

Linear Model 0:

```
> house_lin_model0 <- lm(Price~., NY_House_DF)
> summary(house_lin_model0)
```

Call:
lm(formula = Price ~ ., data = NY_House_DF)

Residuals:

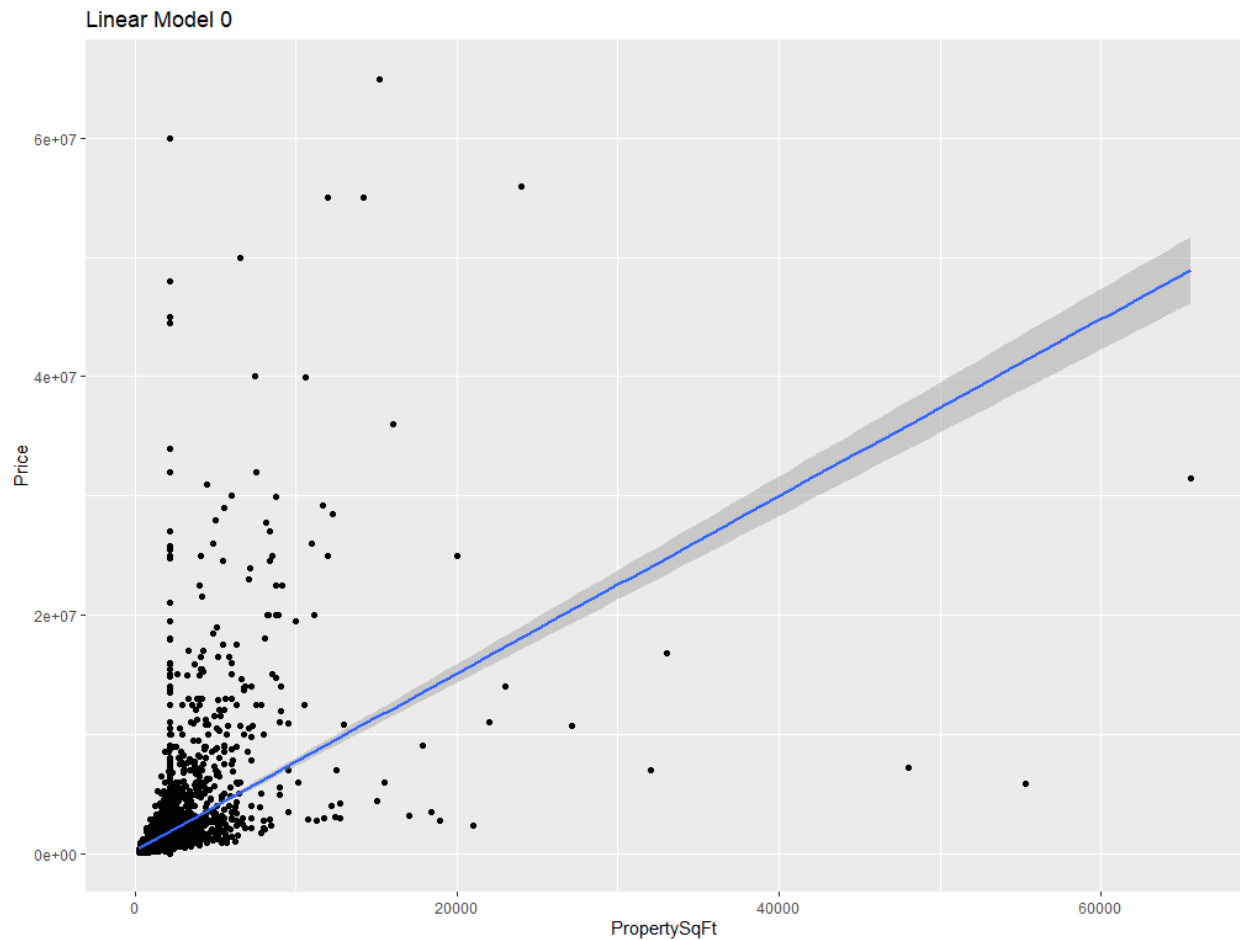
	Min	1Q	Median	3Q	Max
	-17896336	-1279874	-428920	621343	47629493

Coefficients:

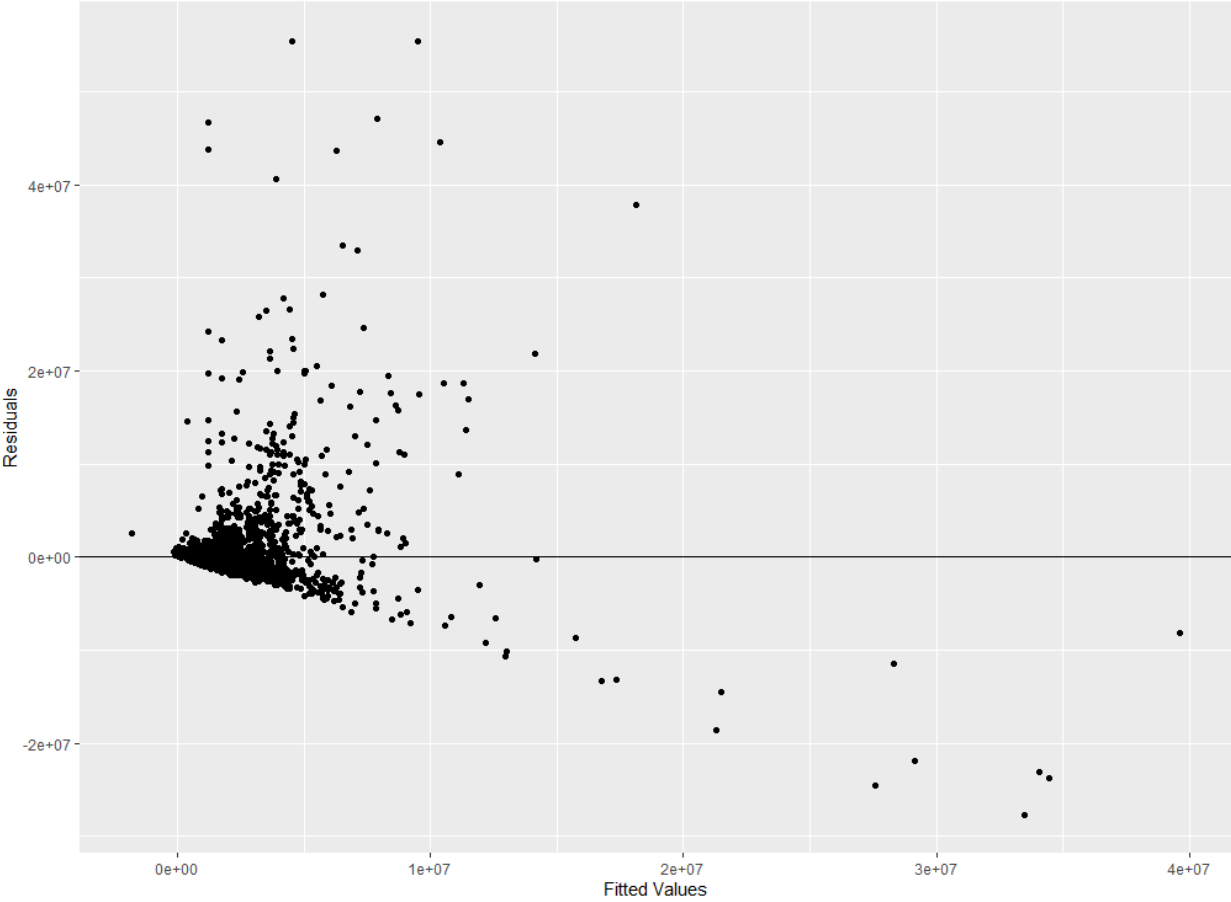
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.502e+07	6.215e+05	-56.345	<2e-16	***
PropertySqFt	3.179e+02	1.990e+01	15.977	<2e-16	***
Beds	-2.120e+05	2.433e+04	-8.713	<2e-16	***
Bath	5.884e+04	3.519e+04	1.672	0.0946	.
LogPrice	6.161e+06	1.092e+05	56.405	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2769000 on 4794 degrees of freedom
Multiple R-squared: 0.5366, Adjusted R-squared: 0.5362
F-statistic: 1388 on 4 and 4794 DF, p-value: < 2.2e-16



Residual vs. Fitted Values Plot (Model 0)



Linear Model 2:

```
> # Create model with log price
> NY_House_DF$LogPrice <- log10(NY_House_DF$Price)
> # Repeat Linear Model Now
> house_lin_model1 <- lm(LogPrice~Bath + Beds + PropertySqFt, NY_House_DF)
> summary(house_lin_model1)
```

Call:

```
lm(formula = LogPrice ~ Bath + Beds + PropertySqFt, data = NY_House_DF)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.2819	-0.2033	-0.0327	0.1691	1.7573

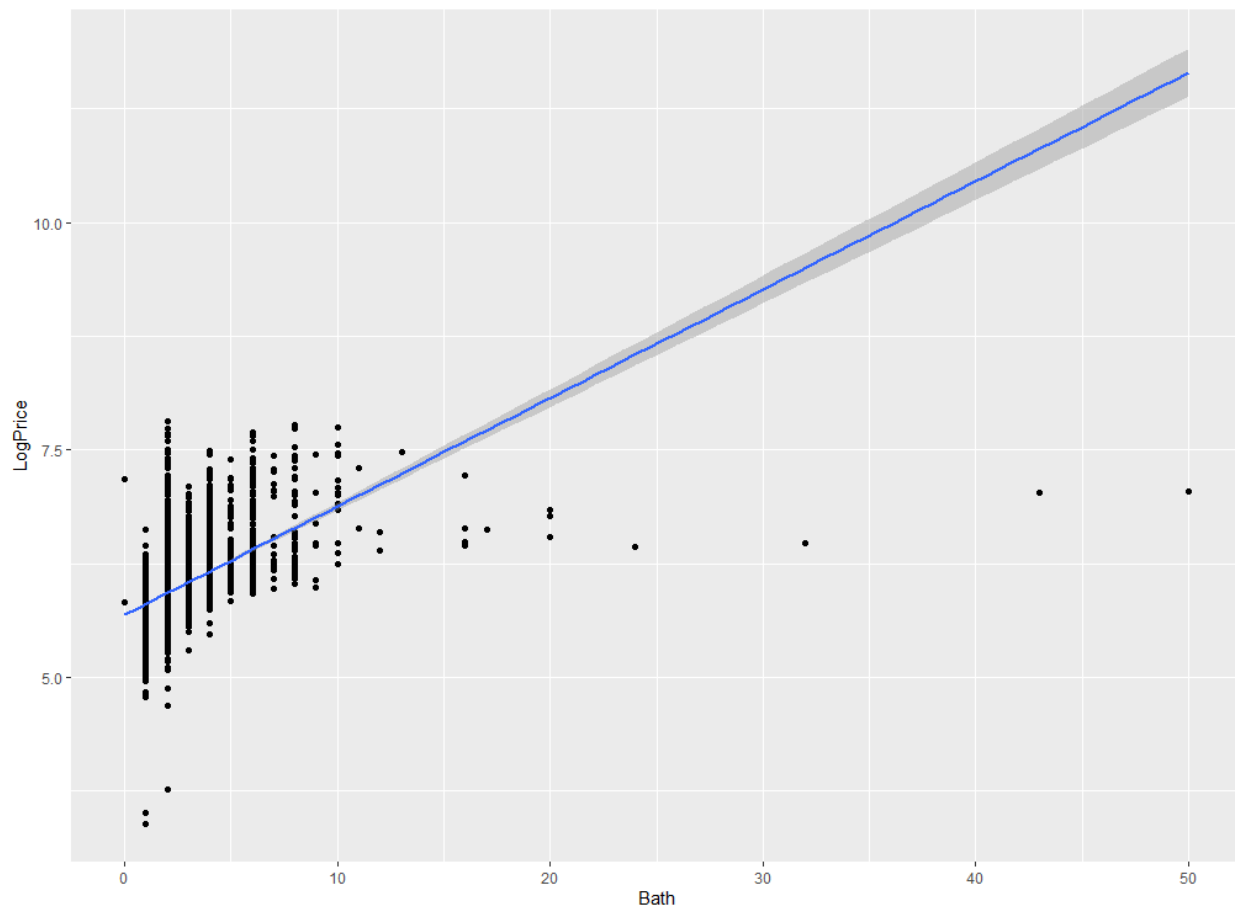
Coefficients:

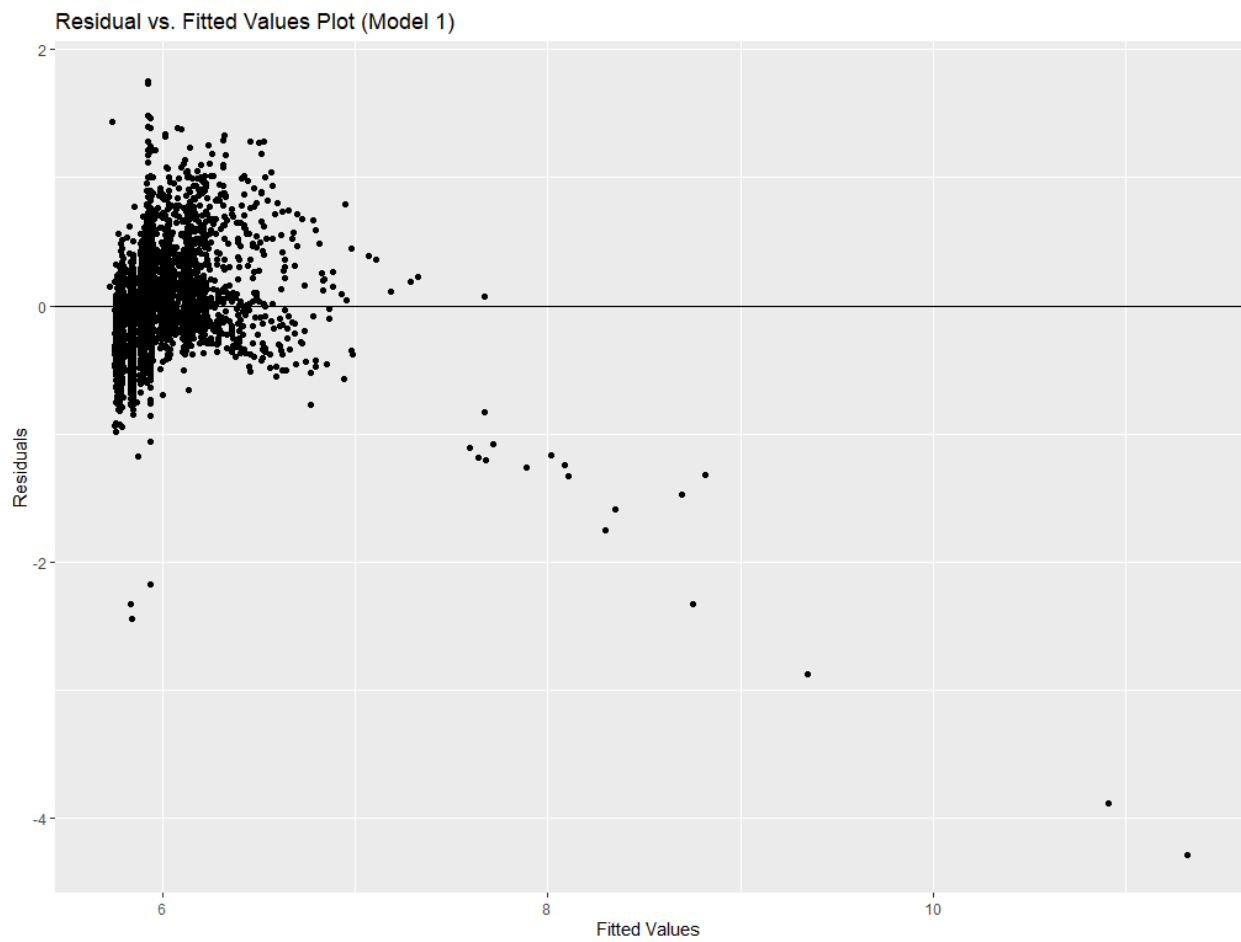
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.656e+00	8.988e-03	629.267	<2e-16 ***
Bath	9.927e-02	4.426e-03	22.428	<2e-16 ***
Beds	-5.952e-03	3.216e-03	-1.851	0.0643 .
PropertySqFt	4.546e-05	2.547e-06	17.846	<2e-16 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3661 on 4795 degrees of freedom
Multiple R-squared: 0.3187, Adjusted R-squared: 0.3183
F-statistic: 747.7 on 3 and 4795 DF, p-value: < 2.2e-16

Linear Model 1





Linear Model 2:

```
> # Remove beds because they have been insignificant  
> house_lin_model2 <- lm(Price~Bath + PropertySqFt, NY_House_DF)  
> summary(house_lin_model2)
```

Call:

```
lm(formula = Price ~ Bath + PropertySqFt, data = NY_House_DF)
```

Residuals:

Min	1Q	Median	3Q	Max
-26618097	-998904	-543599	-29804	55747307

Coefficients:

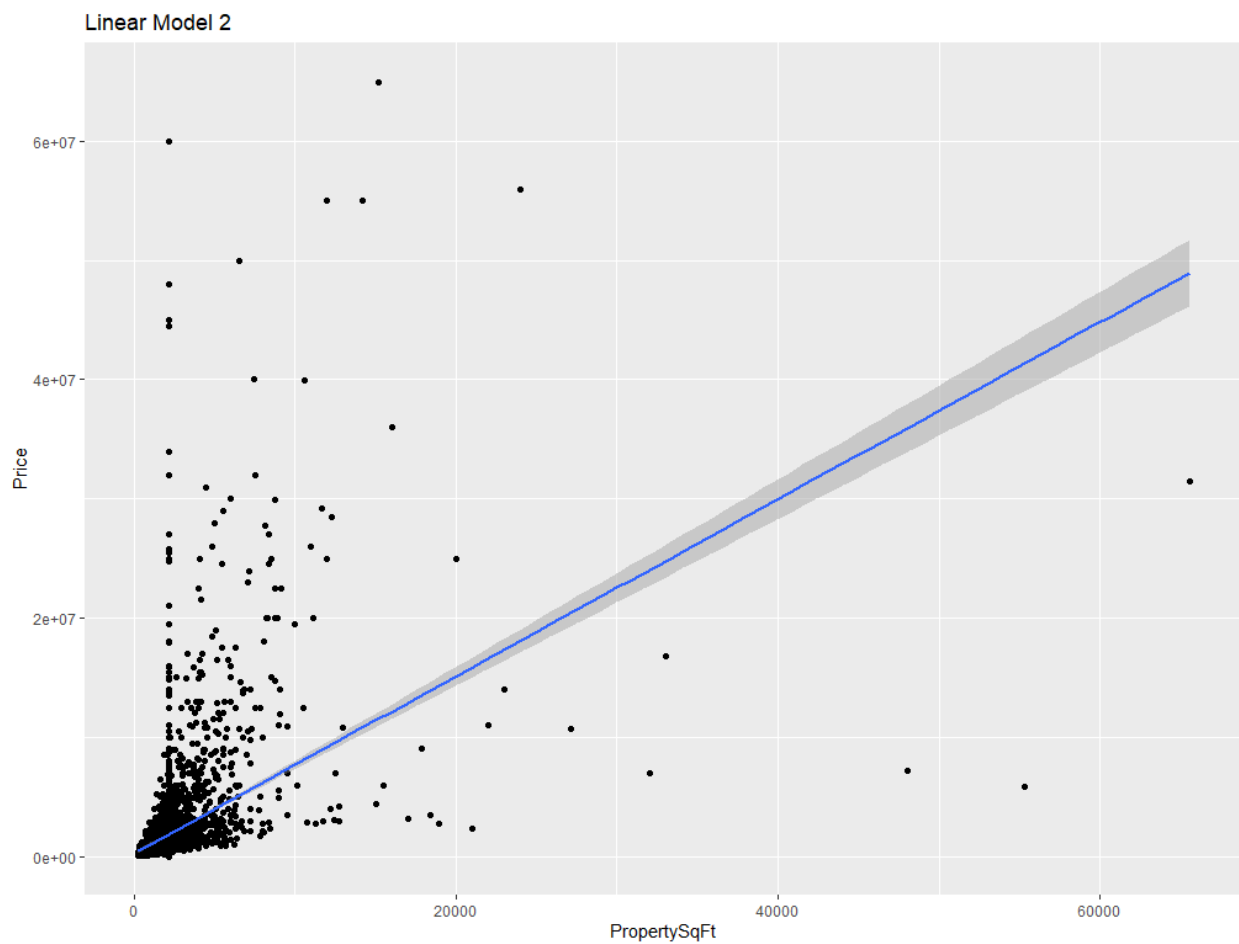
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-387730.27	83917.00	-4.62	3.93e-06	***
Bath	424643.85	30242.00	14.04	< 2e-16	***
PropertySqFt	578.36	24.88	23.24	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

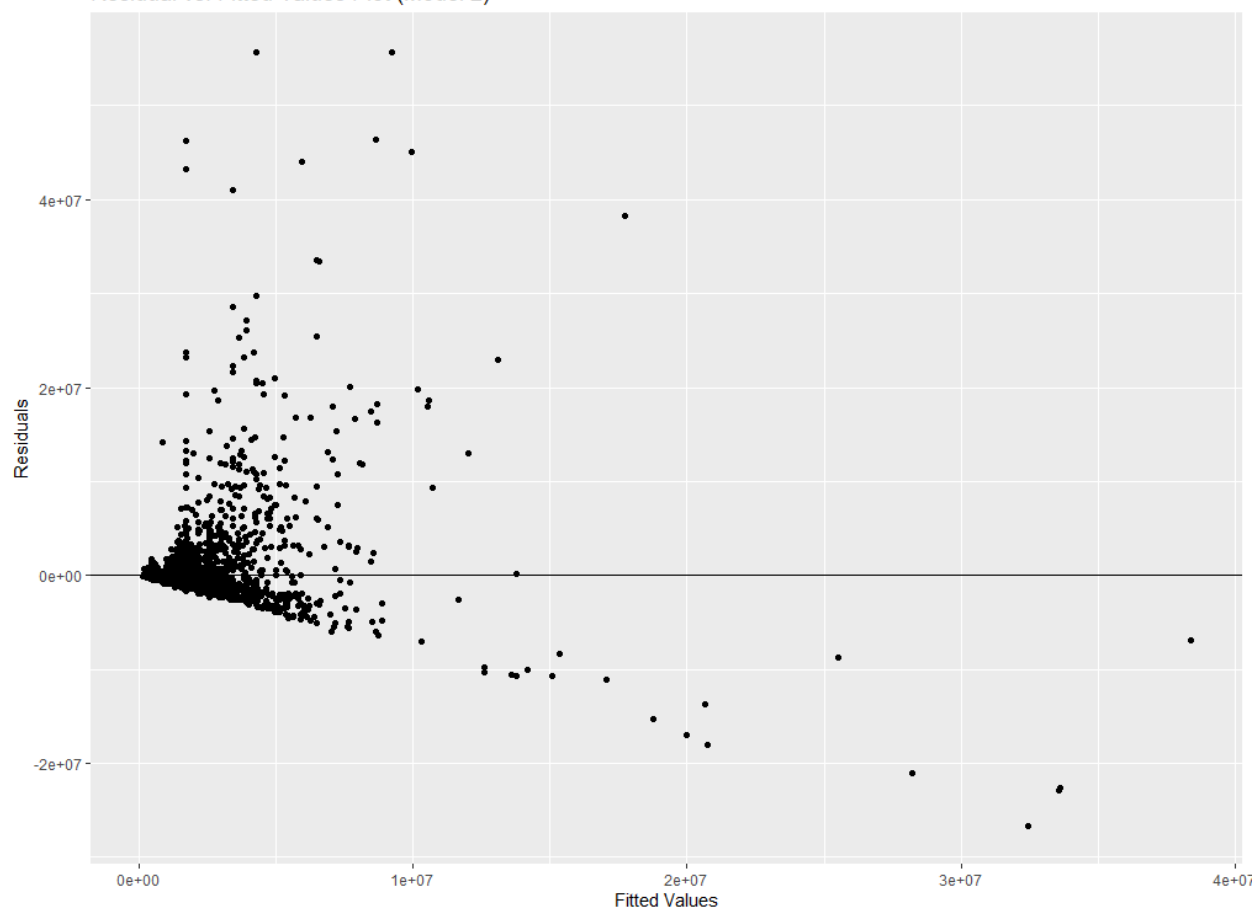
Residual standard error: 3594000 on 4796 degrees of freedom

Multiple R-squared: 0.2189, Adjusted R-squared: 0.2186

F-statistic: 672.1 on 2 and 4796 DF, p-value: < 2.2e-16



Residual vs. Fitted Values Plot (Model 2)



Analysis:

I think the best model is the model with all 3 features unedited. Though the log predictor looks to fit the data better visually, it has worse statistical values like p-value and r squared. Additionally, we can see that removing features like bed makes the model worse despite the feature itself being seemingly insignificant.