Markdown - STAT 6021 Project 1 Group 3

Group Members: Mary Evanston, Brian Wickens, OC Ofoma, Catherine Bielick

Executive Summary

Executive Summary text here.
use two spaces after end of previous line to create a new line.

Dataset and Variable Description

This dataset includes the parameters carat, clarity, color, cut, and prices for 1,214 diamonds available for purchase on bluenile.com.

Carat

description

Clarity

description

Color

description

Cut

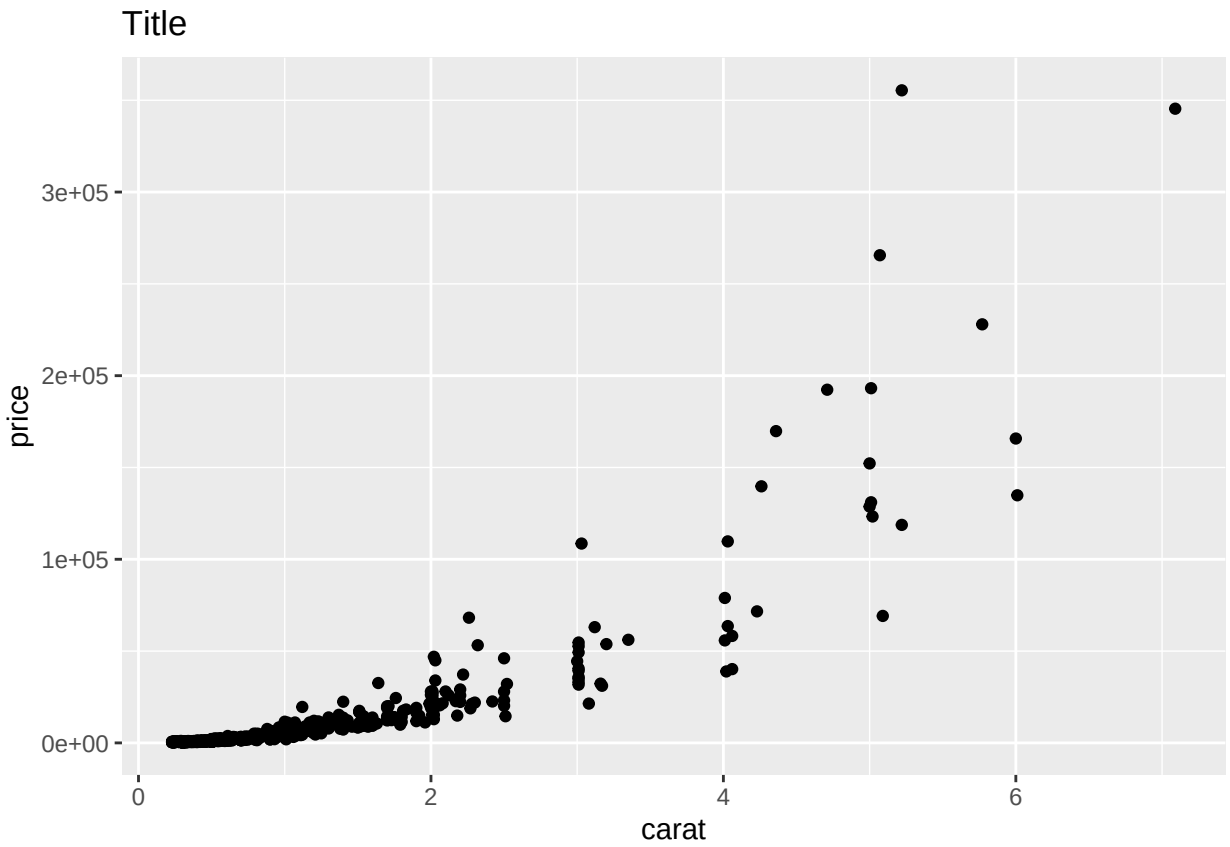
description

Price

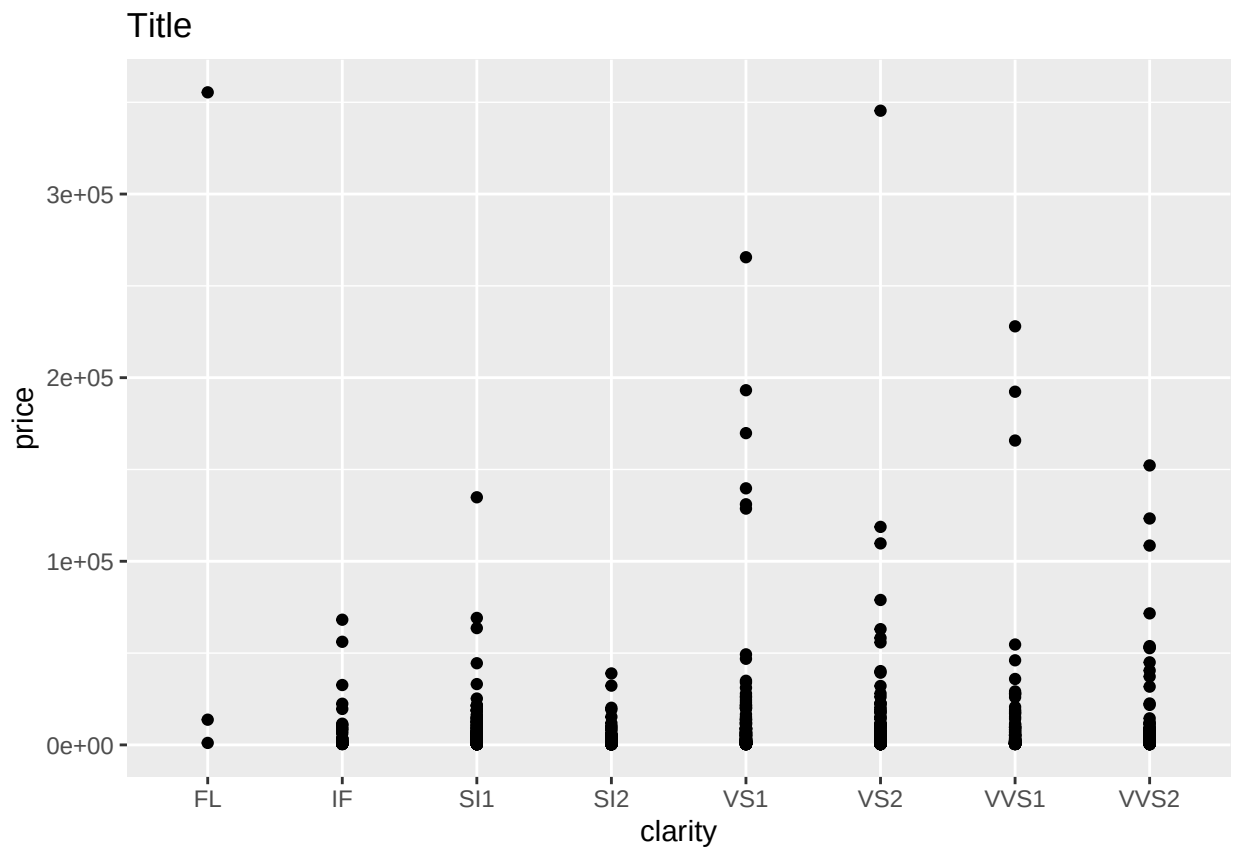
Cost of each diamond in USD. Unless otherwise specified, price will be considered the response variable in this analysis.

Visualizations and Commentary

```
ggplot(data, aes(x = carat, y = price)) +
  geom_point() +
  labs() +
  ggtitle("Title")
```




```
ggplot(data, aes(x = clarity, y = price)) +
  geom_point() +
  labs() +
  ggtitle("Title")
```

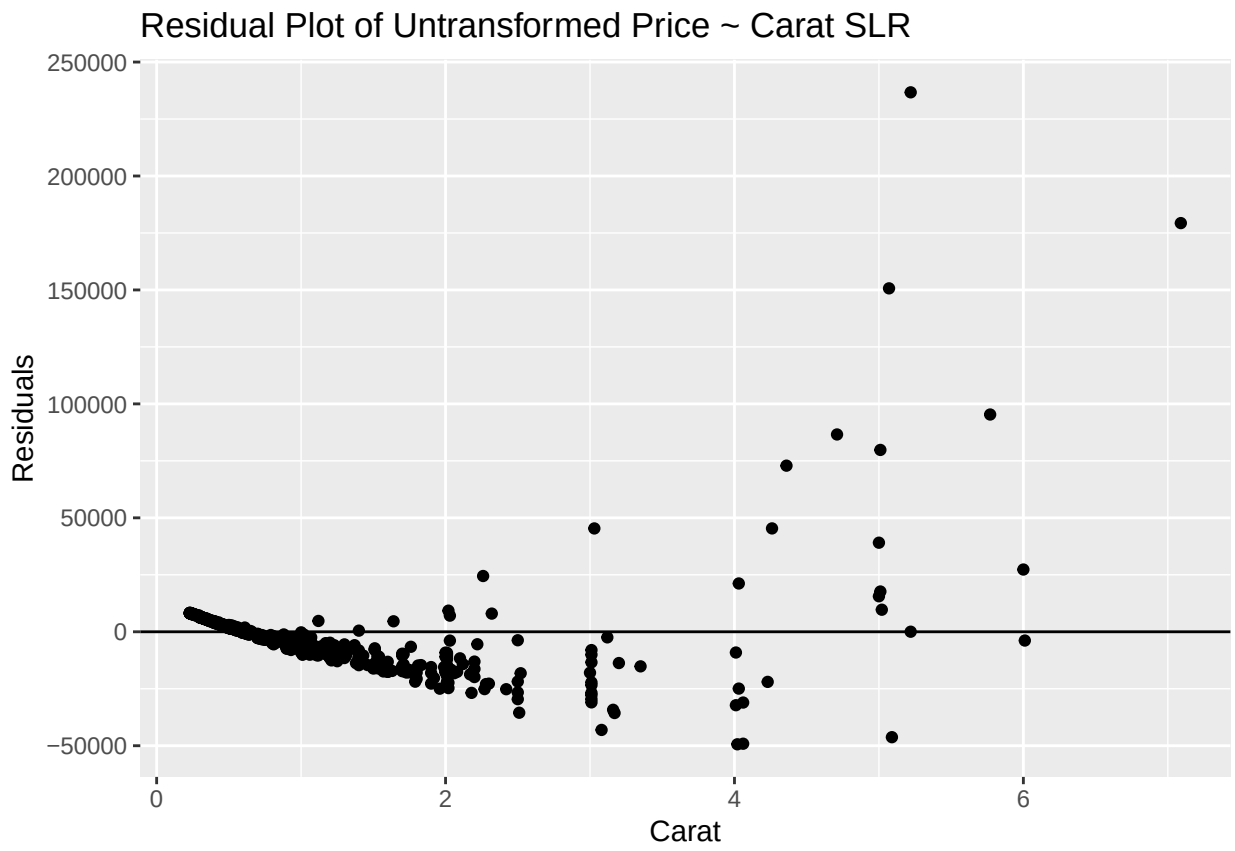


Regression Analysis

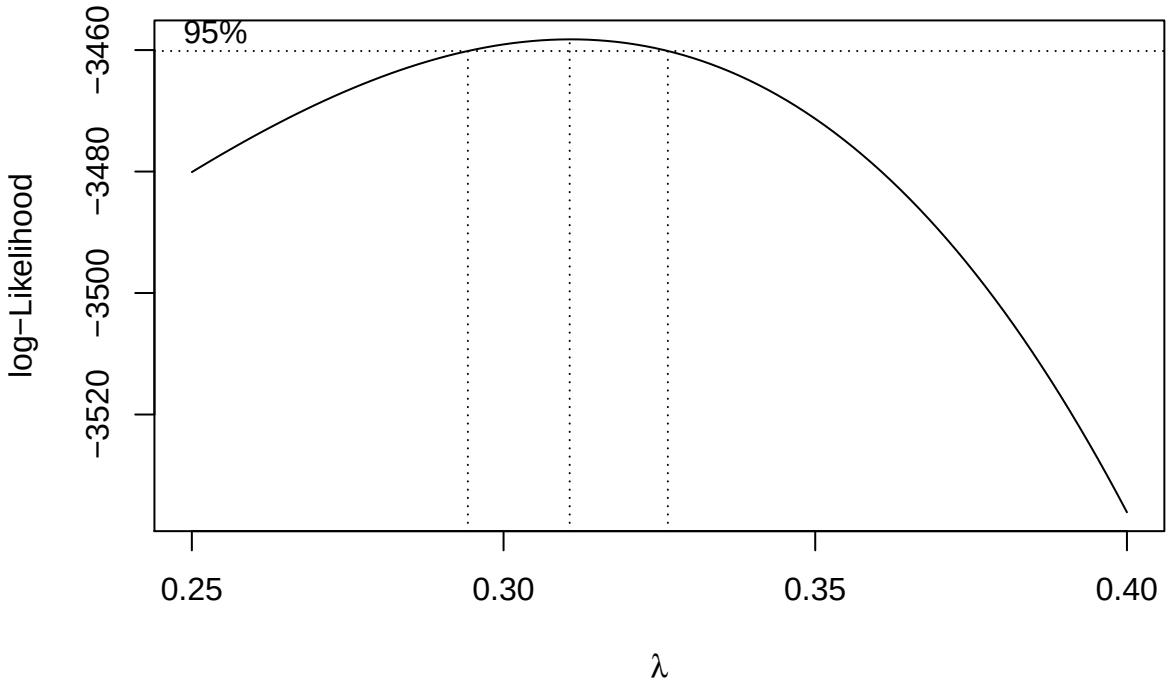
```
#Multiple Linear regression of price vs carat

reg1 = lm(price ~ carat, data = data)

ggplot(data = NULL, mapping = aes(x = reg1$model$carat , y = reg1$residuals)) +
  geom_point() +
  labs(x = 'Carat', y = 'Residuals') +
  geom_hline(yintercept = 0) +
  ggtitle('Residual Plot of Untransformed Price ~ Carat SLR')
```



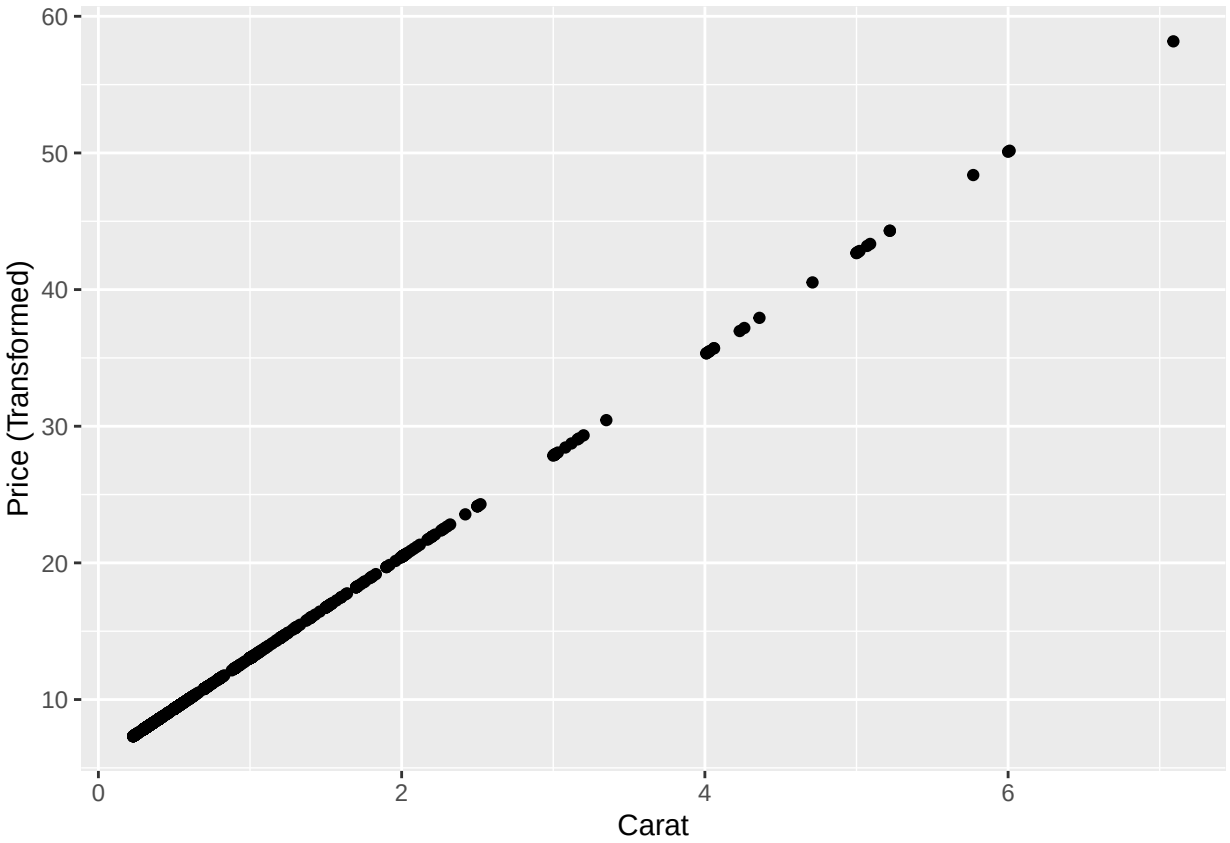

```
#Boxcox for transformation
boxcox(reg1, lambda = seq(0.25, 0.4, by = 0.05))
```




```
data$price.x = data$price^0.31

#Regression with lambda power transformation on price response variable
reg2 = lm(price.x ~ carat, data = data)

ggplot(data, aes(x = carat , y = reg2$fitted.values)) +
  geom_point() +
  labs(x = 'Carat', y = 'Price (Transformed)')
```


Price (Transformed)
60
50
40
30
20
10
0
2
4
6
Carat