**Log(Price) vs. log(SqFt):**

> summary(lin.mod.SqFt)

Call:

lm(formula = log10(NY.house.data.subset.SQFT$PRICE) ~ log10(NY.house.data.subset.SQFT$PROPERTYSQFT),

data = NY.house.data.subset.SQFT)

Residuals:

Min 1Q Median 3Q Max

-0.97152 -0.19195 -0.04776 0.18279 0.95127

Coefficients:

Estimate Std. Error t value

(Intercept) 2.52306 0.05536 45.58

log10(NY.house.data.subset.SQFT$PROPERTYSQFT) 1.08266 0.01727 62.71

Pr(>|t|)

(Intercept) <2e-16 \*\*\*

log10(NY.house.data.subset.SQFT$PROPERTYSQFT) <2e-16 \*\*\*

---

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

Residual standard error: 0.2785 on 3131 degrees of freedom

Multiple R-squared: 0.5567, Adjusted R-squared: 0.5566

F-statistic: 3932 on 1 and 3131 DF, p-value: < 2.2e-16

A graph with a blue line

AI-generated content may be incorrect. A graph showing a number of black dots

AI-generated content may be incorrect.

**Price vs. Beds:**

> summary(lin.mod.BATH)

Call:

lm(formula = NY.house.data.subset.BATH$PRICE ~ NY.house.data.subset.BATH$BATH,

data = NY.house.data.subset.BATH)

Residuals:

Min 1Q Median 3Q Max

-5793930 -720876 -293126 73275 18393709

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -255426 57602 -4.434 9.45e-06 \*\*\*

NY.house.data.subset.BATH$BATH 782151 21882 35.744 < 2e-16 \*\*\*

---

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

Residual standard error: 2050000 on 4710 degrees of freedom

Multiple R-squared: 0.2134, Adjusted R-squared: 0.2132

F-statistic: 1278 on 1 and 4710 DF, p-value: < 2.2e-16

A graph with a line going up

AI-generated content may be incorrect.

A graph of a graph showing a number of values

AI-generated content may be incorrect.

**Price vs. Baths:**

> summary(lin.mod.BATH)

Call:

lm(formula = NY.house.data.subset.BATH$PRICE ~ NY.house.data.subset.BATH$BATH,

data = NY.house.data.subset.BATH)

Residuals:

Min 1Q Median 3Q Max

-5793930 -720876 -293126 73275 18393709

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -255426 57602 -4.434 9.45e-06 \*\*\*

NY.house.data.subset.BATH$BATH 782151 21882 35.744 < 2e-16 \*\*\*

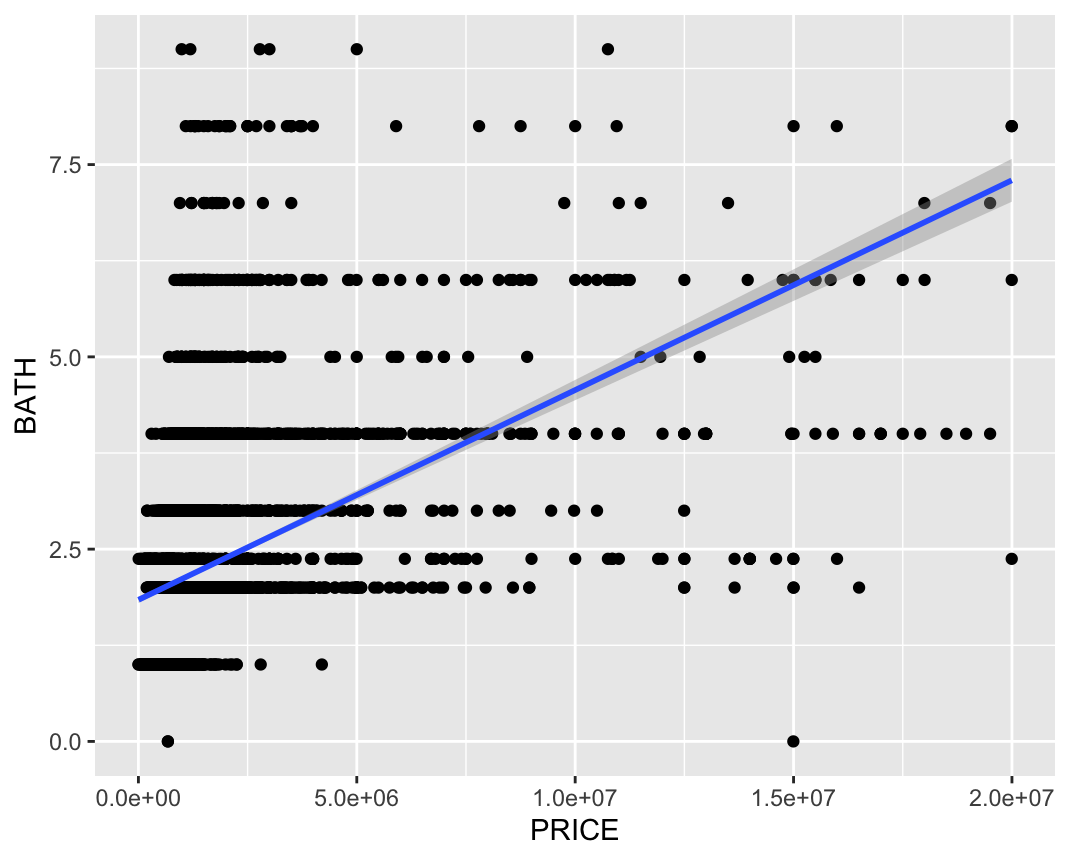
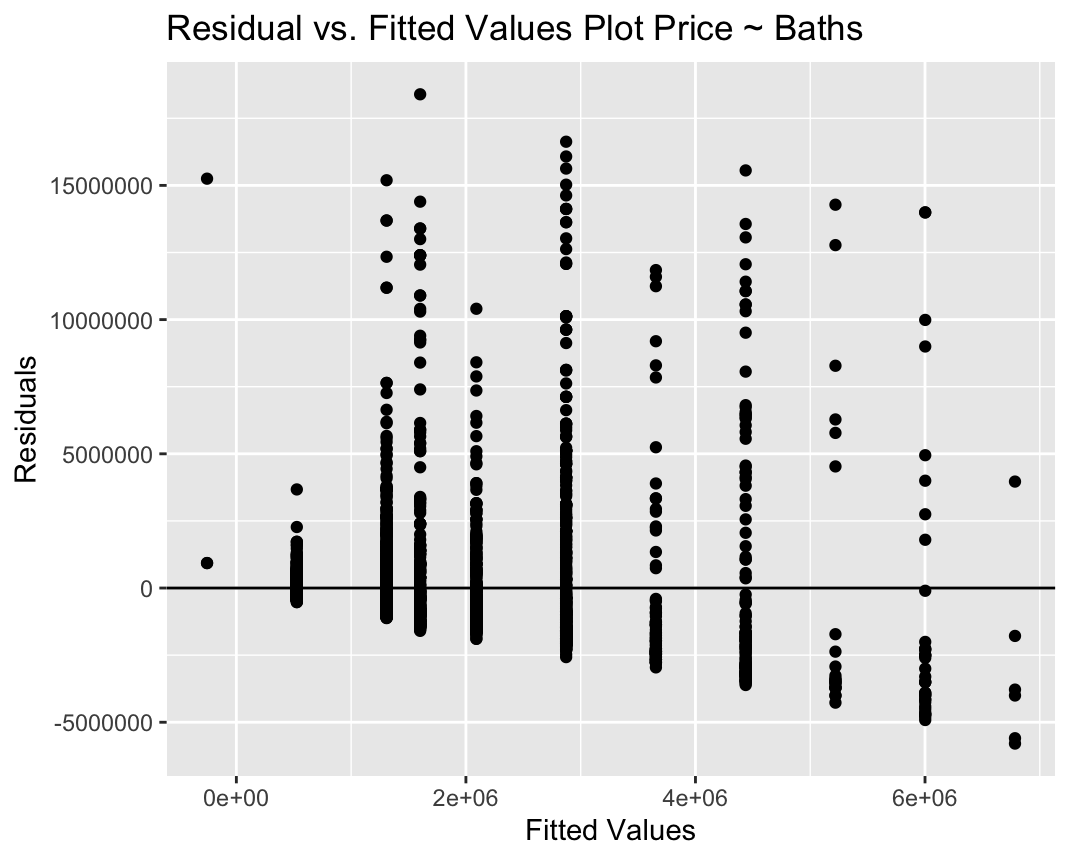
---

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

Residual standard error: 2050000 on 4710 degrees of freedom

Multiple R-squared: 0.2134, Adjusted R-squared: 0.2132

F-statistic: 1278 on 1 and 4710 DF, p-value: < 2.2e-16

**Takeaway:**

From the 3 models represented here, the most significant Price-correlated variable is Square Feet, specifically log(price) vs log(SqFt). Each model uses a subset of the total data where several outliers or discrepancies are removed: namely, prices over 20 million and the large number of houses with a reported square footage of exactly 2184.208.