> summary(housing.data)

BROKERTITLE TYPE PRICE BEDS

BATH

Length: 4801 Length: 4801 Min. : 2.494e+03 Min. : 1.000

Min. : 0.000

Qu.: 1.000

Mode :character Mode :character Median :8.250e+05 Median : 3.000

Median : 2.000

Mean :2.357e+06 Mean : 3.357

Mean : 2.374

3rd Qu.:1.495e+06 3rd Qu.: 4.000 3rd

Qu.: 3.000

Max. :2.147e+09 Max. :50.000

Max. :50.000

PROPERTYSQFT ADDRESS STATE MAIN\_ADDRESS
Min. : 230 Length:4801 Length:4801 Length:4801
1st Qu.: 1200 Class :character Class :character Class :character

Median: 2184 Mode :character Mode :character Mode :character

Mean : 2184 3rd Qu.: 2184 Max. :65535

ADMINISTRATIVE\_AREA\_LEVEL\_2 LOCALITY SUBLOCALITY STREET\_NAME Length:4801 Length:4801 Length:4801 Length:4801

:character

Mode :character Mode :character Mode :character Mode

:character

LONG\_NAME FORMATTED\_ADDRESS LATITUDE LONGITUDE
Length:4801 Length:4801 Min. :40.50 Min. :-74.25
Class :character Class :character 1st Qu.:40.64 1st Qu.:-73.99
Mode :character Mode :character Median :40.73 Median :-73.95
Mean :40.71 Mean :-73.94
3rd Qu.:40.77 3rd Qu.:-73.87
Max. :40.91 Max. :-73.70

## > summary(model1)

Call:

lm(formula = PRICE ~ PROPERTYSQFT, data = housing.data)

Residuals:

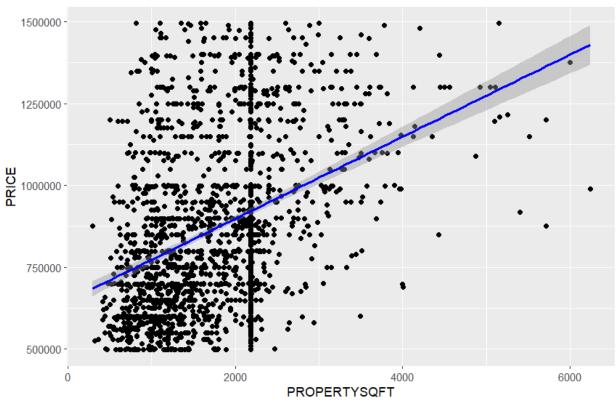
Min 1Q Median 3Q Max -488204 -184799 -48453 166120 745792

Coefficients:

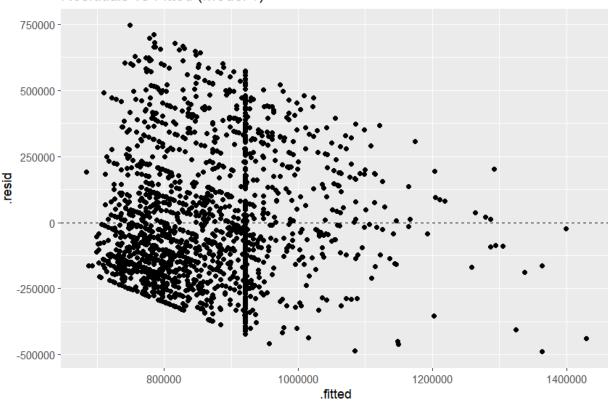
```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.472e+05 1.367e+04 47.34 <2e-16 ***
PROPERTYSQFT 1.254e+02 6.909e+00 18.14 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 248700 on 2413 degrees of freedom Multiple R-squared: 0.12, Adjusted R-squared: 0.1197 F-statistic: 329.2 on 1 and 2413 DF, p-value: < 2.2e-16





# Residuals vs Fitted (Model 1)



## > summary(model2)

```
Call:
```

lm(formula = PRICE ~ PROPERTYSQFT + BEDS, data = housing.data)

#### Residuals:

Min 1Q Median 3Q Max -493279 -182697 -54960 155569 770366

### Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.271e+05 1.364e+04 45.976 <2e-16 \*\*\*

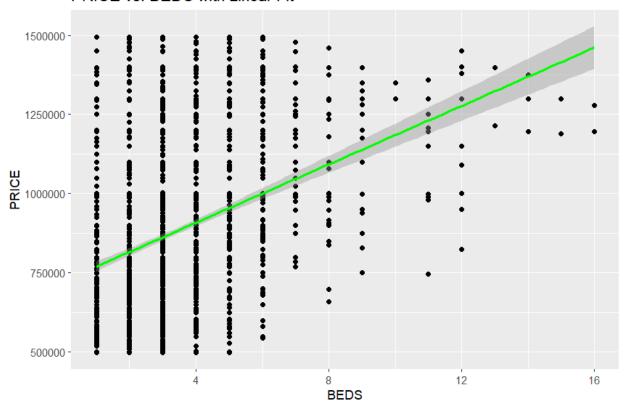
PROPERTYSQFT 8.540e+01 8.135e+00 10.499 <2e-16 \*\*\*

BEDS 2.802e+04 3.133e+03 8.944 <2e-16 \*\*\*

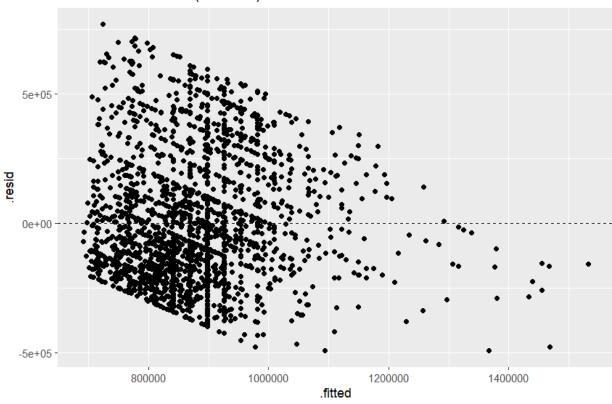
--
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 244700 on 2412 degrees of freedom Multiple R-squared: 0.1483, Adjusted R-squared: 0.1476 F-statistic: 210 on 2 and 2412 DF, p-value: < 2.2e-16

PRICE vs. BEDS with Linear Fit



# Residuals vs Fitted (Model 2)



# > summary(model3)

## Call:

lm(formula = PRICE ~ PROPERTYSQFT + BEDS + BATH, data = housing.data)

#### Residuals:

Min 1Q Median 3Q Max -668919 -183467 -54426 151801 777403

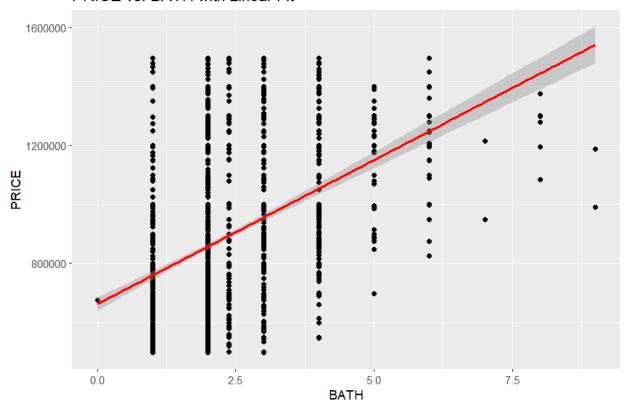
#### Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 592551.41 13816.13 42.888 < 2e-16 \*\*\*
PROPERTYSQFT 64.82 8.24 7.867 5.44e-15 \*\*\*
BEDS 5676.12 3806.80 1.491 0.136
BATH 66605.26 6704.40 9.935 < 2e-16 \*\*\*

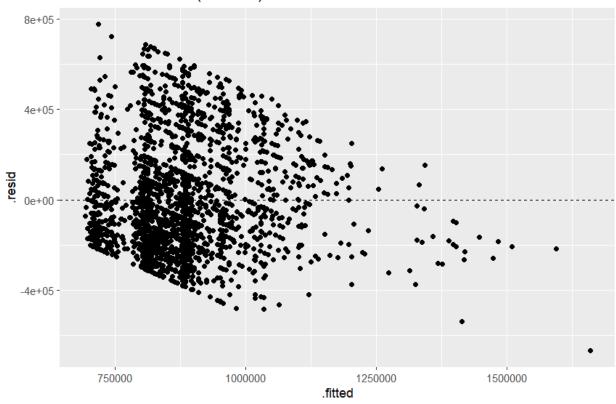
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1

Residual standard error: 239900 on 2411 degrees of freedom Multiple R-squared: 0.1818, Adjusted R-squared: 0.1808 F-statistic: 178.6 on 3 and 2411 DF, p-value: < 2.2e-16

PRICE vs. BATH with Linear Fit



# Residuals vs Fitted (Model 3)



# > AIC(model1, model2, model3)

df AIC model1 3 66865.36 model2 4 66788.57 model3 5 66693.68

# > BIC(model1, model2, model3)

df BIC model1 3 66882.73 model2 4 66811.73 model3 5 66722.63