() commands

```
names(exL) <- c("fluid_vel", "viscosity","mesh_open","fluid_gas_vel","pressure_drop")
attach(exL)
```

Output data to make sure it reads properly

```
out <- as.data.frame(c(exL))
colnames(out) <- c("fluid_vel", "viscosity","mesh_open","fluid_gas_vel","pressure_drop")
tab <- (xtable(out, digits=c(0,2,1,2,3,1)))
print(tab, type="html")
```

	fluid_vel	viscosity	mesh_open	fluid_gas_vel	pressure_drop

1	2.14	10.0	0.34	1.000	28.9
2	4.14	10.0	0.34	1.000	31.0
3	8.15	10.0	0.34	1.000	26.4
4	2.14	10.0	0.34	0.246	27.2
5	4.14	10.0	0.34	0.379	26.1
6	8.15	10.0	0.34	0.474	23.2
7	2.14	10.0	0.34	0.141	19.7
8	4.14	10.0	0.34	0.234	22.1
9	8.15	10.0	0.34	0.311	22.8
10	2.14	10.0	0.34	0.076	29.2
11	4.14	10.0	0.34	0.132	23.6
12	8.15	10.0	0.34	0.184	23.6
13	2.14	2.6	0.34	0.679	24.2
14	4.14	2.6	0.34	0.804	22.1
15	8.15	2.6	0.34	0.890	20.9
16	2.14	2.6	0.34	0.514	17.6
17	4.14	2.6	0.34	0.672	15.7
18	8.15	2.6	0.34	0.801	15.8
19	2.14	2.6	0.34	0.346	14.0
20	4.14	2.6	0.34	0.506	17.1
21	8.15	2.6	0.34	0.669	18.3
22	2.14	2.6	0.34	1.000	33.8
23	4.14	2.6	0.34	1.000	31.7
24	8.15	2.6	0.34	1.000	28.1
25	5.60	1.2	0.34	0.848	18.1
26	5.60	1.2	0.34	0.737	16.5
27	5.60	1.2	0.34	0.651	15.4
28	5.60	1.2	0.34	0.554	15.0
29	4.30	2.6	0.34	0.748	19.1
30	4.30	2.6	0.34	0.682	16.2
31	4.30	2.6	0.34	0.524	16.3
32	4.30	2.6	0.34	0.472	15.8
33	4.30	2.6	0.34	0.398	15.4
34	5.60	10.1	0.25	0.789	19.2
35	5.60	10.1	0.25	0.677	8.4
36	5.60	10.1	0.25	0.590	15.0
37	5.60	10.1	0.25	0.523	12.0
38	5.60	10.1	0.34	0.789	21.9
39	5.60	10.1	0.34	0.677	21.3
40	5.60	10.1	0.34	0.590	21.6
41	5.60	10.1	0.34	0.523	19.8
42	4.30	10.1	0.34	0.741	21.6
43	4.30	10.1	0.34	0.617	17.3
44	4.30	10.1	0.34	0.524	20.0
45	4.30	10.1	0.34	0.457	18.6
46	2.40	10.1	0.34	0.615	22.1

47	2.40	10.1	0.34	0.473	14.7
48	2.40	10.1	0.34	0.381	15.8
49	2.40	10.1	0.34	0.320	13.2
50	5.60	10.1	0.55	0.789	30.8
51	5.60	10.1	0.55	0.677	27.5
52	5.60	10.1	0.55	0.590	25.2
53	5.60	10.1	0.55	0.523	22.8
54	2.14	112.0	0.34	0.680	41.7
55	4.14	112.0	0.34	0.803	33.7
56	8.15	112.0	0.34	0.889	29.7
57	2.14	112.0	0.34	0.514	41.8
58	4.14	112.0	0.34	0.672	37.1
59	8.15	112.0	0.34	0.801	40.1
60	2.14	112.0	0.34	0.306	42.7
61	4.14	112.0	0.34	0.506	48.6
62	8.15	112.0	0.34	0.668	42.4

```
# Output data structure and dimensions
str(exL)
```

'data.frame': 62 obs. of 5 variables: $ fluid_vel : num 2.14 4.14 8.15 2.14 4.14 8.15 2.14 4.14 8.15 2.14 … $ viscosity : num 10 10 10 10 10 10 10 10 10 10 … $ mesh_open : num 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 … $ fluid_gas_vel: num 1 1 1 0.246 0.379 0.474 0.141 0.234 0.311 0.076 … $ pressure_drop: num 28.9 31 26.4 27.2 26.1 23.2 19.7 22.1 22.8 29.2 …

```
dim(exL)
```

[1] 62 5

create multiple least squares model

```
model <- lm(pressure_drop ~ fluid_vel + viscosity + mesh_open + fluid_gas_vel)

summary(model)
```

Call: lm(formula = pressure_drop ~ fluid_vel + viscosity + mesh_open + fluid_gas_vel)

Residuals: Min 1Q Median 3Q Max -9.9958 -3.3092 -0.2419 3.3924 10.5668

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.89453 4.32508 1.363 0.17828
fluid_vel -0.47790 0.34002 -1.406 0.16530
viscosity 0.18271 0.01718 10.633 3.78e-15 ***mesh_open 35.40284 11.09960 3.190 0.00232*** *fluid_gas_vel 5.84391 2.90978 2.008 0.04935*
— Signif. codes: 0 '*' ***0.001*** '*' *0.01* '*'* 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.014 on 57 degrees of freedom Multiple R-squared: 0.6914, Adjusted R-squared: 0.6697 F-statistic: 31.92 on 4 and 57 DF, p-value: 5.818e-14

```
xtable(summary(model))
```

	Estimate <dbl>	Std. Error <dbl>	t value <dbl>	Pr(>|t|) <dbl>
(Intercept)	5.8945253	4.32507771	1.362872	1.782831e-01
fluid_vel	-0.4779013	0.34001900	-1.405514	1.652960e-01
viscosity	0.1827137	0.01718375	10.632936	3.779436e-15
mesh_open	35.4028387	11.09960045	3.189560	2.316482e-03
fluid_gas_vel	5.8439137	2.90977841	2.008371	4.935047e-02

5 rows

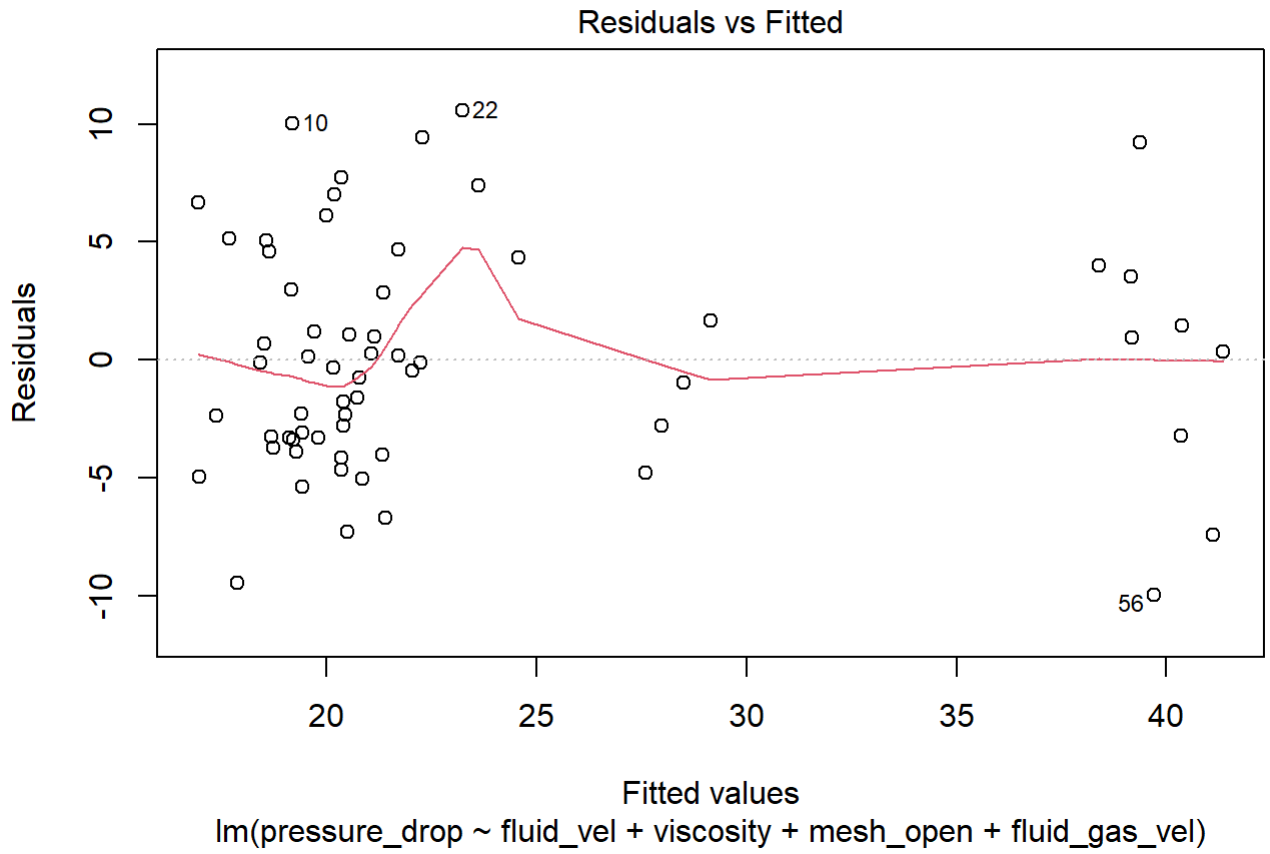
```
xtable(anova(model))
```

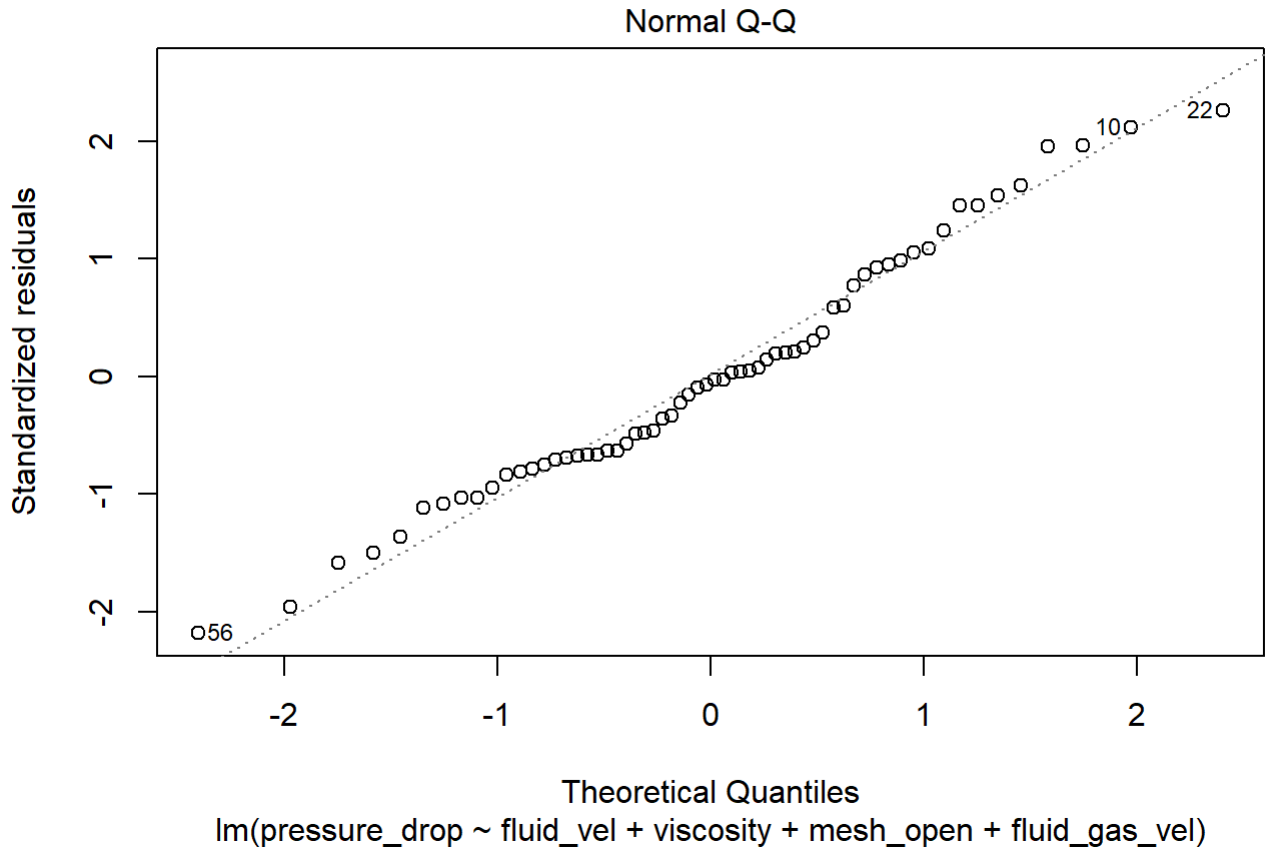
	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
fluid_vel	1	9.595968	9.595968	0.3817523	5.391278e-01
viscosity	1	2839.780203	2839.780203	112.9737594	3.834506e-15
mesh_open	1	258.951631	258.951631	10.3017618	2.183922e-03
fluid_gas_vel	1	101.389947	101.389947	4.0335528	4.935047e-02
Residuals	57	1432.788219	25.136635	*NA*	*NA*

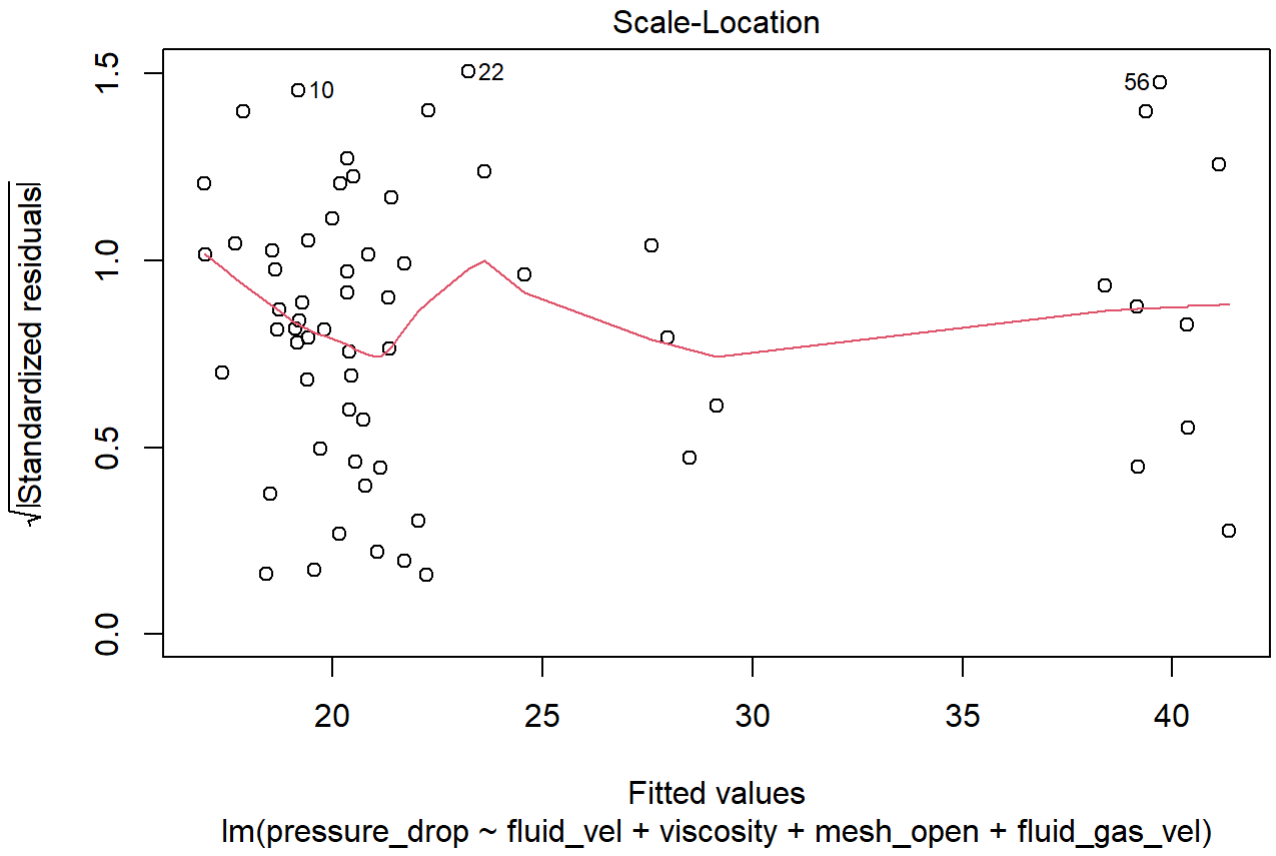
5 rows

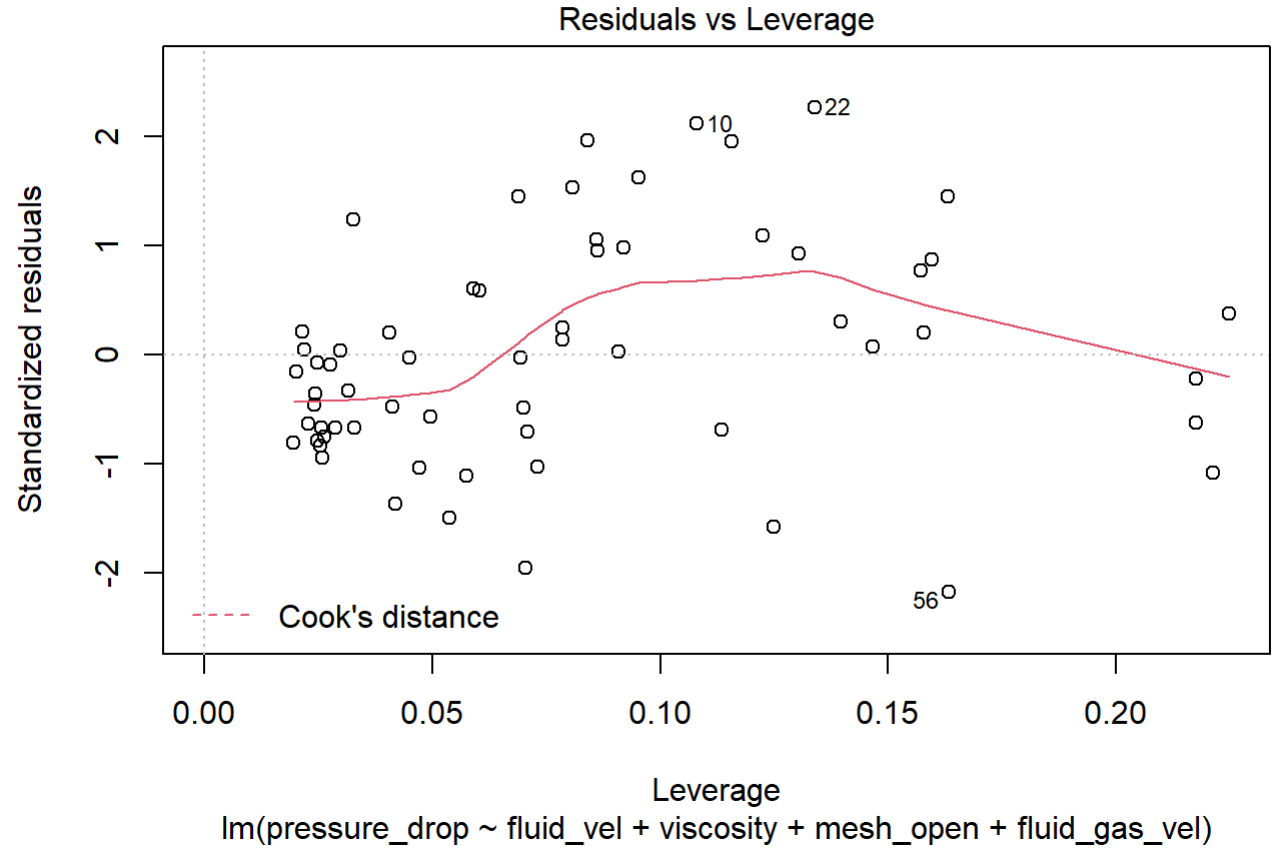
make a plot of the multiple least squares model

```
plot(model)
```

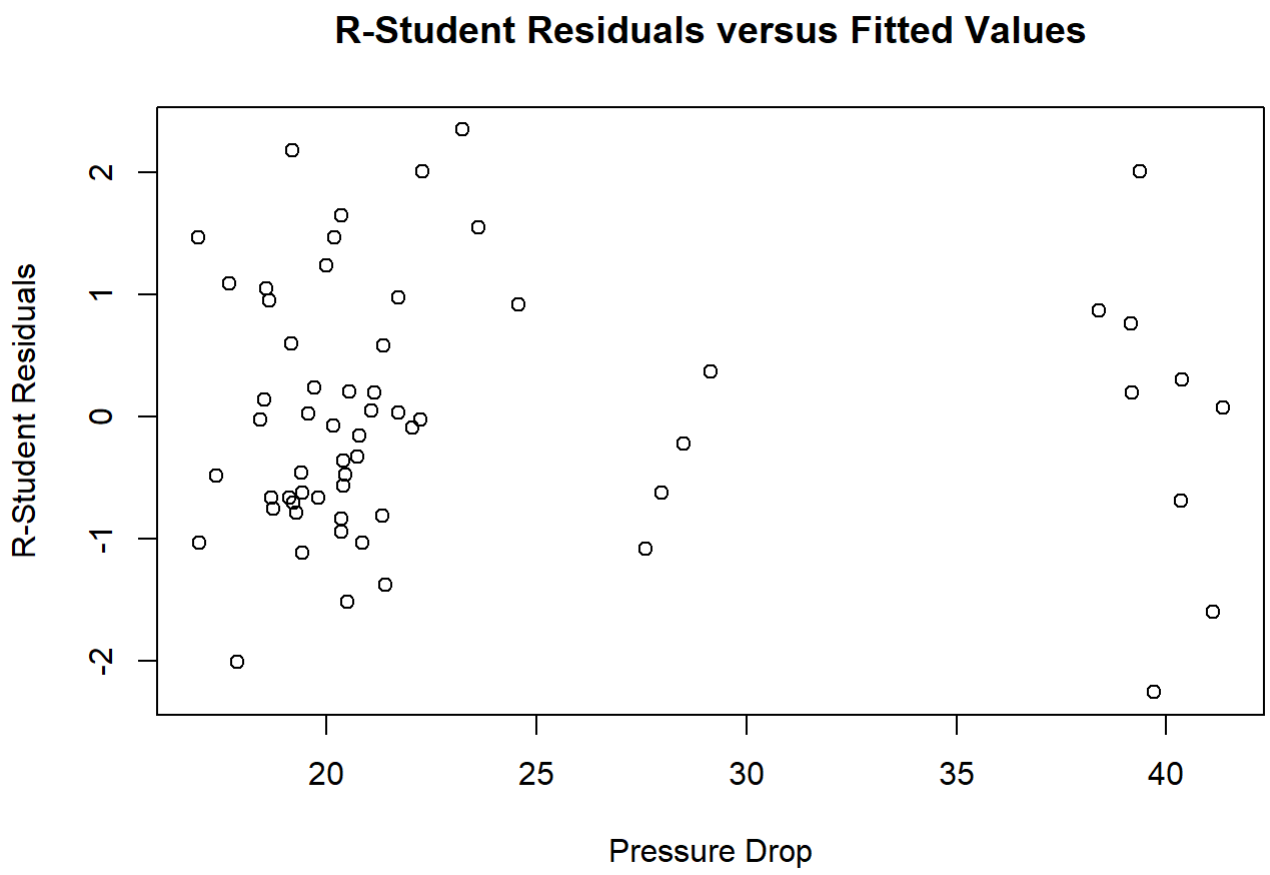

Residuals vs Fitted
Residuals
10
22
56
10
5
0
-5
-10
20
25
30
35
40
Fitted values
lm(pressure_drop ~ fluid_vel + viscosity + mesh_open + fluid_gas_vel)


Normal Q-Q
Standardized residuals
56
10
22
2
1
0
-1
-2
-2
-1
0
1
2
Theoretical Quantiles
lm(pressure_drop ~ fluid_vel + viscosity + mesh_open + fluid_gas_vel)

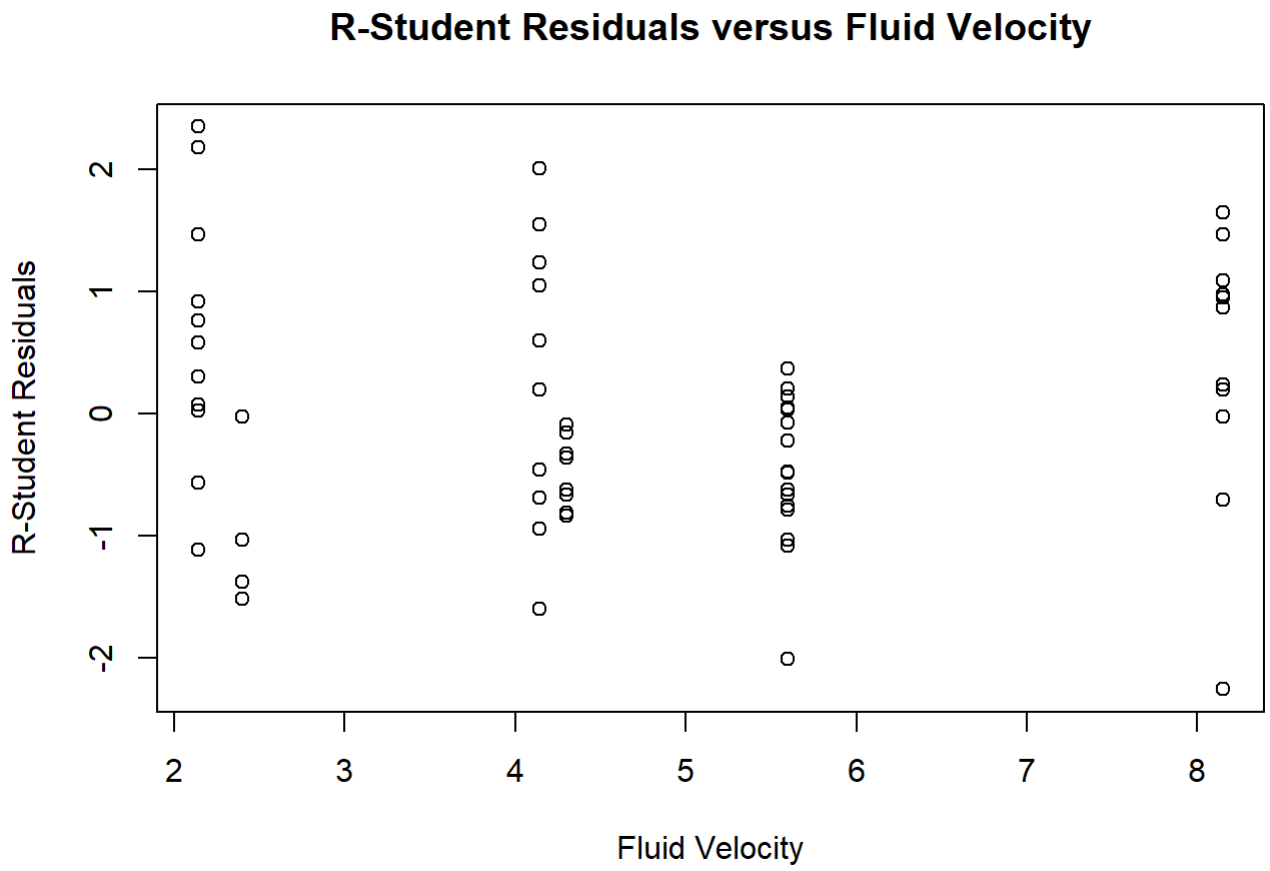

Scale-Location
√|Standardized residuals|
1.5
1.0
0.5
0.0
10
22
56
20
25
30
35
40
Fitted values
lm(pressure_drop ~ fluid_vel + viscosity + mesh_open + fluid_gas_vel)


Residuals vs Leverage
Standardized residuals
2
1
0
-1
-2
10
22
56
Cook's distance
0.00
0.05
0.10
0.15
0.20
Leverage
lm(pressure_drop ~ fluid_vel + viscosity + mesh_open + fluid_gas_vel)

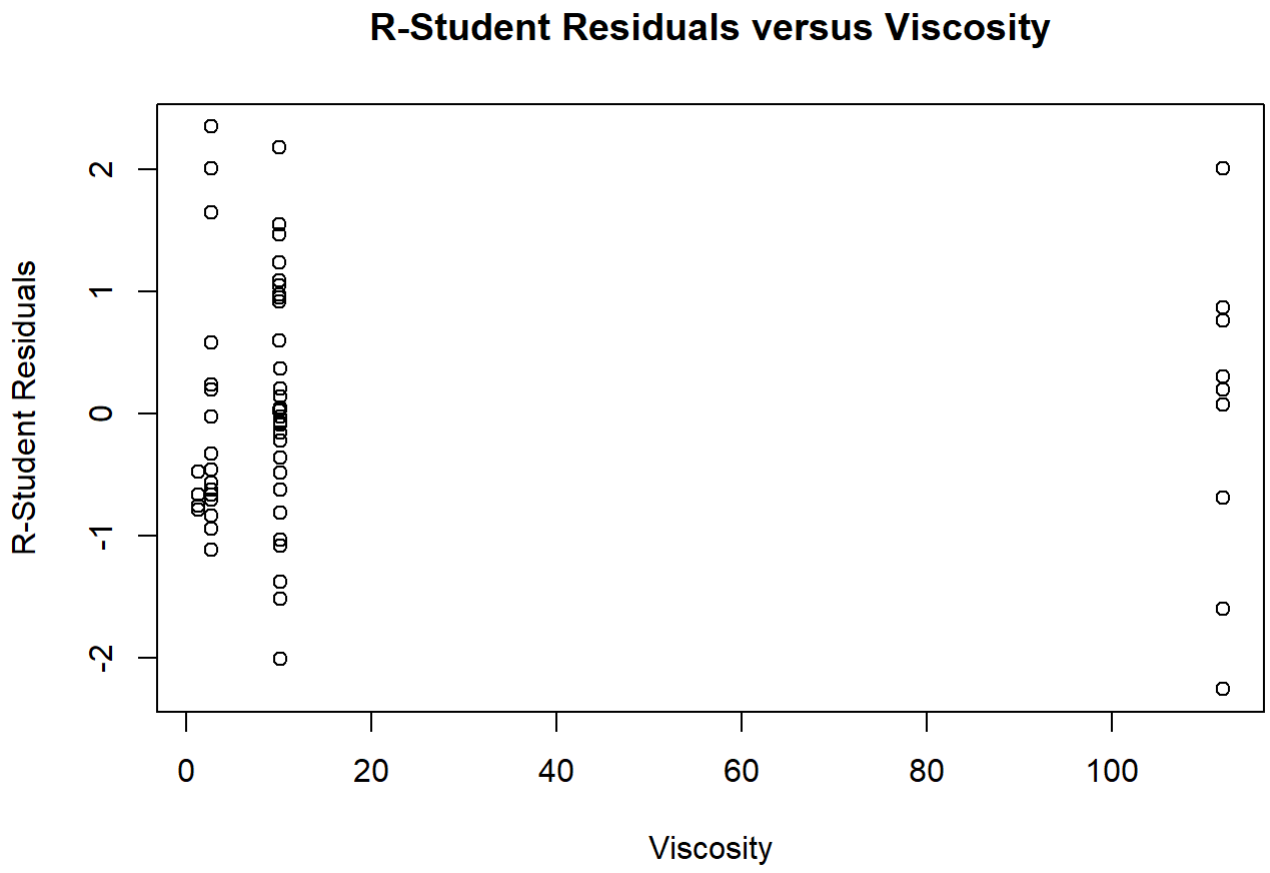
```
# Let's look at a plot of the R-student residuals versus the fitted values

R_Student_Residuals <- rstudent(model)
y_hat <- model$fitted.values
plot(y_hat, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Pressure Drop", main = "R
-Student Residuals versus Fitted Values")
```



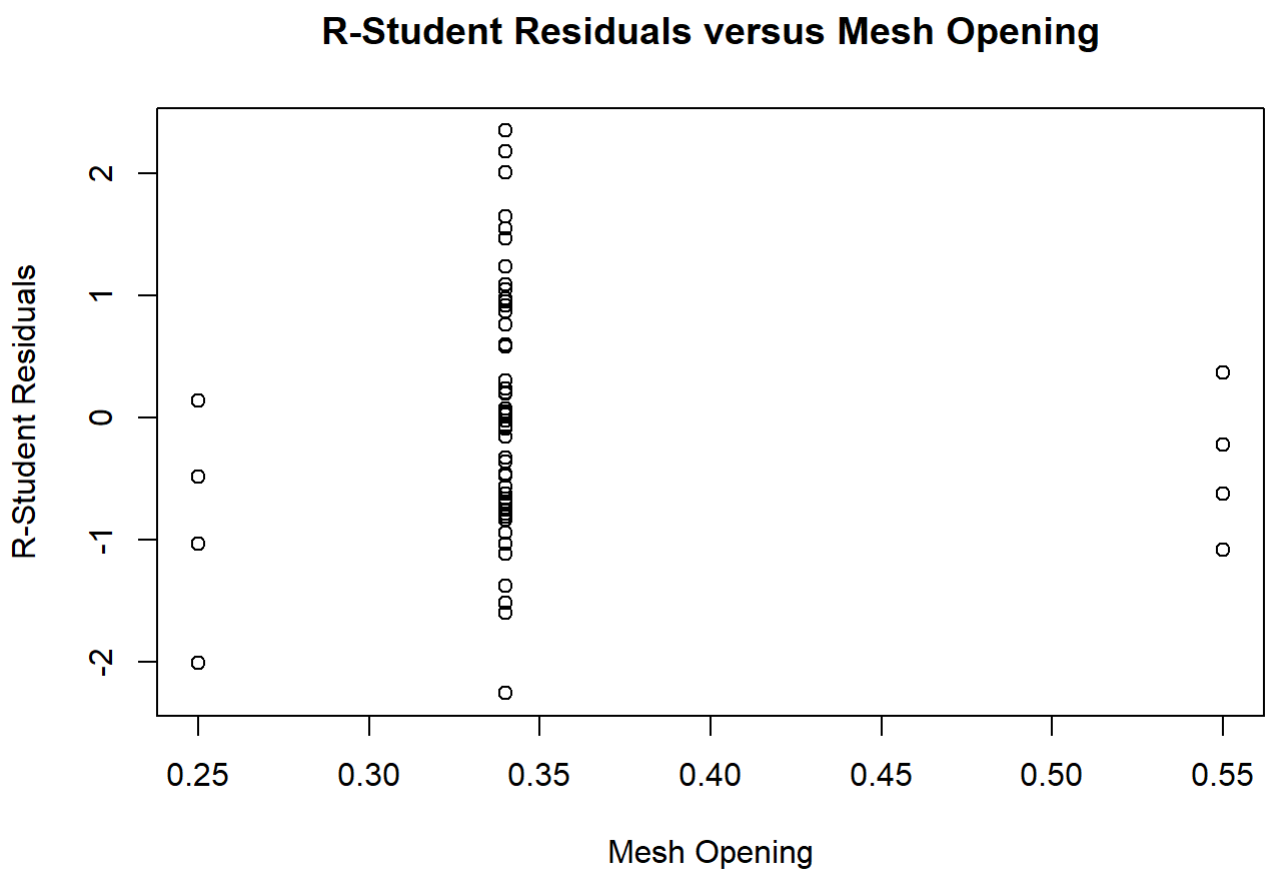

```
plot(fluid_vel, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Fluid Velocity", main
= "R-Student Residuals versus Fluid Velocity")
```



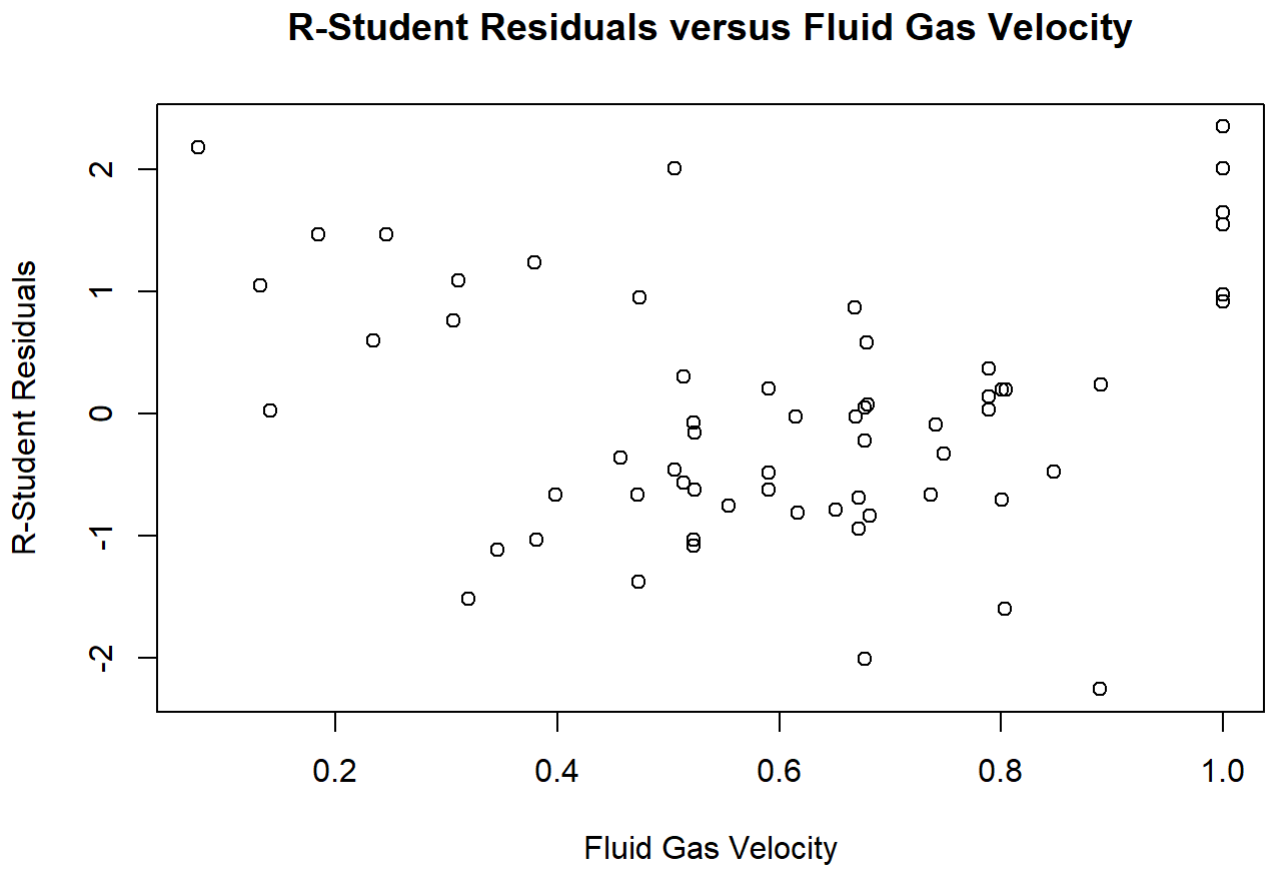

```
plot(viscosity, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Viscosity", main = "R
-Student Residuals versus Viscosity")
```




```
plot(mesh_open, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Mesh Opening", main =
"R-Student Residuals versus Mesh Opening")
```




```
plot(fluid_gas_vel, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Fluid Gas Velocit
y", main = "R-Student Residuals versus Fluid Gas Velocity")
```



Since there's a huge gap in the residuals versus viscosity data, we'll try a natural log on the regressor. Redo the exercise above, but, with the lm() using log of vicosity.

```
model <- lm(pressure_drop ~ fluid_vel + log(viscosity) + mesh_open + fluid_gas_vel)

summary(model)
```

Call: lm(formula = pressure_drop ~ fluid_vel + log(viscosity) + mesh_open + fluid_gas_vel)

Residuals: Min 1Q Median 3Q Max -13.4818 -3.5114 -0.2417 2.9716 12.8961

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.8207 4.9007 -0.167 0.86760
fluid_vel -0.5492 0.3739 -1.469 0.14739
log(viscosity) 5.0960 0.5574 9.143 9.03e-13 ***mesh_open 28.0374 12.1975 2.299 0.02521***
fluid_gas_vel 10.3162 3.2096 3.214 0.00215 — Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.514 on 57 degrees of freedom Multiple R-squared: 0.6267, Adjusted R-squared: 0.6005
F-statistic: 23.92 on 4 and 57 DF, p-value: 1.201e-11

```
xtable(summary(model))
```

	Estimate <dbl>	Std. Error <dbl>	t value <dbl>	Pr(>|t|) <dbl>
(Intercept)	-0.8206807	4.9007332	-0.1674608	8.676002e-01
fluid_vel	-0.5492321	0.3739394	-1.4687728	1.473920e-01
log(viscosity)	5.0960394	0.5573821	9.1428120	9.031359e-13
mesh_open	28.0374015	12.1974870	2.2986211	2.521404e-02
fluid_gas_vel	10.3162044	3.2095719	3.2141995	2.154800e-03

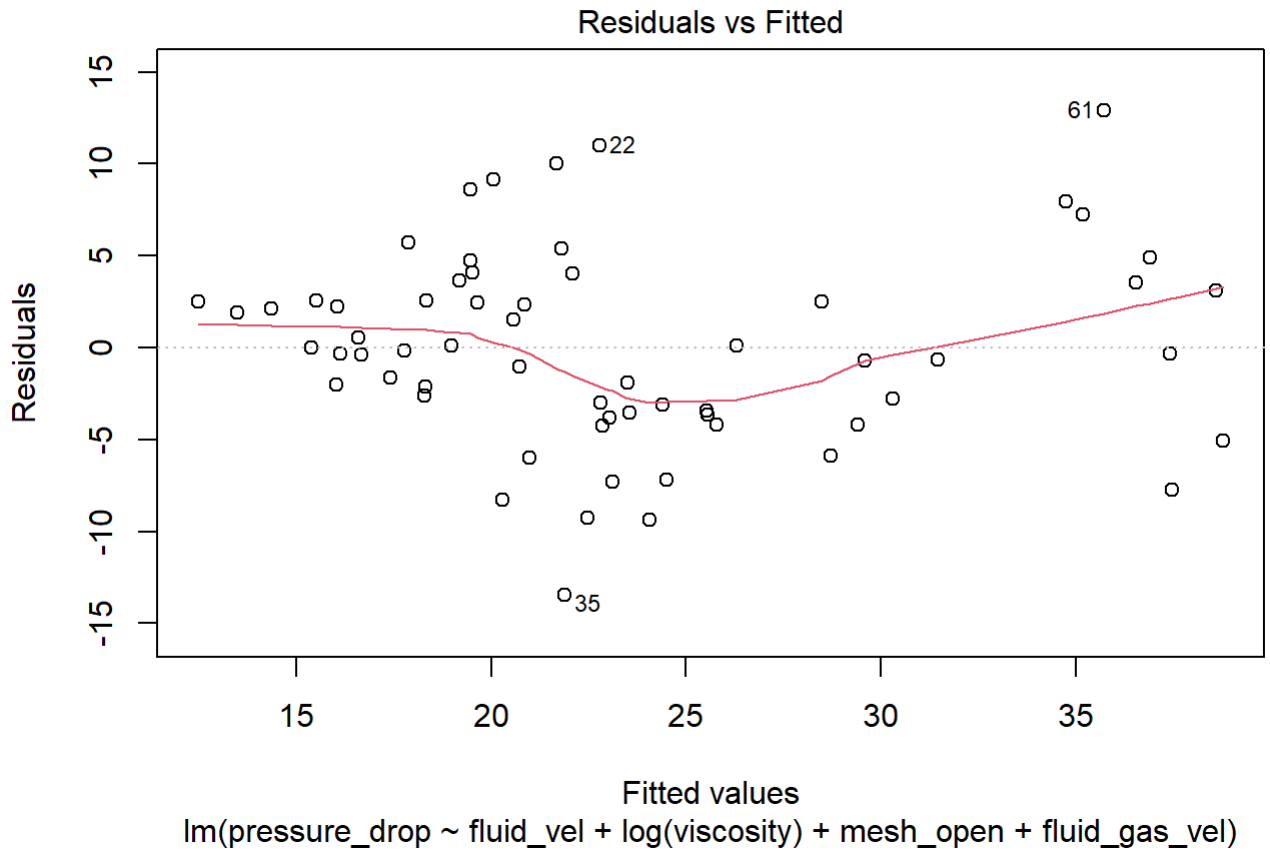
5 rows

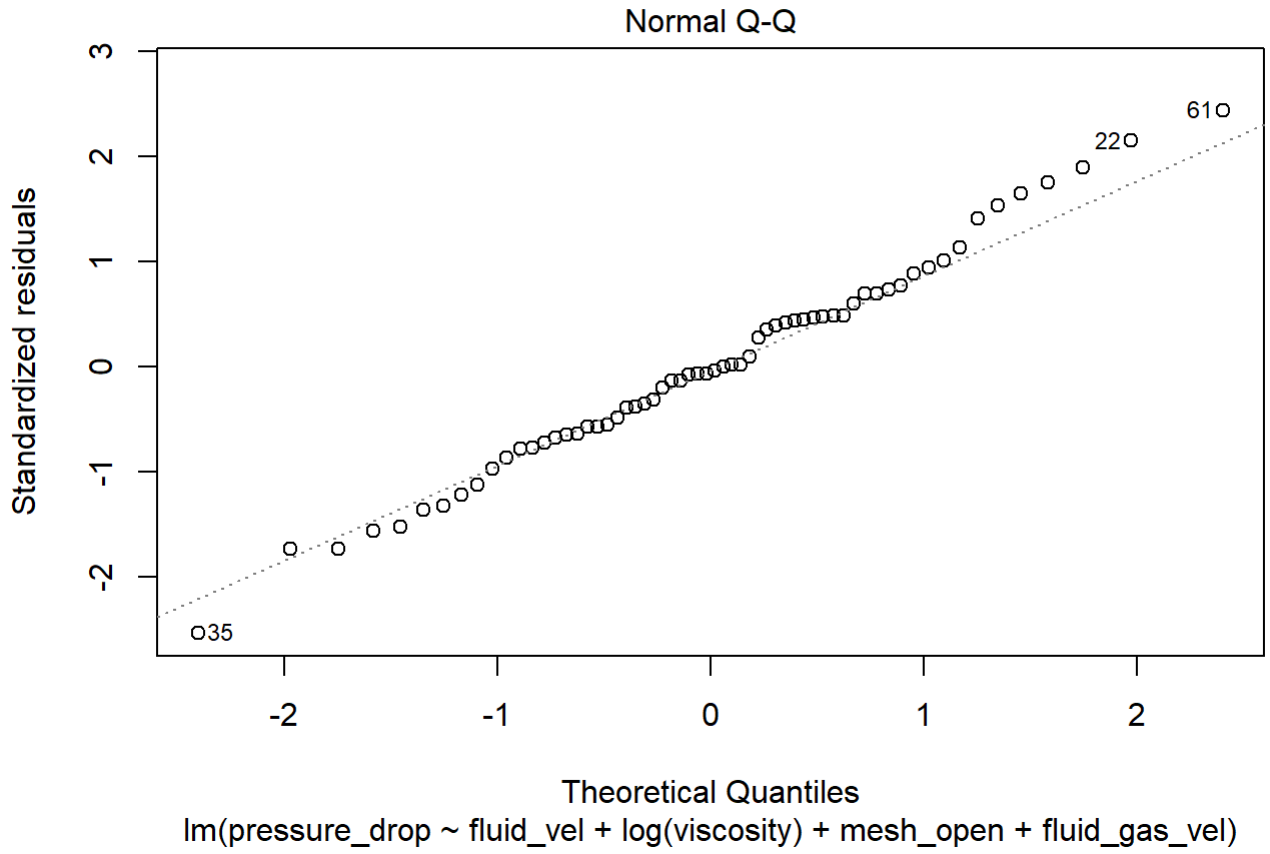
```
xtable(anova(model))
```

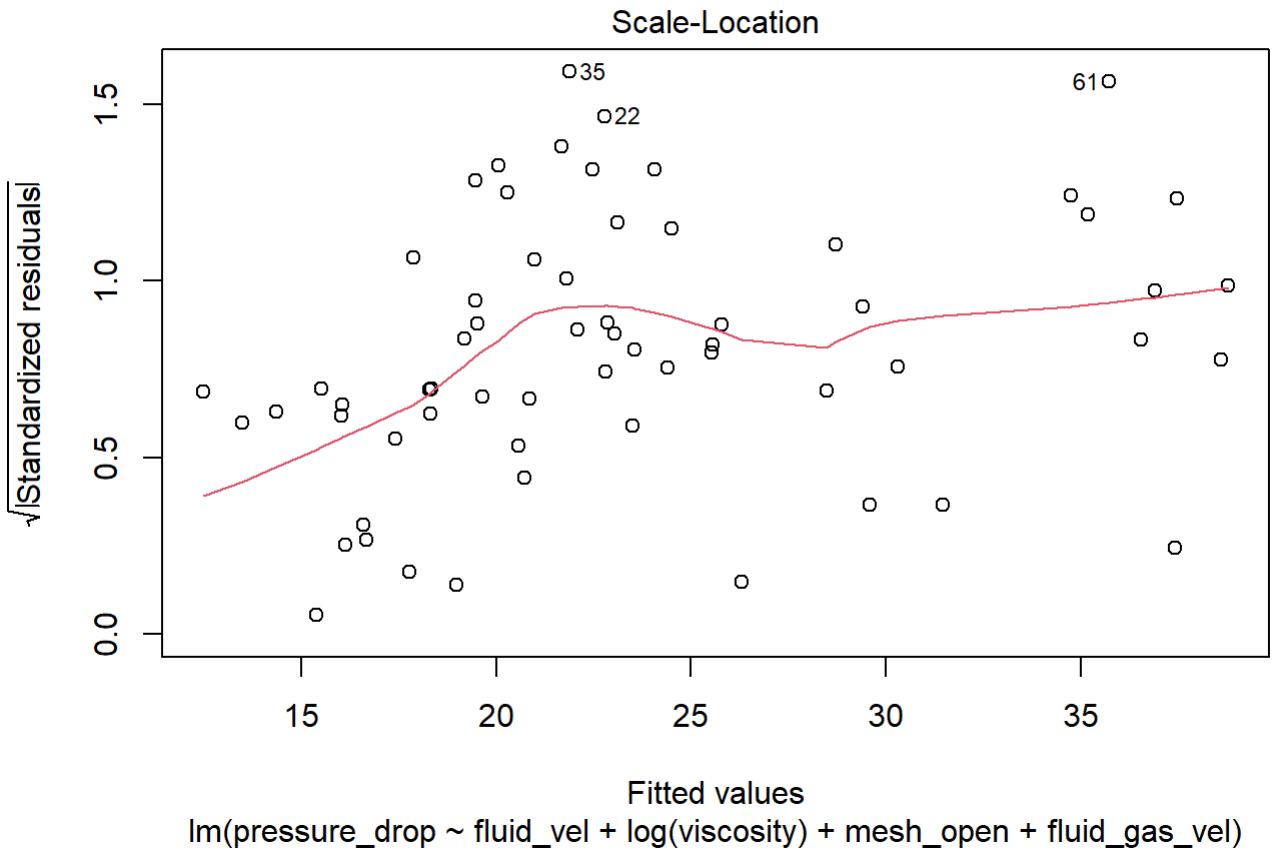
	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
fluid_vel	1	9.595968	9.595968	0.3156013	5.764656e-01
log(viscosity)	1	2420.863436	2420.863436	79.6196421	2.063142e-12
mesh_open	1	164.821263	164.821263	5.4207973	2.346952e-02
fluid_gas_vel	1	314.120098	314.120098	10.3310783	2.154800e-03
Residuals	57	1733.105202	30.405354	*NA*	*NA*

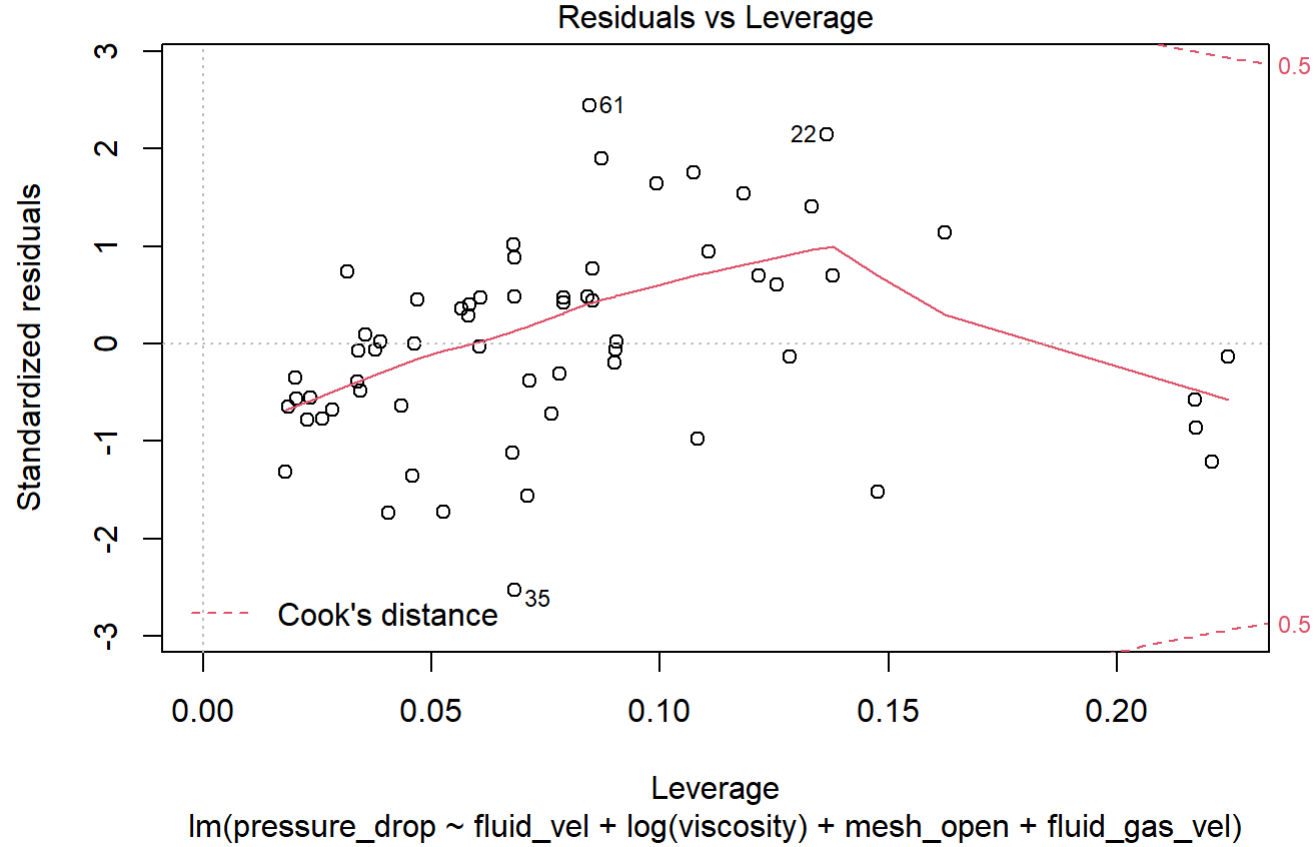
5 rows

```
plot(model)
```

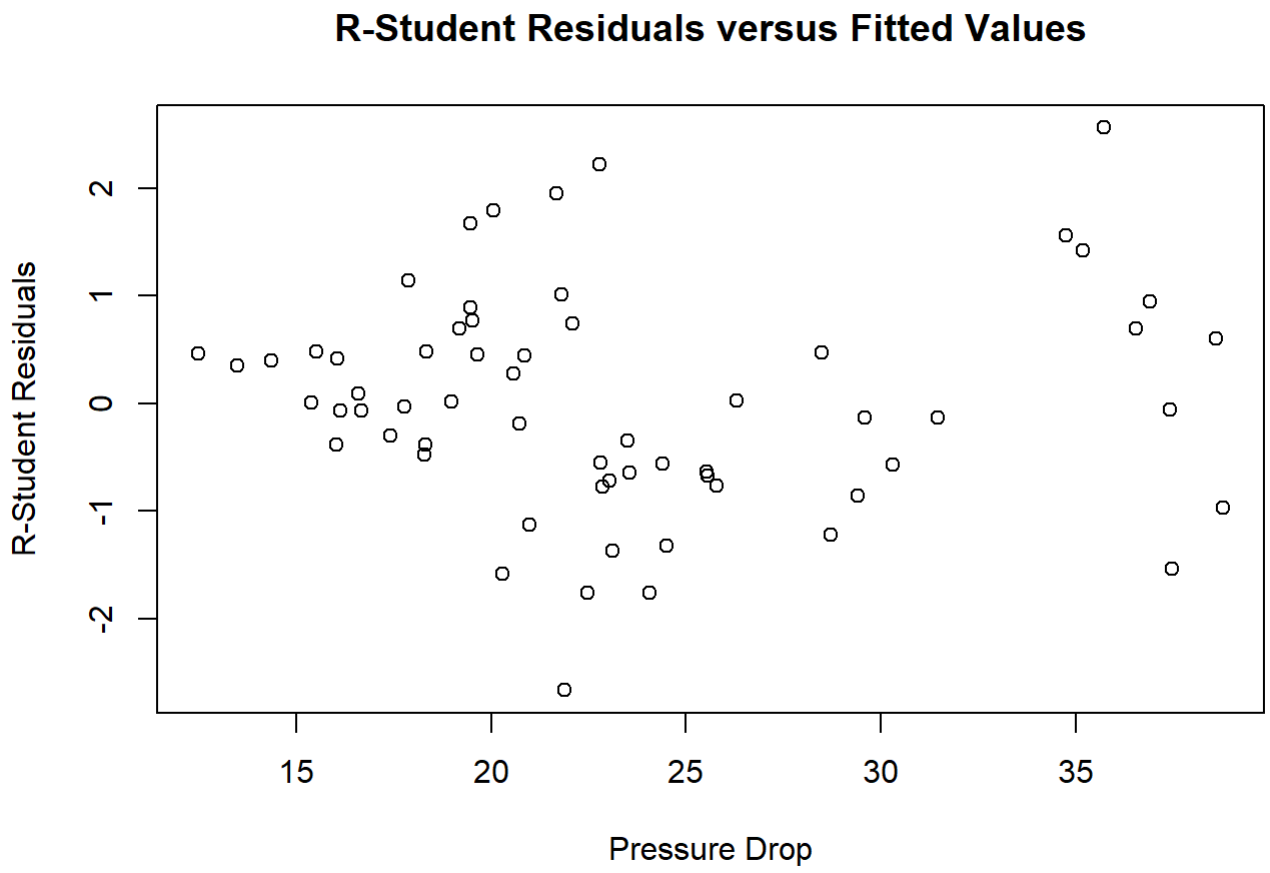

Residuals vs Fitted
Residuals
Fitted values
lm(pressure_drop ~ fluid_vel + log(viscosity) + mesh_open + fluid_gas_vel)
61
22
35
15
10
5
0
-5
-10
-15
15
20
25
30
35


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(pressure_drop ~ fluid_vel + log(viscosity) + mesh_open + fluid_gas_vel)
61
22
35
3
2
1
0
-1
-2
-2
-1
0
1
2

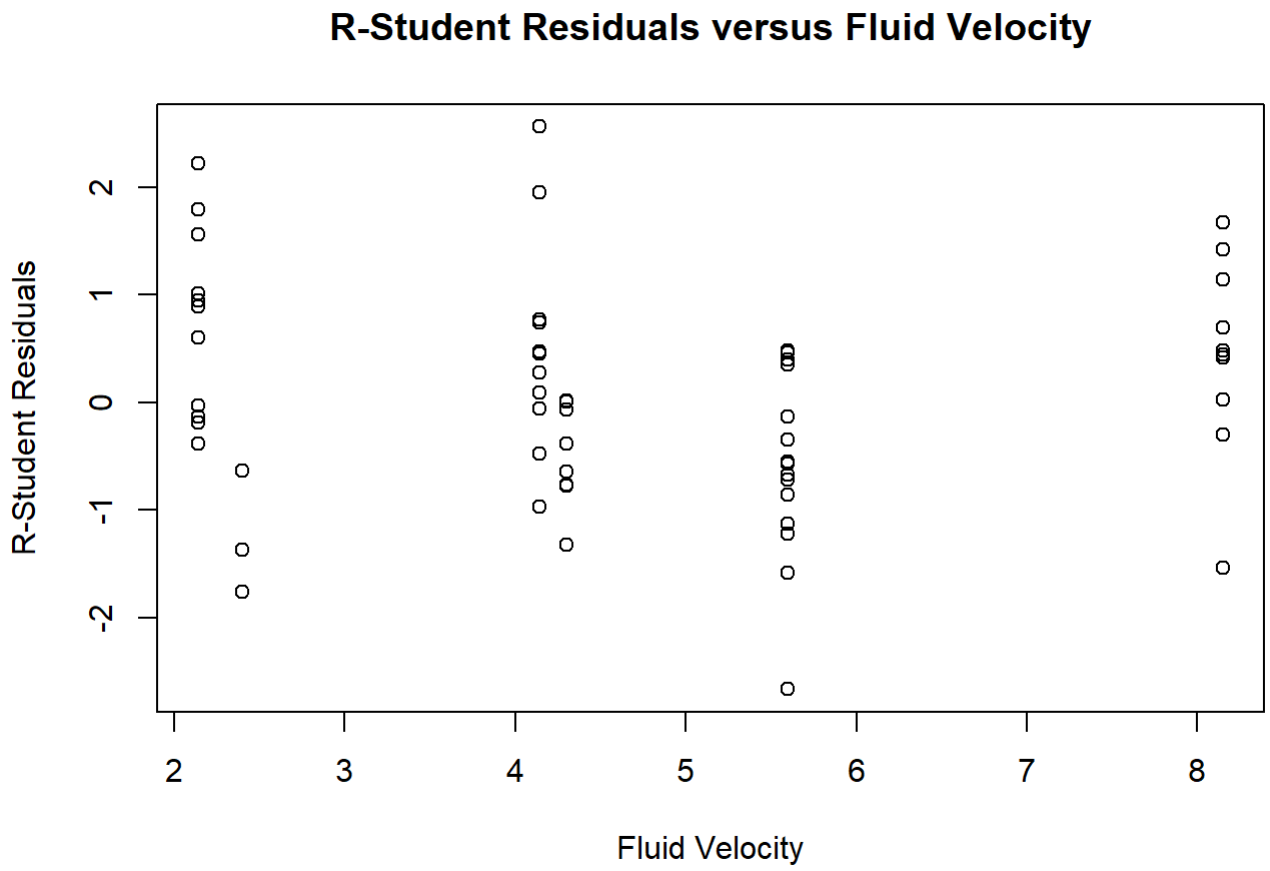

Scale-Location
√|Standardized residuals|
35
22
61
0.0
0.5
1.0
1.5
15
20
25
30
35
Fitted values
lm(pressure_drop ~ fluid_vel + log(viscosity) + mesh_open + fluid_gas_vel)


Residuals vs Leverage
Standardized residuals
61
22
35
Cook's distance
0.5
0.5
-3
-2
-1
0
1
2
3
0.00
0.05
0.10
0.15
0.20
Leverage
lm(pressure_drop ~ fluid_vel + log(viscosity) + mesh_open + fluid_gas_vel)

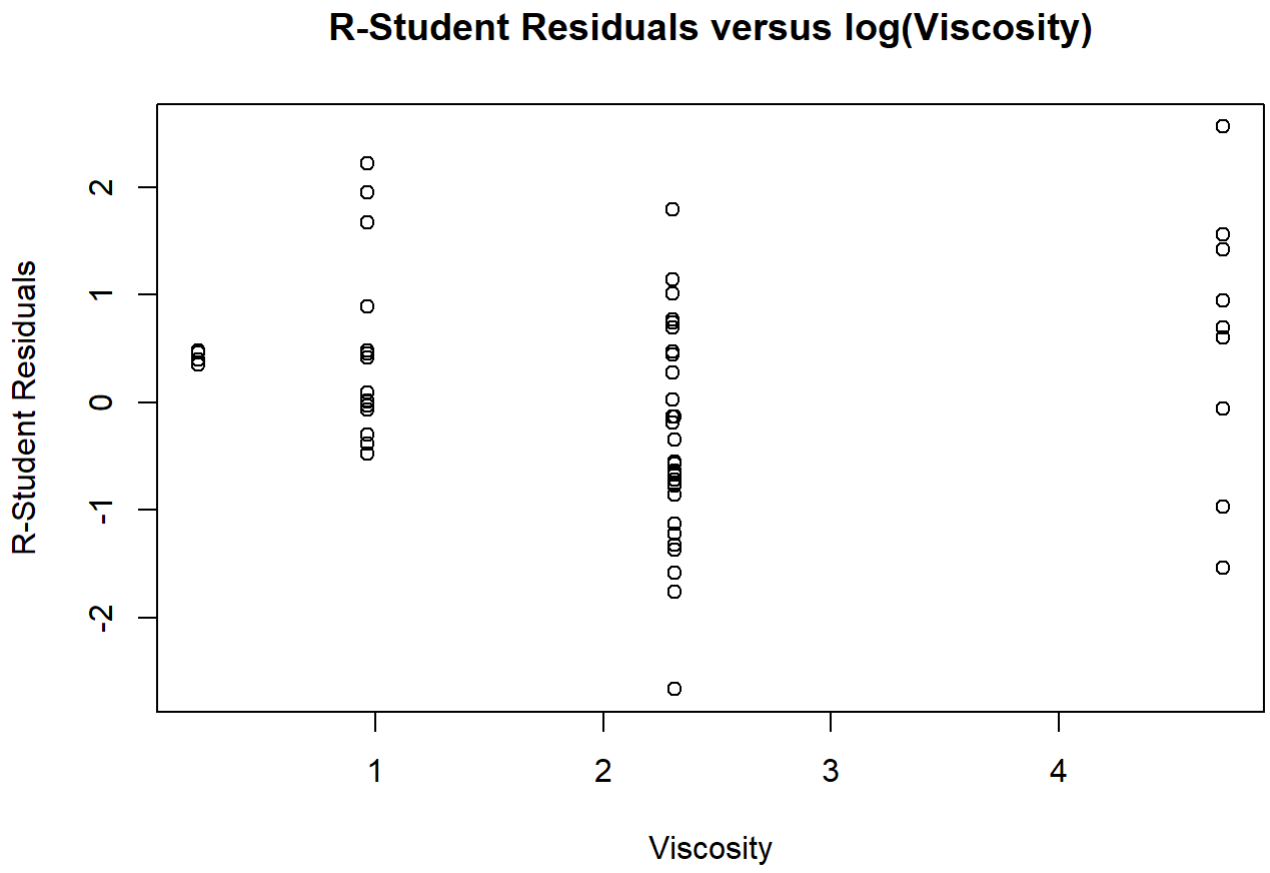
```
R_Student_Residuals <- rstudent(model)
y_hat <- model$fitted.values
plot(y_hat, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Pressure Drop", main = "R
-Student Residuals versus Fitted Values")
```



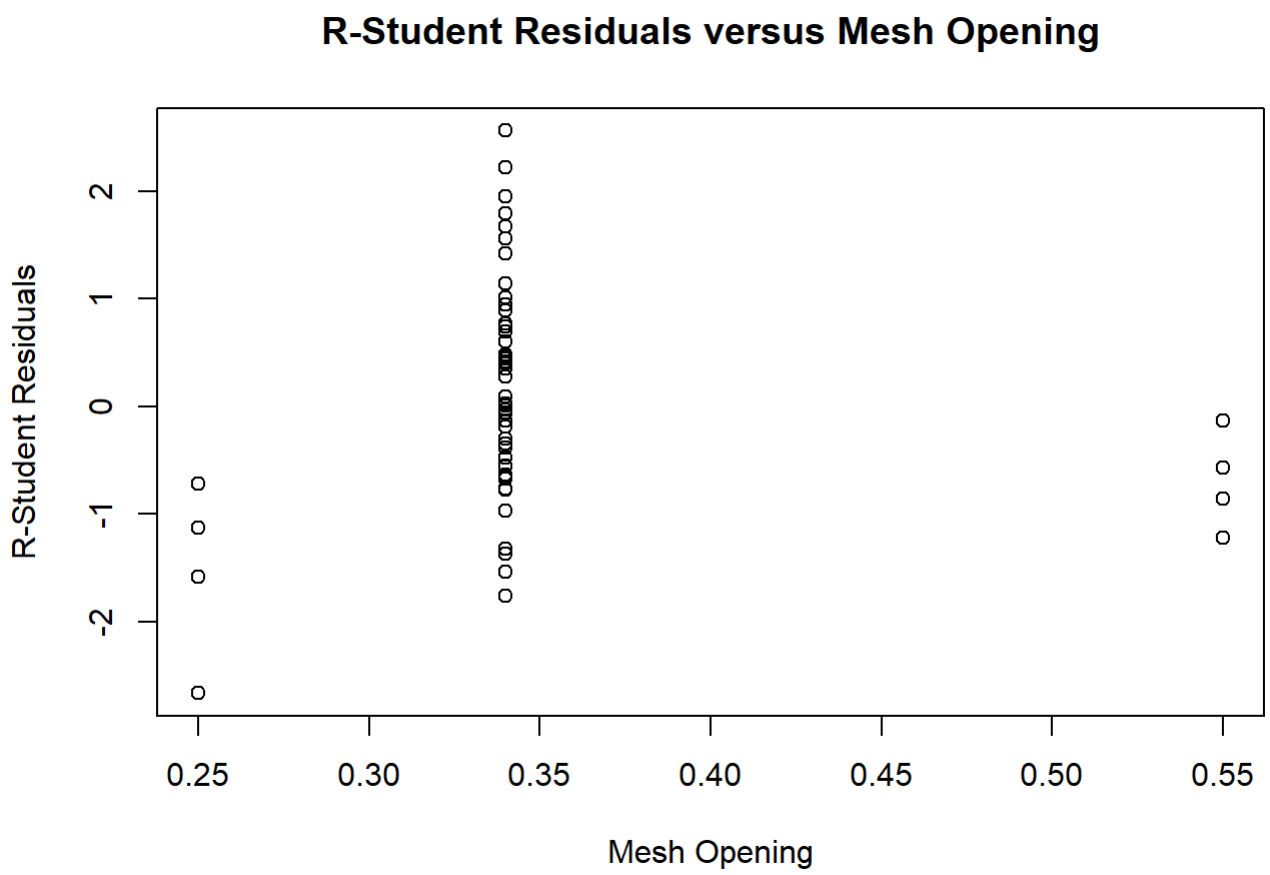

```
plot(fluid_vel, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Fluid Velocity", main
= "R-Student Residuals versus Fluid Velocity")
```



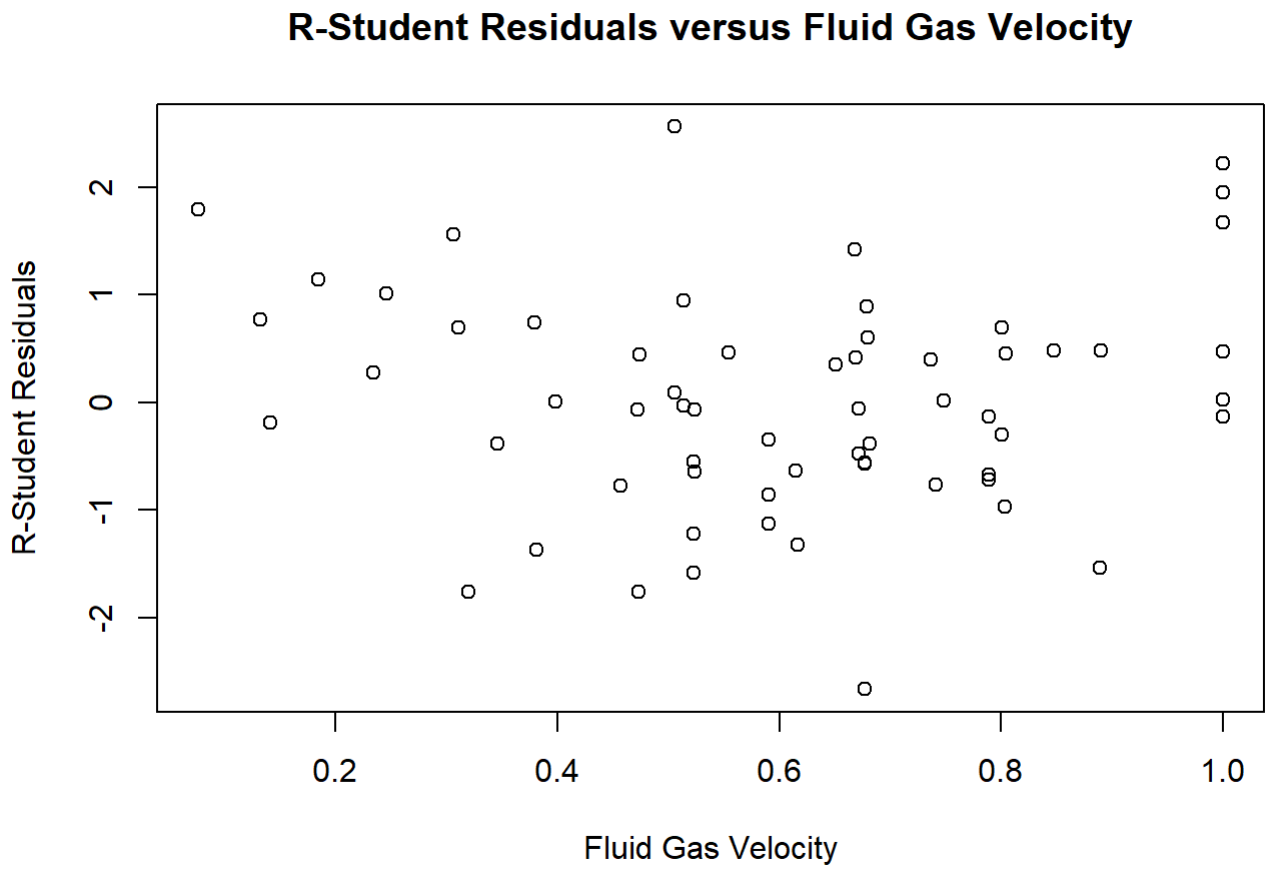

```
plot(log(viscosity), R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Viscosity", main
= "R-Student Residuals versus log(Viscosity)")
```




```
plot(mesh_open, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Mesh Opening", main =
"R-Student Residuals versus Mesh Opening")
```




```
plot(fluid_gas_vel, R_Student_Residuals, ylab = "R-Student Residuals", xlab = "Fluid Gas Velocit
y", main = "R-Student Residuals versus Fluid Gas Velocity")
```


R-Student Residuals versus Fluid Gas Velocity
R-Student Residuals
-2
-1
0
1
2
0.2
0.4
0.6
0.8
1.0
Fluid Gas Velocity