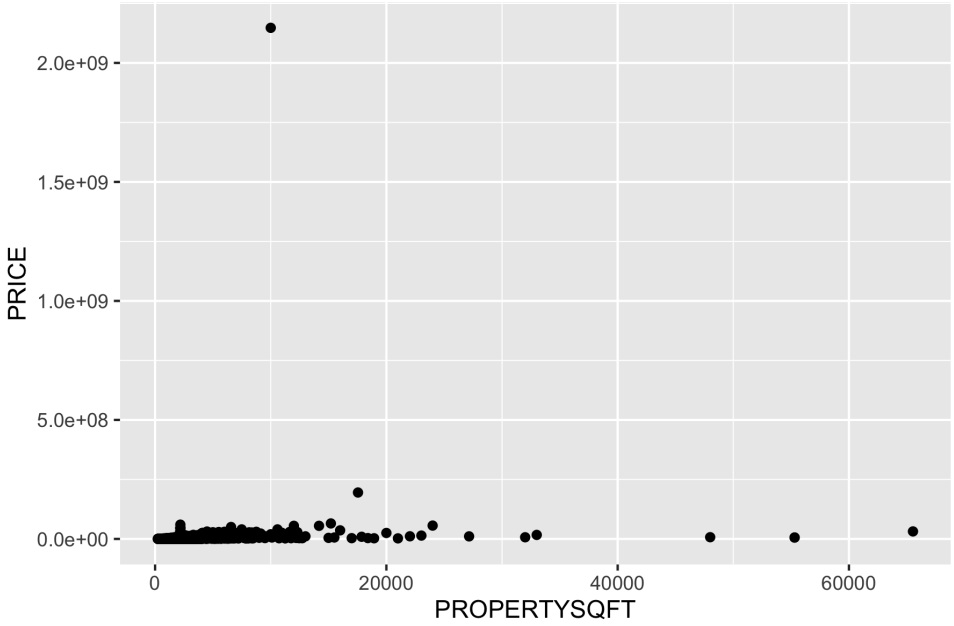
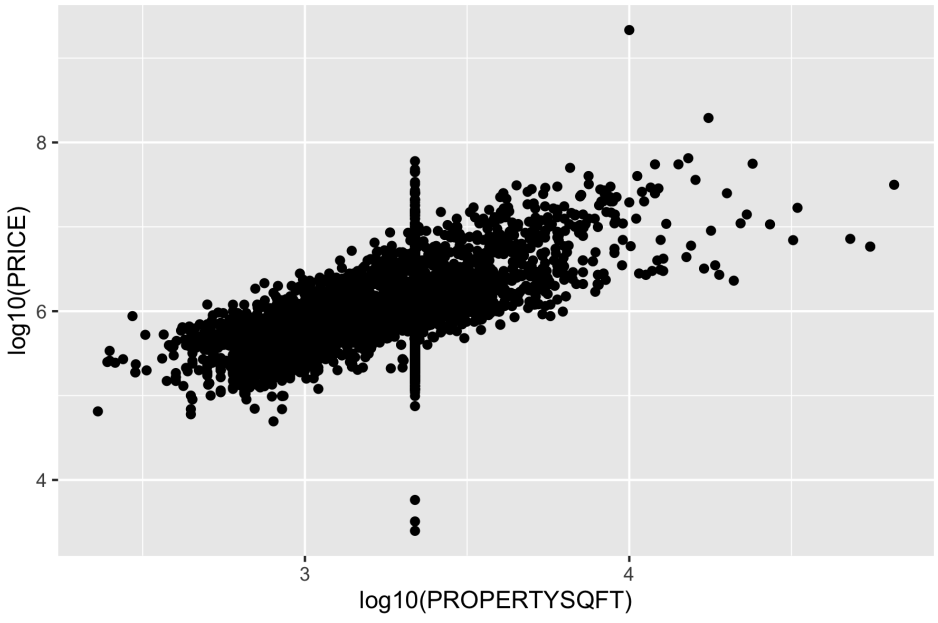
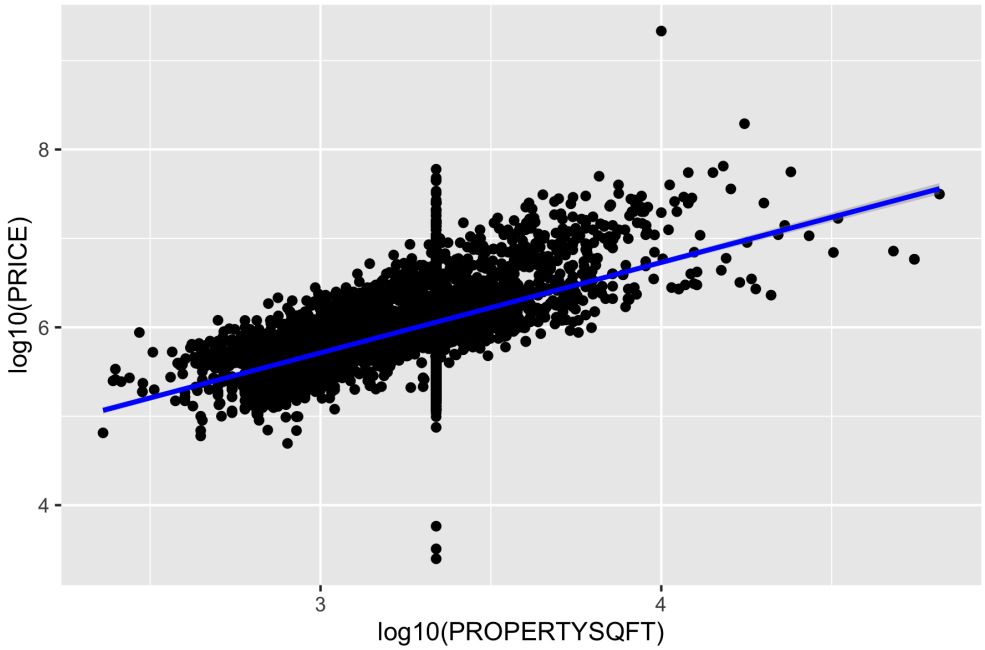
Support Vector Machine Classification

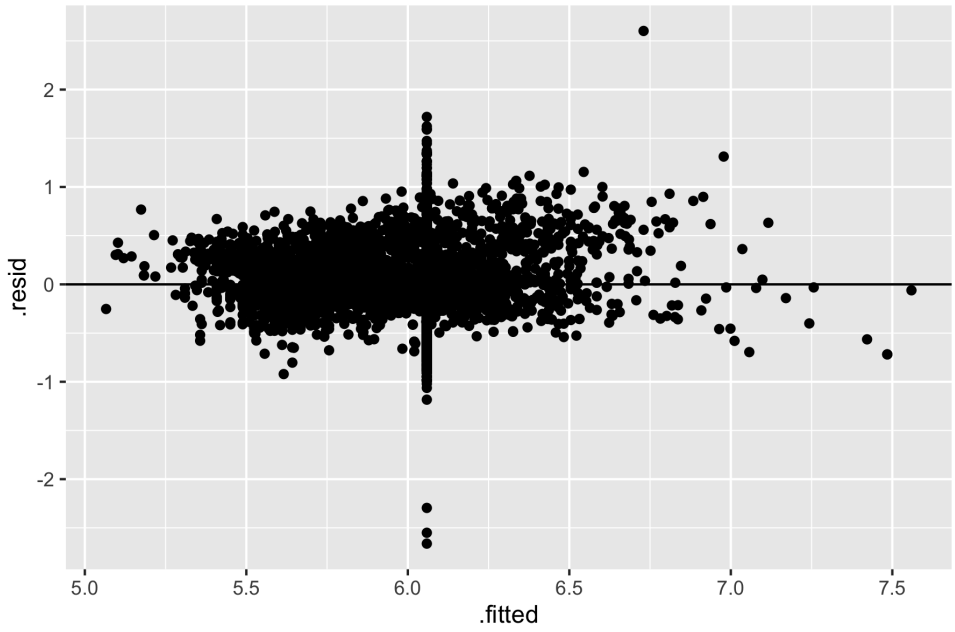
Using the NY House dataset:

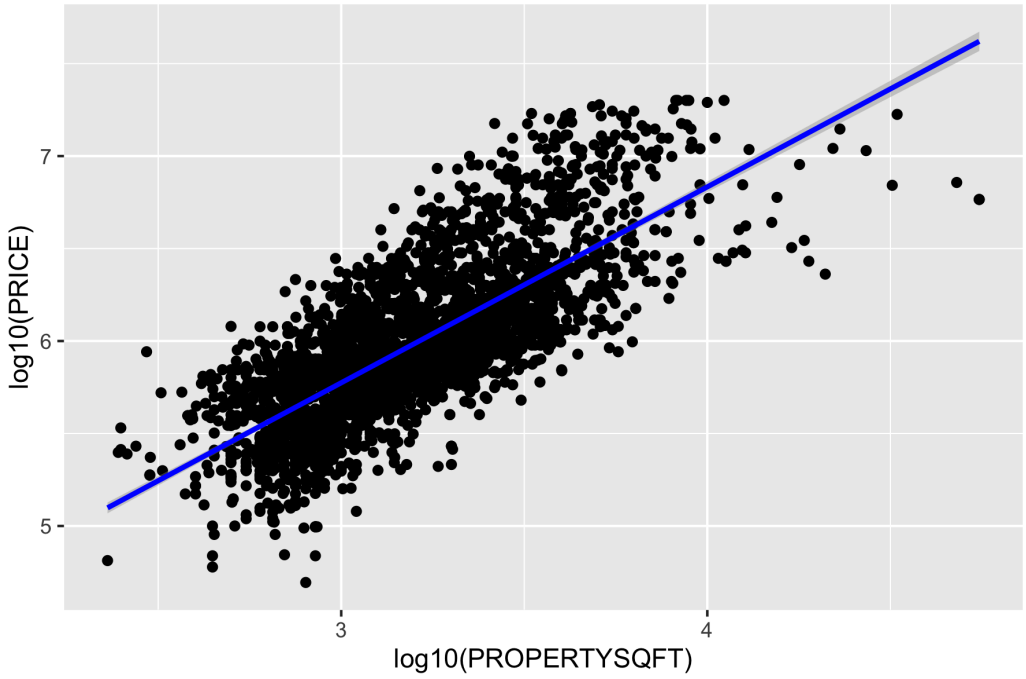
- Train 3 regression models each using a different algorithm to predict price from square footage. Evaluate model performance using the MAE, MSE and RMSE metrics.

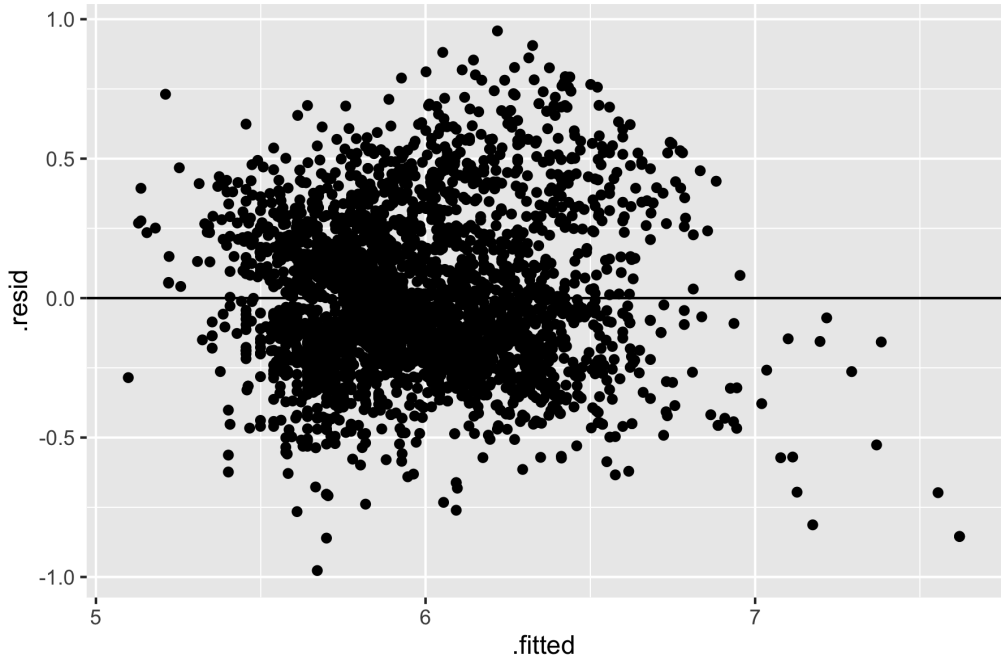




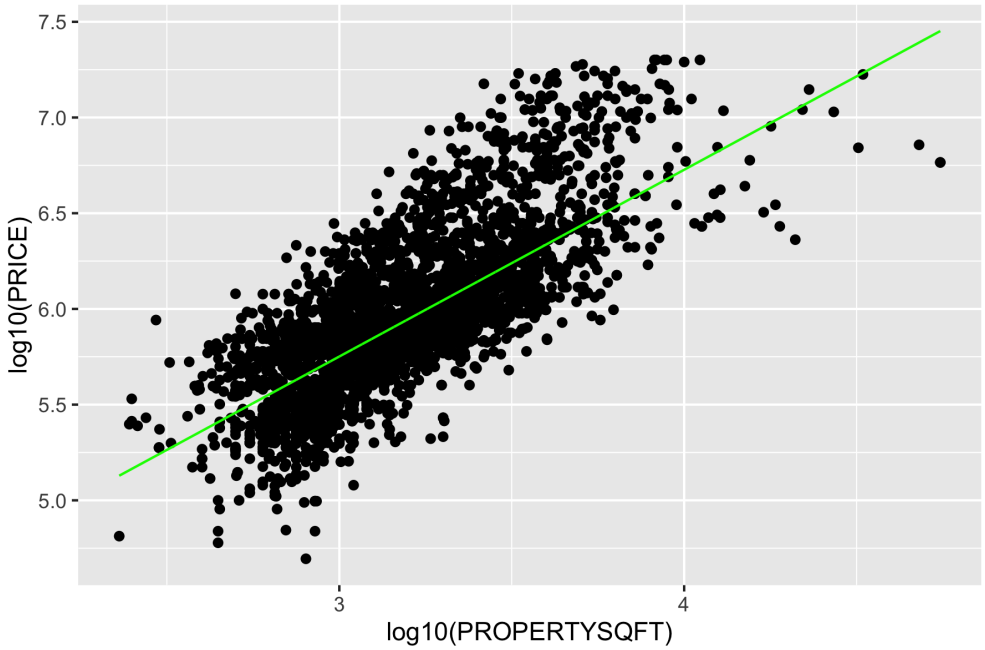

log10(PRICE)
8
6
4
3
4
log10(PROPERTYSQFT)

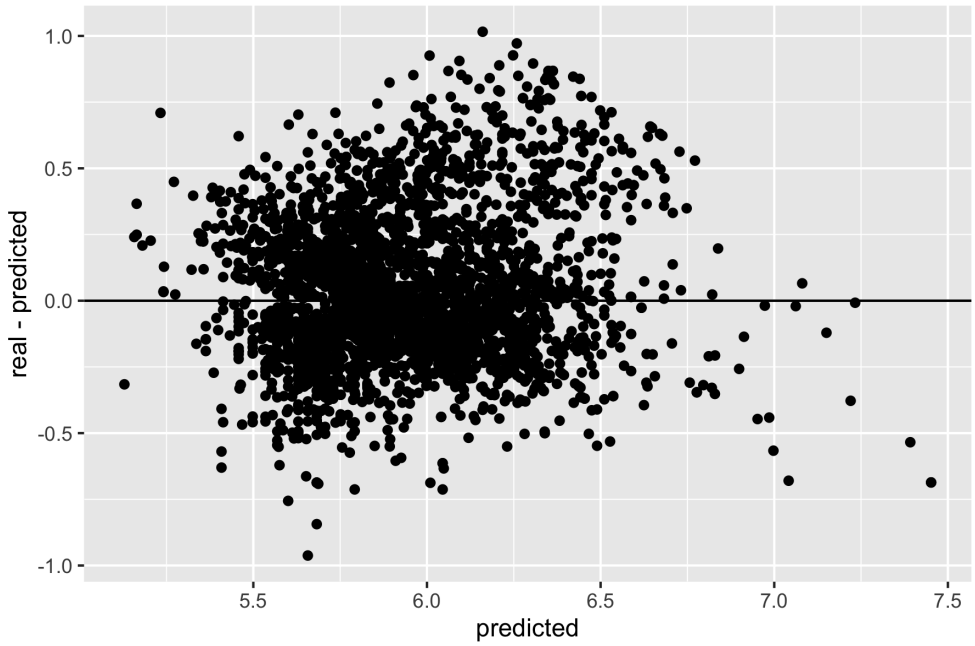

.resid
2
1
0
-1
-2
5.0
5.5
6.0
6.5
7.0
7.5
.fitted

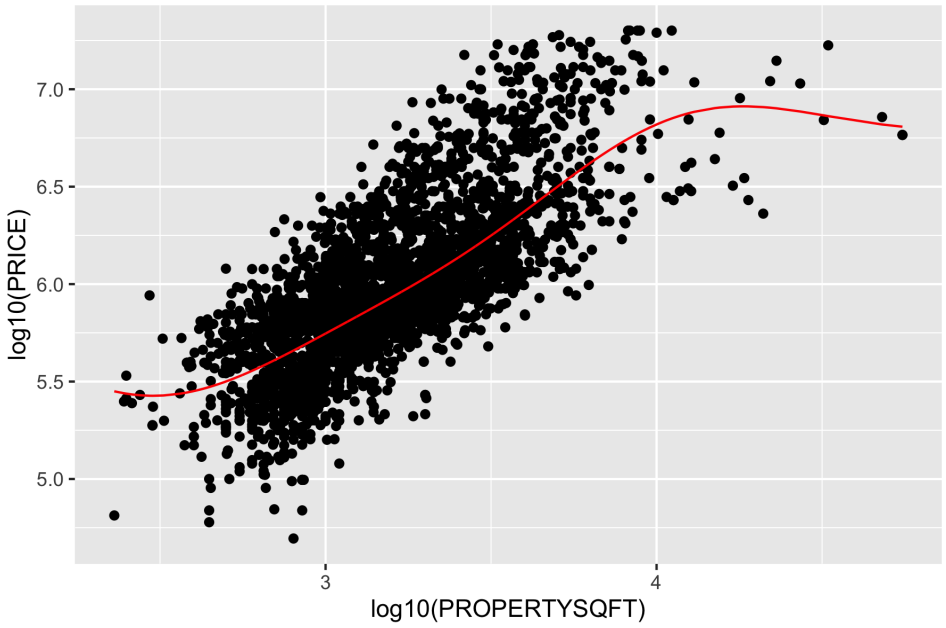

log10(PRICE)
7
6
5
3
4
log10(PROPERTYSQFT)

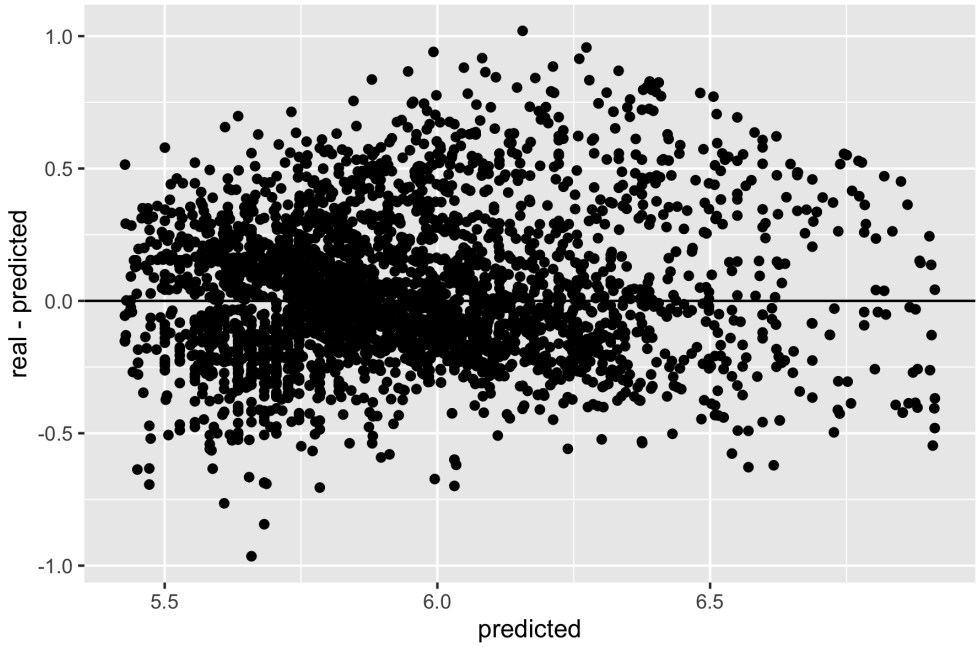

.resid
1.0
0.5
0.0
-0.5
-1.0
5
6
7
.fitted

```
> summary(lin.mod0)

Call:
lm(formula = log10(PRICE) ~ log10(PROPERTYSQFT), data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-2.66188 -0.20813 -0.02737  0.19770  2.60205

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.66695    0.06625   40.26   <2e-16 ***
log10(PROPERTYSQFT) 1.01573    0.02031   50.00   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3627 on 4799 degrees of freedom
Multiple R-squared:  0.3425,    Adjusted R-squared:  0.3424
F-statistic:  2500 on 1 and 4799 DF,  p-value: < 2.2e-16

> summary(lin.mod1)

Call:
lm(formula = log10(PRICE) ~ log10(PROPERTYSQFT), data = dataset.sub0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97692 -0.19308 -0.04588  0.18168  0.95796

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.59643    0.05408   48.01   <2e-16 ***
log10(PROPERTYSQFT) 1.05925    0.01684   62.89   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2799 on 3140 degrees of freedom
Multiple R-squared:  0.5574,    Adjusted R-squared:  0.5573
F-statistic:  3955 on 1 and 3140 DF,  p-value: < 2.2e-16
```


7.5
7.0
6.5
6.0
5.5
5.0
log10(PRICE)
3
4
log10(PROPERTYSQFT)


1.0
0.5
0.0
-0.5
-1.0
real - predicted
5.5
6.0
6.5
7.0
7.5
predicted






```
> summary(cv.res)
10-Fold Cross Validation
method: Woodbury
criterion: mse
cross-validation criterion = 0.07884688
bias-adjusted cross-validation criterion = 0.07883774
95% CI for bias-adjusted CV criterion = (0.07424872, 0.08342675)
full-sample criterion = 0.07867297
```

```
> print(results)
                           Model       MAE        MSE      RMSE
1              Linear Regression 0.2235978 0.07793381 0.2790709
2                     Linear SVM 0.2205650 0.08021474 0.2831010
3 RBF SVM (tuned, global params) 0.2185808 0.07861065 0.2802605
```

Model performance was assessed using MAE, MSE, and RMSE, which capture average absolute deviation, squared error, and its root, respectively. These metrics quantify both bias and variance in the predictions, providing a clear measure of overall accuracy. Consistent with the regression analysis, the tuned RBF SVM achieved the lowest errors, indicating a marginal but measurable improvement in predictive performance.