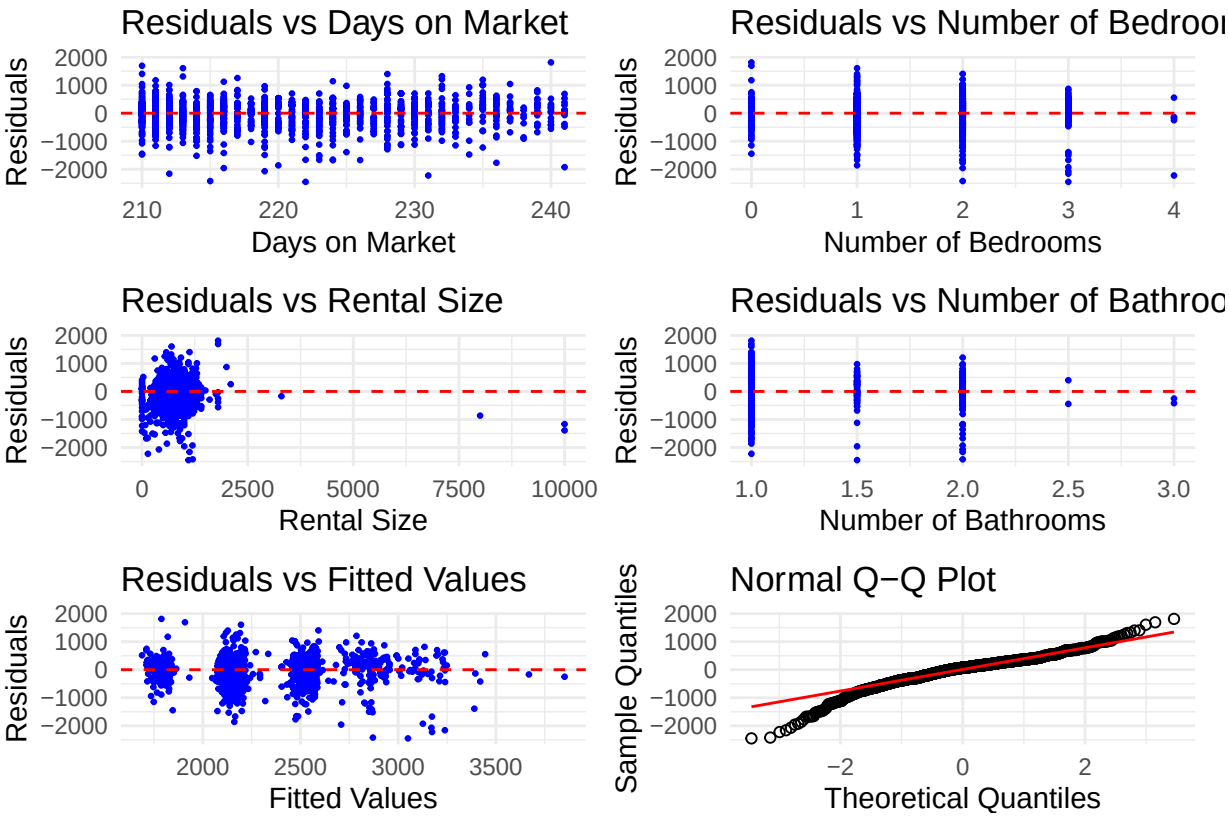
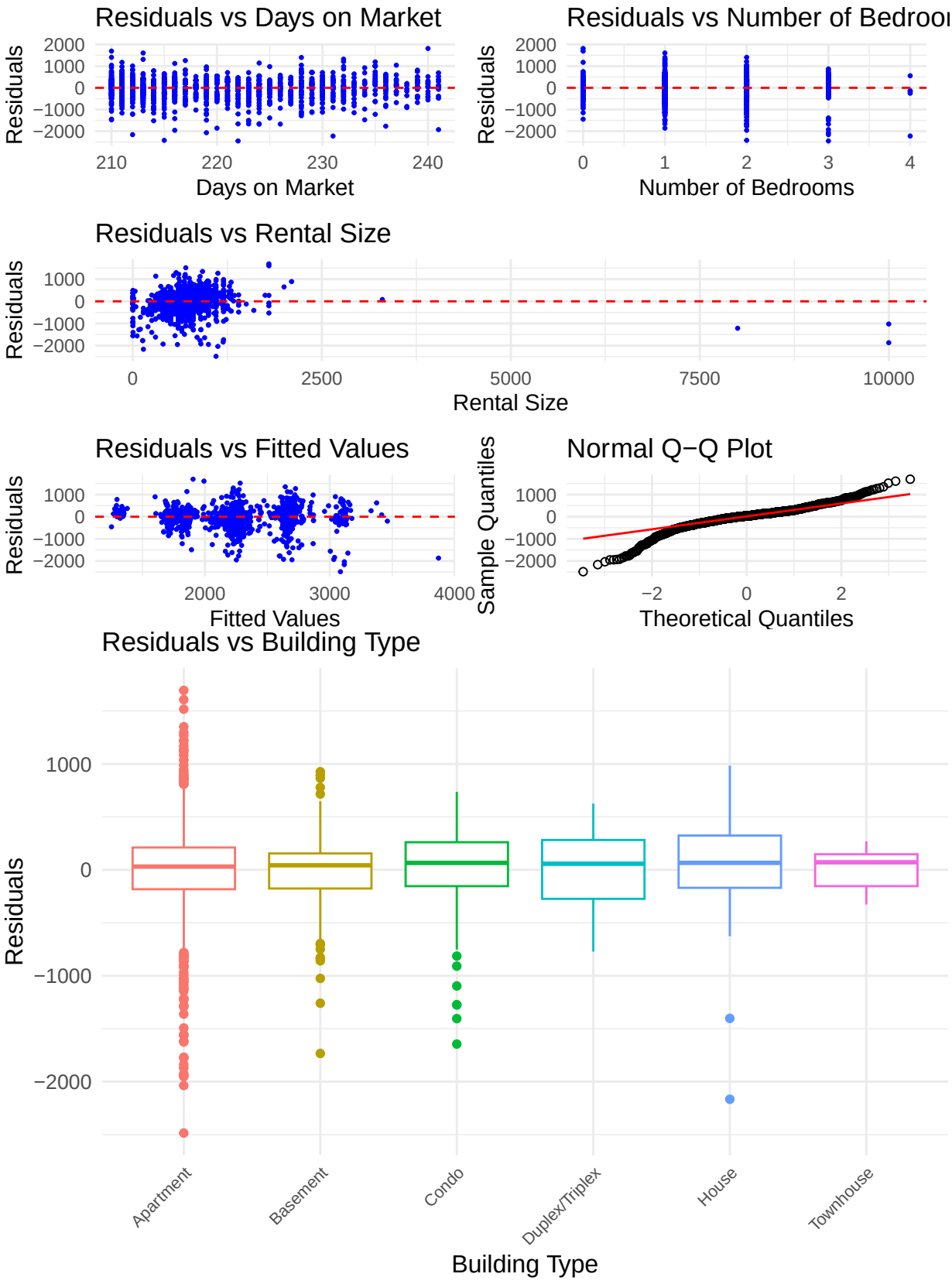
STA302 Project part 1

Hilary

2024-10-11

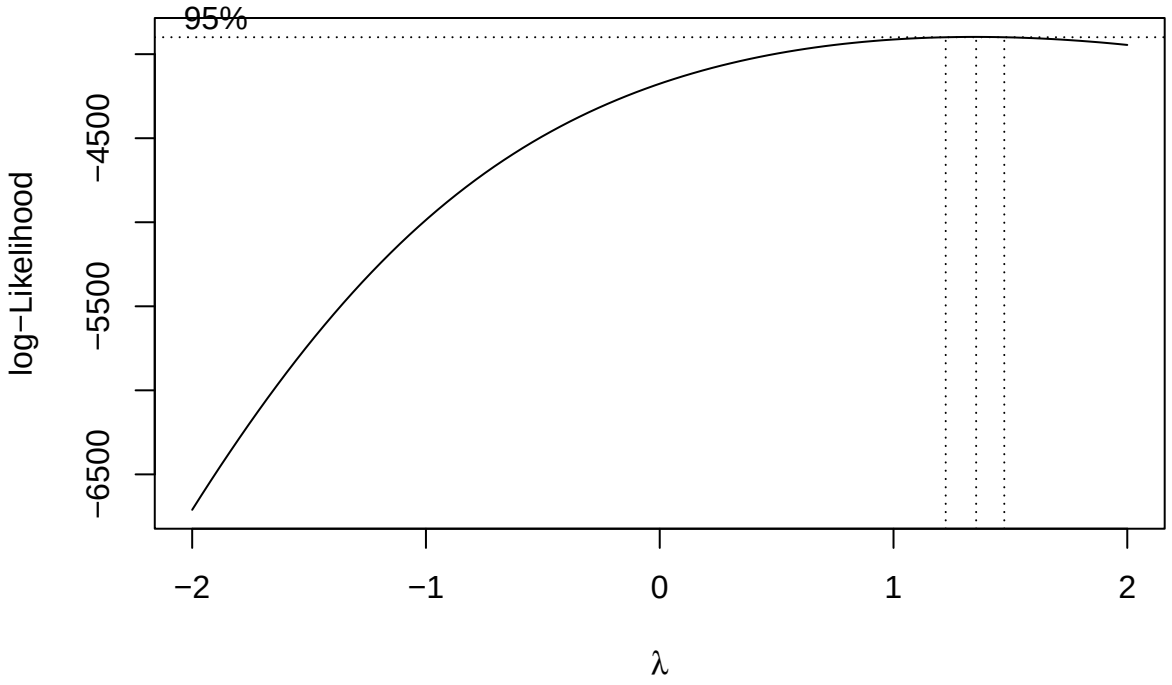
```
#write.csv(data_clean, "~/Course Materials/STA302/cleaned_data.csv")
```





1. Residuals vs. Fitted Values In the plot, residuals are plotted against the fitted values . The residuals do not show a clear pattern but cluster slightly, suggesting the linearity assumption is fitted.

2. Histogram of Residuals The histogram shows the distribution of residuals. While the residuals appear close to normal, small deviations affect confidence intervals and p-values. If deviations from normality are significant, we can apply a Box-Cox to help with analysis.
3. Normal Q-Q Plot The residuals mostly follow the reference line, but there are slight deviations at the tails. This suggests that the residuals are approximately normal, but the deviations at the tails might indicate some outliers.




```
summary(model1)
```

```
##
## Call:
## lm(formula = Price ~ Days_on_Market + Number_of_Bedrooms + Number_of_Bathrooms +
##     Rental_Size, data = data_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2450.13  -249.25    70.87   271.00  1811.84
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2356.6545   271.2164   8.689  < 2e-16 ***
## Days_on_Market        -4.0176     1.2386  -3.244   0.0012 **
## Number_of_Bedrooms   360.5152    15.6104  23.095  < 2e-16 ***
## Number_of_Bathrooms  294.7471    42.1571   6.992  3.8e-12 ***
## Rental_Size            0.0561     0.0259   2.166   0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 449.1 on 1815 degrees of freedom
## Multiple R-squared:  0.3504, Adjusted R-squared:  0.349
## F-statistic: 244.8 on 4 and 1815 DF,  p-value: < 2.2e-16
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = Price ~ Days_on_Market + Number_of_Bedrooms + Building_Type +
##     Rental_Size, data = data_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2485.26  -181.61    33.86   212.10  1694.57
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2442.99103  250.15709   9.766  < 2e-16 ***
## Days_on_Market                -2.90905    1.14564  -2.539 0.011193 *
## Number_of_Bedrooms           396.64347   13.58128  29.205  < 2e-16 ***
## Building_TypeBasement       -534.19880   27.65160 -19.319  < 2e-16 ***
## Building_TypeCondo            84.76074   40.92452   2.071 0.038486 *
## Building_TypeDuplex/Triplex -127.05225  125.23295  -1.015 0.310467
## Building_TypeHouse          -253.98219   66.79910  -3.802 0.000148 ***
## Building_TypeTownhouse      -184.32283  156.64821  -1.177 0.239483
## Rental_Size                   0.08923    0.02389   3.736 0.000193 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 412.5 on 1811 degrees of freedom
## Multiple R-squared:  0.4533, Adjusted R-squared:  0.4509
## F-statistic: 187.7 on 8 and 1811 DF,  p-value: < 2.2e-16
```

```
## [1] "ANOVA Test Results:"
```

```
## Analysis of Variance Table
##
## Model 1: Price ~ Days_on_Market + Number_of_Bedrooms + Number_of_Bathrooms +
##     Rental_Size
## Model 2: Price ~ Days_on_Market + Number_of_Bedrooms + Building_Type +
##     Rental_Size
##   Res.Df       RSS Df Sum of Sq      F    Pr(>F)
## 1   1815 366123959
## 2   1811 308119829  4  58004130 85.231 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "VIF for Model 1:"
```

```
##      Days_on_Market  Number_of_Bedrooms Number_of_Bathrooms         Rental_Size
##            1.012598            1.348865            1.171498            1.176335
```

```
## [1] "VIF for Model 2:"
```

```
##                        GVIF Df GVIF^(1/(2*Df))
## Days_on_Market     1.027165  1        1.013491
## Number_of_Bedrooms 1.210527  1        1.100240
## Building_Type      1.054538  5        1.005324
## Rental_Size        1.185961  1        1.089018
```