1. From the scatterplot of TotalPrice vs Carat, we observe a nonlinear relationship that appears to be quadratic, which perhaps suggest that a second-degree polynomial regression model may fit the data well.
2. , where represents the predictor Carat. The R-squared value is reported as 0.9257, while the adjusted R-squared is 0.9253.
3. From the residuals plot, we notice some grouping at smaller values, and fanning at higher values, which is reflected in the Normal Q-Q plot at the further quantiles. However, the data do seem to be fairly well behaved overall, leading us to conclude that the model conditions are sufficiently satisfied.
4. , where represents the predictor Carat. The R-squared value is reported as 0.9257, while the adjusted R-squared is 0.9251. However, we note that the cubic coefficient is not statistically significant, implying that we need only include up to second-order terms.
5. Similar to part (c), there appears to still be some fanning and grouping in the residuals plot, and some deviation from the norm in the Normal Q-Q plot. However, overall the data appear to be well behaved, thus we conclude that the model assumptions are sufficiently satisfied.

library(readr)

## Warning: package 'readr' was built under R version 4.1.2

data <- read\_csv("Diamonds.csv")

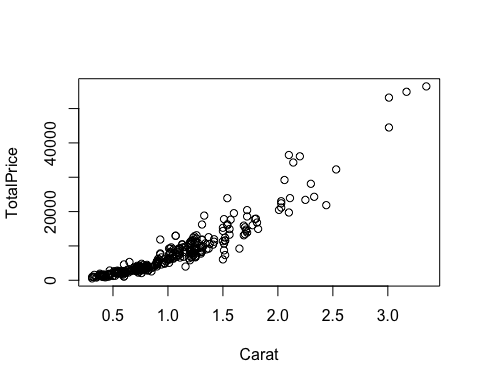
## Rows: 351 Columns: 6  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Color, Clarity  
## dbl (4): Carat, Depth, PricePerCt, TotalPrice  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

attach(data)

head(data)

## # A tibble: 6 × 6  
## Carat Color Clarity Depth PricePerCt TotalPrice  
## <dbl> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 1.08 E VS1 68.6 6693. 7229.  
## 2 0.31 F VVS1 61.9 3159 979.  
## 3 0.31 H VS1 62.1 1755 544.  
## 4 0.32 F VVS1 60.8 3159 1011.  
## 5 0.33 D IF 60.8 4759. 1570.  
## 6 0.33 G VVS1 61.5 2896. 956.

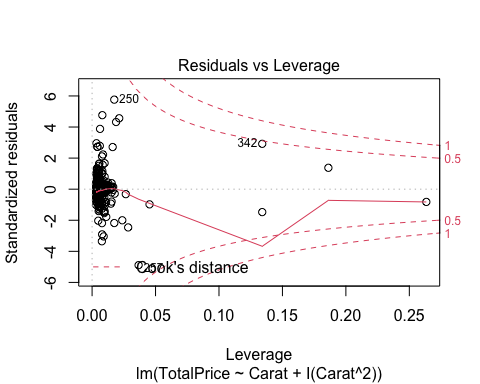
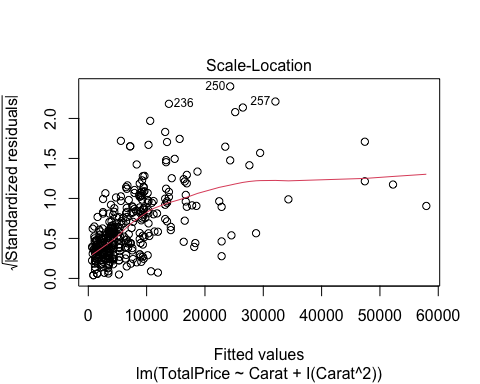
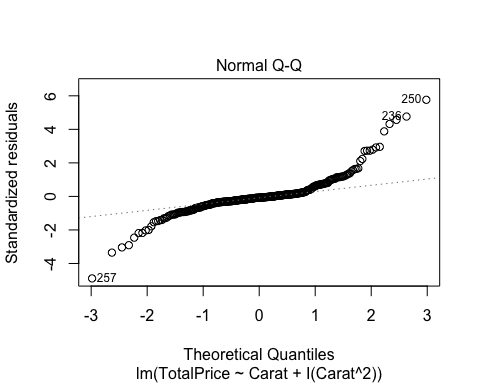
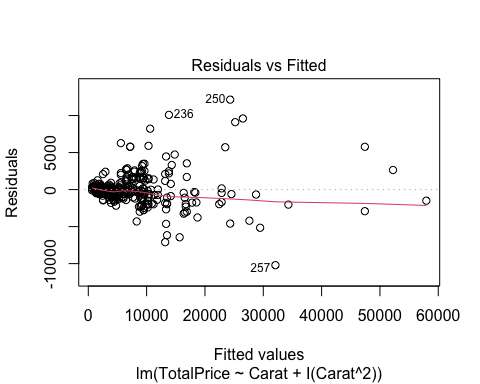
plot(TotalPrice~Carat)



model = lm(TotalPrice~Carat+I(Carat^2))  
summary(model)

##   
## Call:  
## lm(formula = TotalPrice ~ Carat + I(Carat^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10207.4 -711.6 -167.9 355.0 12147.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -522.7 466.3 -1.121 0.26307   
## Carat 2386.0 752.5 3.171 0.00166 \*\*   
## I(Carat^2) 4498.2 263.0 17.101 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2127 on 348 degrees of freedom  
## Multiple R-squared: 0.9257, Adjusted R-squared: 0.9253   
## F-statistic: 2168 on 2 and 348 DF, p-value: < 2.2e-16

plot(model)



model = lm(TotalPrice~Carat+I(Carat^2)+I(Carat^3))  
summary(model)

##   
## Call:  
## lm(formula = TotalPrice ~ Carat + I(Carat^2) + I(Carat^3))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10136.8 -725.2 -182.1 380.5 12220.8   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -723.44 875.50 -0.826 0.40919   
## Carat 2942.02 2185.44 1.346 0.17912   
## I(Carat^2) 4077.65 1573.80 2.591 0.00997 \*\*  
## I(Carat^3) 87.92 324.38 0.271 0.78652   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2130 on 347 degrees of freedom  
## Multiple R-squared: 0.9257, Adjusted R-squared: 0.9251   
## F-statistic: 1442 on 3 and 347 DF, p-value: < 2.2e-16

plot(model)

