Model 1: log(PRICE) ~ log(PROPERTYSQFT) + BEDS + BATH

This model explains about 58.9% of the variance in housing prices (Adjusted $R^2 = 0.5884$).

- Property size (log-transformed) is strongly and positively associated with price ($\beta = 1.081$, $p < 0.001$).
- Beds show a small but negative effect on price ($\beta = -- 0.050$, $p < 0.001$).
- Baths have a positive effect on price ($\beta = 0.068$, $p < 0.001$).
  Residuals are fairly tight ($SE \approx 0.27$), and all predictors are statistically significant.

```
> summary(mod1)

Call:

lm(formula = log10(PRICE) ~ log10(PROPERTYSQFT) + BEDS + BATH,

    data = nydataset)

Residuals:

     Min       1Q   Median       3Q      Max

-0.99063 -0.18610 -0.03317  0.17488  0.97089

Coefficients:

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

(Intercept)          2.535254   0.075273   33.68   <2e-16 ***

log10(PROPERTYSQFT)  1.081017   0.026741   40.43   <2e-16 ***

BEDS                -0.049919   0.003742  -13.34   <2e-16 ***

BATH                 0.068009   0.005707   11.92   <2e-16 ***

---

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

Residual standard error: 0.2715 on 3120 degrees of freedom

Multiple R-squared:  0.5888,  Adjusted R-squared:  0.5884

F-statistic:  1489 on 3 and 3120 DF,  p-value: < 2.2e-16
```

Model 2: log(PRICE) ~ log(PROPERTYSQFT) + BEDS

This simpler model explains about 57.0% of the variance (Adjusted $R^2 = 0.5699$).

- Property size remains highly significant ($\beta = 1.240$, $p < 0.001$).
- Beds is again negative (β = -- 0.031, $p < 0.001$), but its effect size is smaller.
  The model fit is slightly weaker than Model 1, with a higher residual error (SE ≈ 0.28).

```
summary(mod2)

Call:

lm(formula = log10(PRICE) ~ log10(PROPERTYSQFT) + BEDS, data = nydataset)

Residuals:

     Min       1Q   Median       3Q      Max

-1.14827 -0.19379 -0.03883  0.18404  0.88413

Coefficients:

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

(Intercept)          2.124750   0.068423  31.053   <2e-16 ***

log10(PROPERTYSQFT)  1.240209   0.023682  52.369   <2e-16 ***

BEDS                -0.031000   0.003464  -8.949   <2e-16 ***

---

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

Residual standard error: 0.2775 on 3121 degrees of freedom

Multiple R-squared:  0.5701,  Adjusted R-squared:  0.5699

F-statistic:  2070 on 2 and 3121 DF,  p-value: < 2.2e-16
```

Model 3: log(PRICE) ~ BEDS + BATH

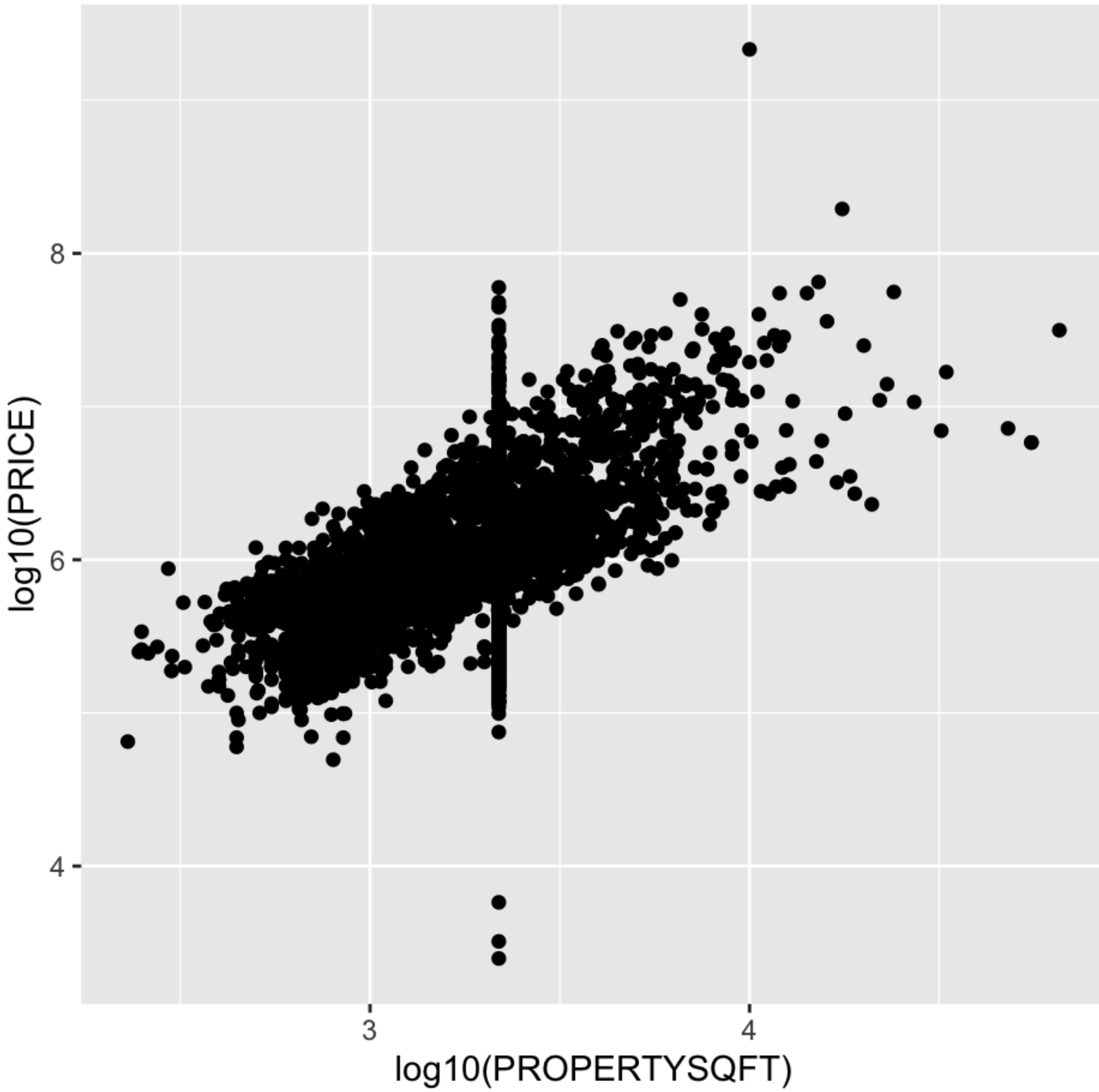
This model explains only 37.3% of the variance (Adjusted $R^2 = 0.3731$).

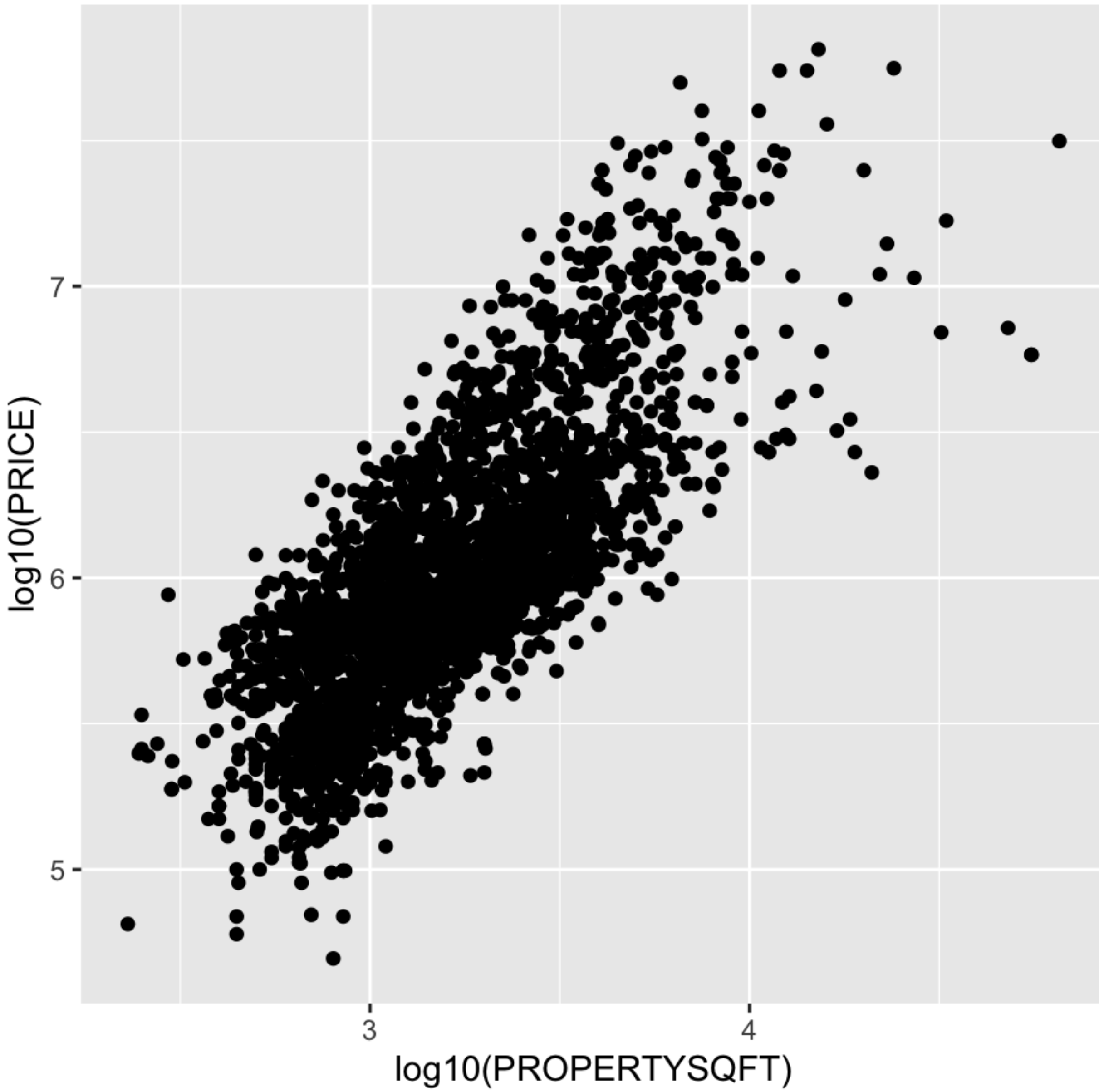
- Beds are not significant ($\beta \approx 0$, $p = 0.948$).
- Baths are strongly positive ($\beta = 0.183$, $p < 0.001$).
  Model 3 fits the data noticeably worse than the other two, with the largest residual error (SE ≈ 0.34).

```
> summary(mod3)

Call:

lm(formula = log10(PRICE) ~ BEDS + BATH, data = nydataset)

Residuals:

     Min       1Q   Median       3Q      Max

-1.29091 -0.21116 -0.04134  0.17962  1.39373

Coefficients:

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

(Intercept)  5.5513224  0.0123299 450.232   <2e-16 ***

BEDS        -0.0002832  0.0043627  -0.065    0.948

BATH         0.1832649  0.0061018  30.034   <2e-16 ***

---

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

Residual standard error: 0.3351 on 3121 degrees of freedom

Multiple R-squared:  0.3735,  Adjusted R-squared:  0.3731

F-statistic: 930.2 on 2 and 3121 DF,  p-value: < 2.2e-16
```

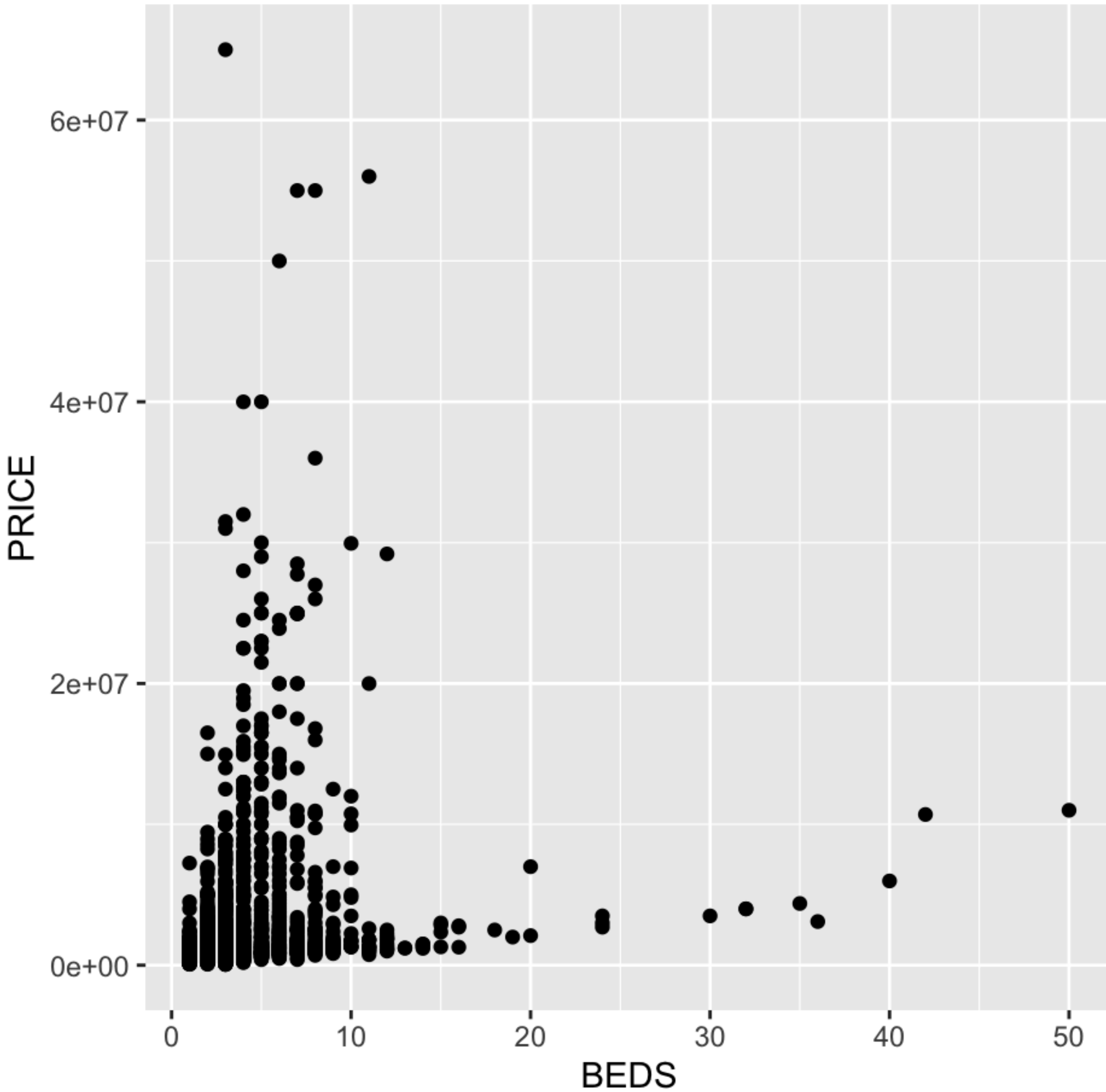
Comparison:

- Model 1 provides the best overall fit, balancing predictors and explaining the most variance.
- Model 2 is nearly as good but omits Baths, which appear to be an important predictor.
- Model 3 is weakest, as it excludes property size, which is clearly the dominant factor.

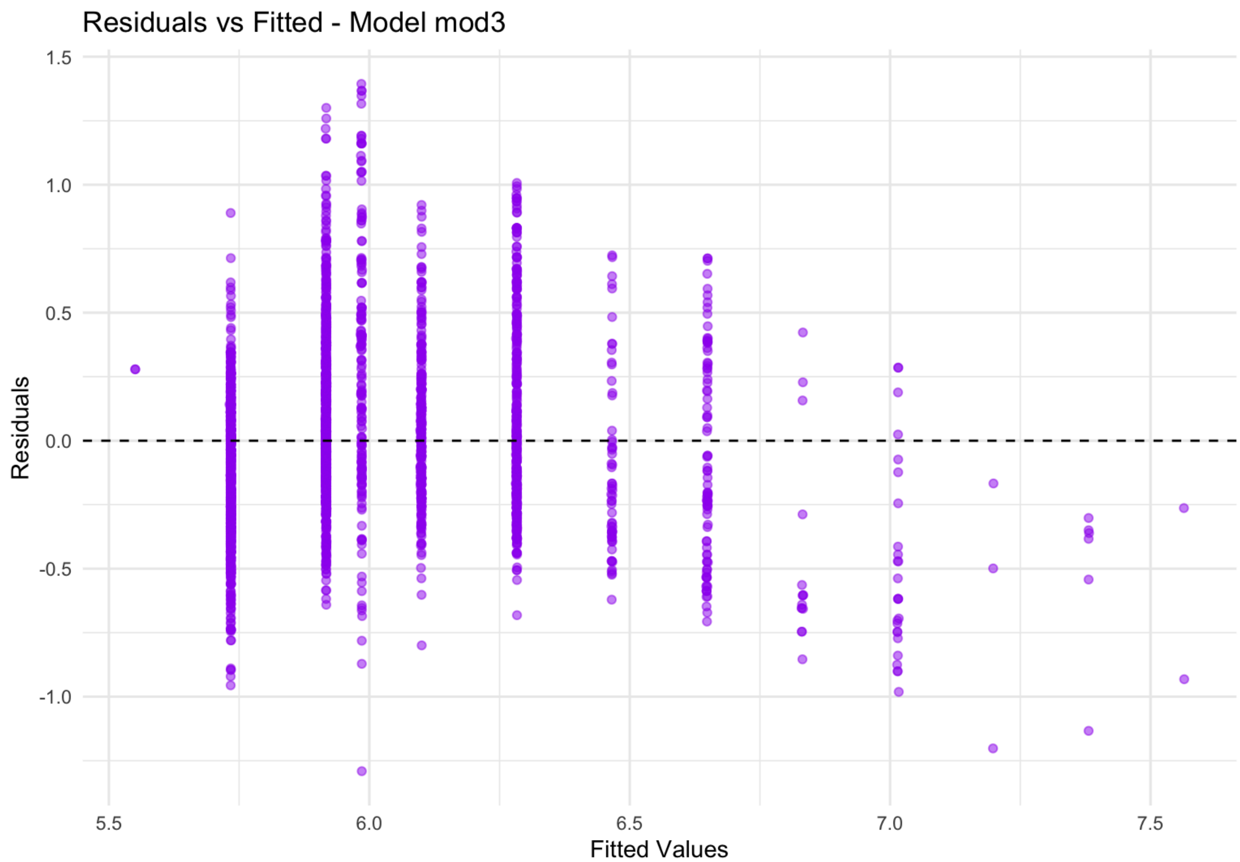

8
6
4
log10(PRICE)
3
4
log10(PROPERTYSQFT)

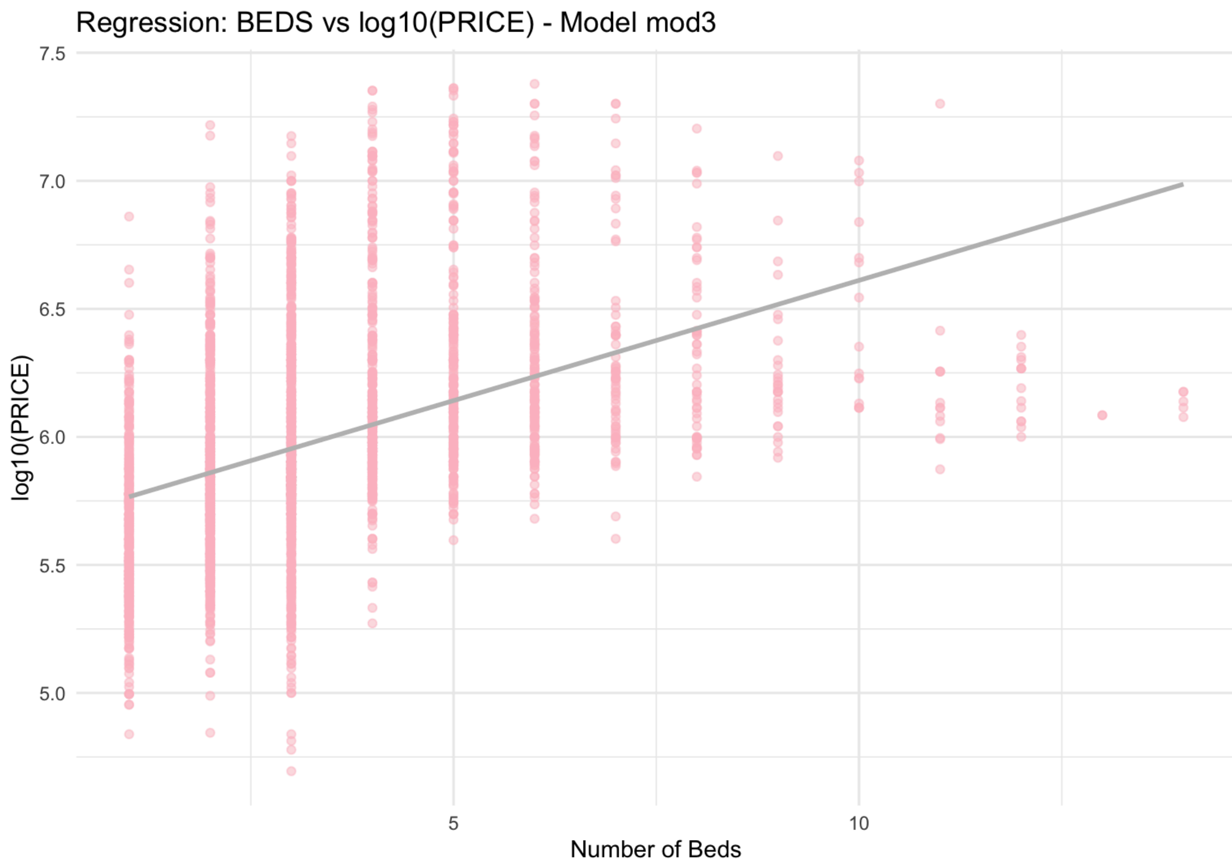


7
6
5
log10(PRICE)
3
4
log10(PROPERTYSQFT)

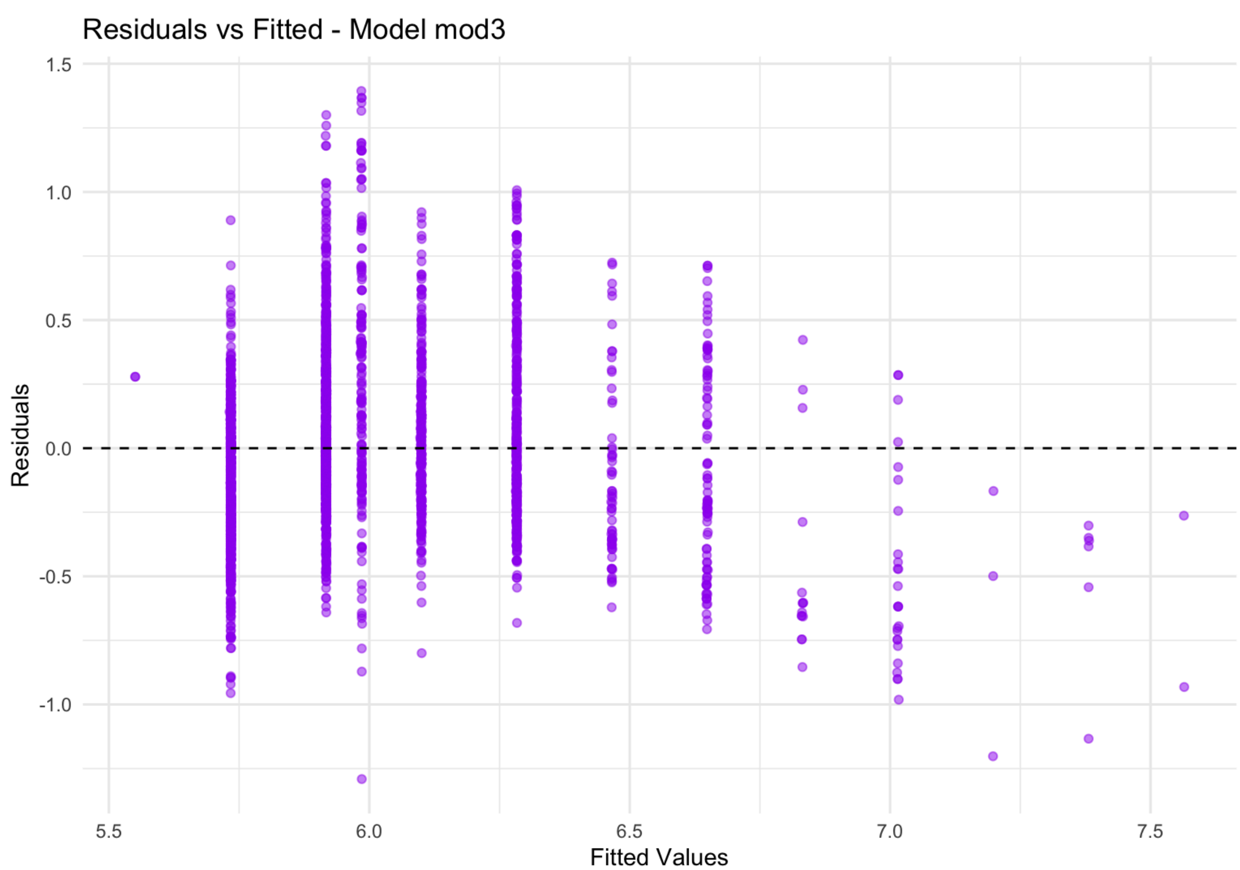


6e+07
4e+07
2e+07
0e+00
PRICE
0
10
20
30
40
50
BEDS


Residuals vs Fitted - Model mod3
Residuals
Fitted Values
1.5
1.0
0.5
0.0
-0.5
-1.0
5.5
6.0
6.5
7.0
7.5



Regression: BEDS vs log10(PRICE) - Model mod3
7.5
7.0
6.5
6.0
5.5
5.0
log10(PRICE)
5
10
Number of Beds



Residuals vs Fitted - Model mod3
1.5
1.0
0.5
0.0
-0.5
-1.0
Residuals
5.5
6.0
6.5
7.0
7.5
Fitted Values