Train Ensemble Model with Multiple Train/Test Splits

Anna Shcherbina

Dec 7, 2018

Load Feature Matrix

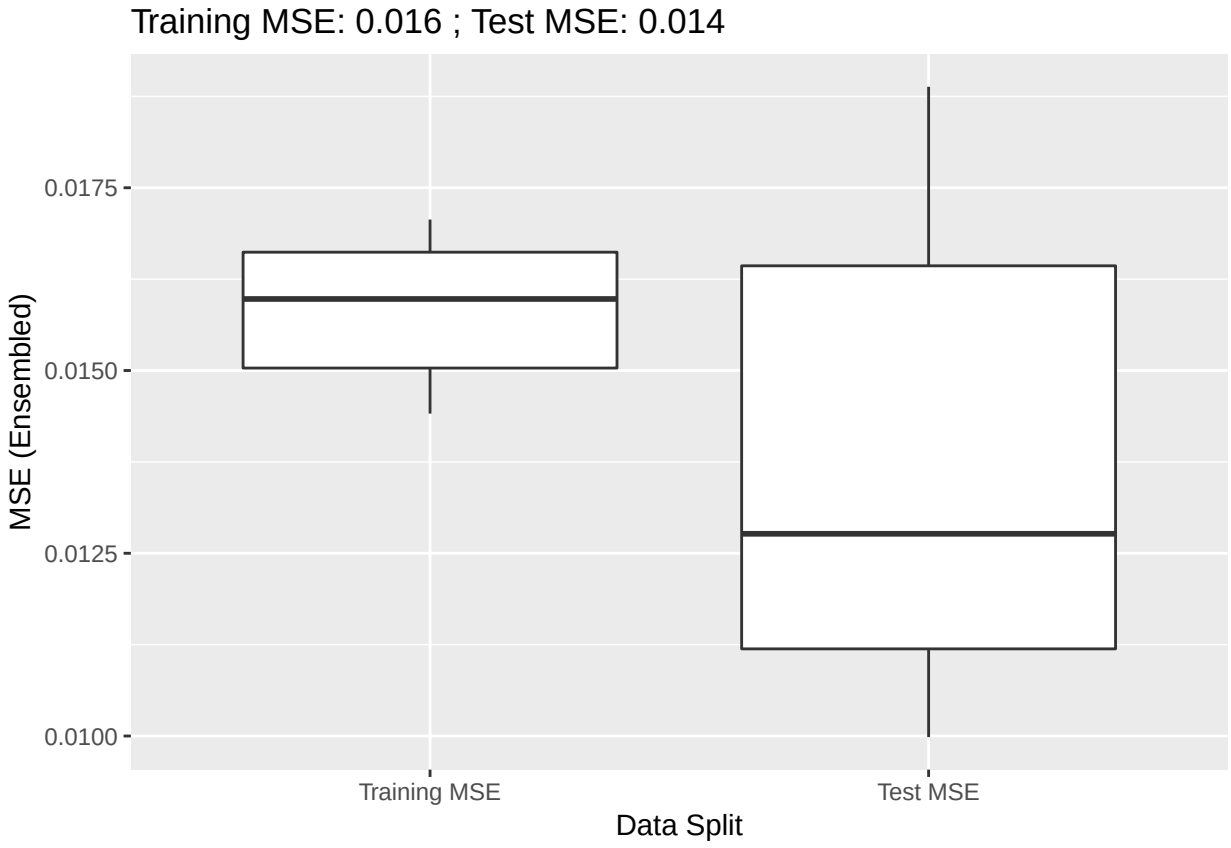
```
## [1] "seed: 1234"
```

Train and Test Splits

```
## [1] "n_iter: 10"
## [1] "train_split_size: 0.8"
## [1] "test_split_size: 0.2"
```

Quantify the MSE

```
## [1] "Training MSE (Ensembled): 0.0158356435033462"
## [1] "Test MSE (Ensembled): 0.0137324203760116"
## No id variables; using all as measure variables
```



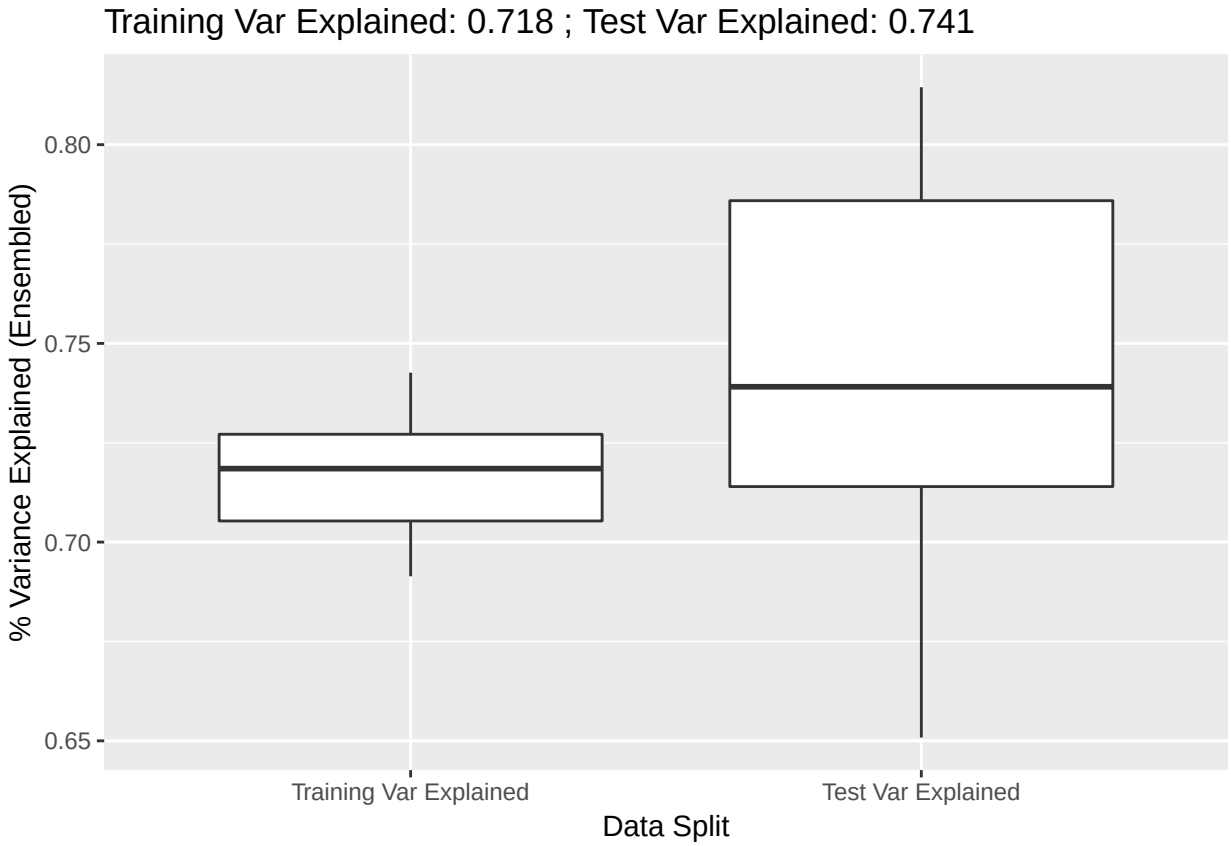
```
## pdf
##   2
```

Quantify % of Variance Explained in the Data by the Random Forest

```
## [1] "Training Var Explained (Ensembled): 0.717767287586099"
```

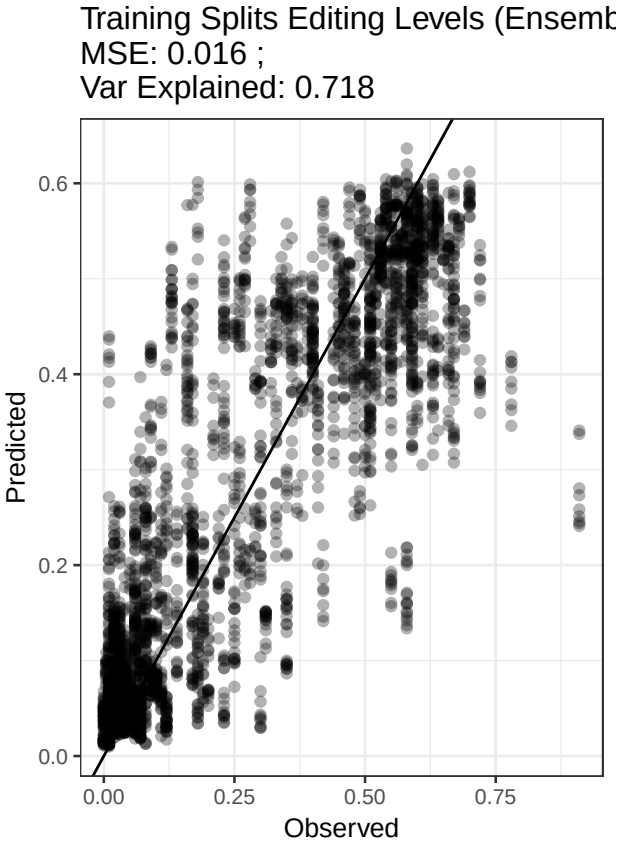
```
## [1] "Test Var Explained (Ensembled): 0.740963410310106"
```

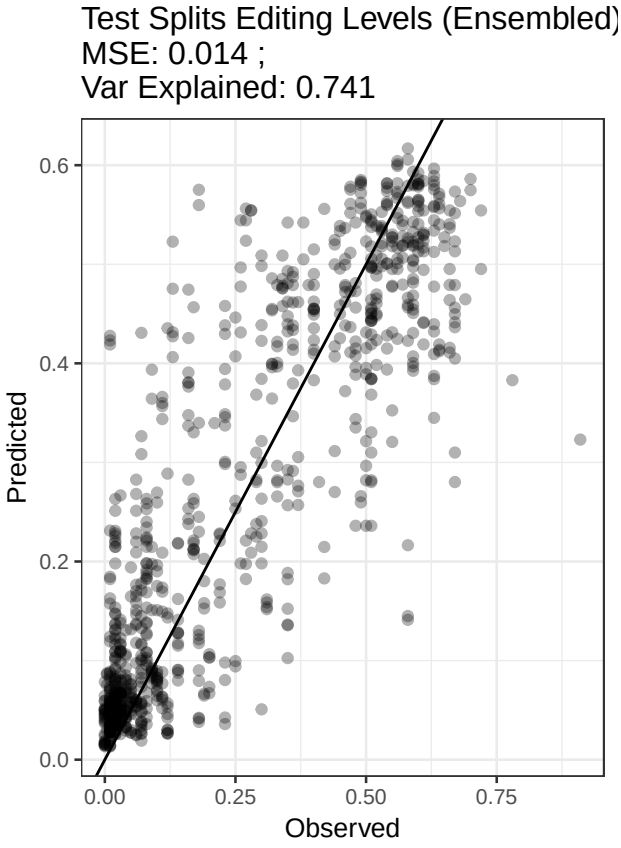
```
## No id variables; using all as measure variables
```




```
## pdf
##   2
```

Train and Test Split Predictions






```
## pdf
##   2
```

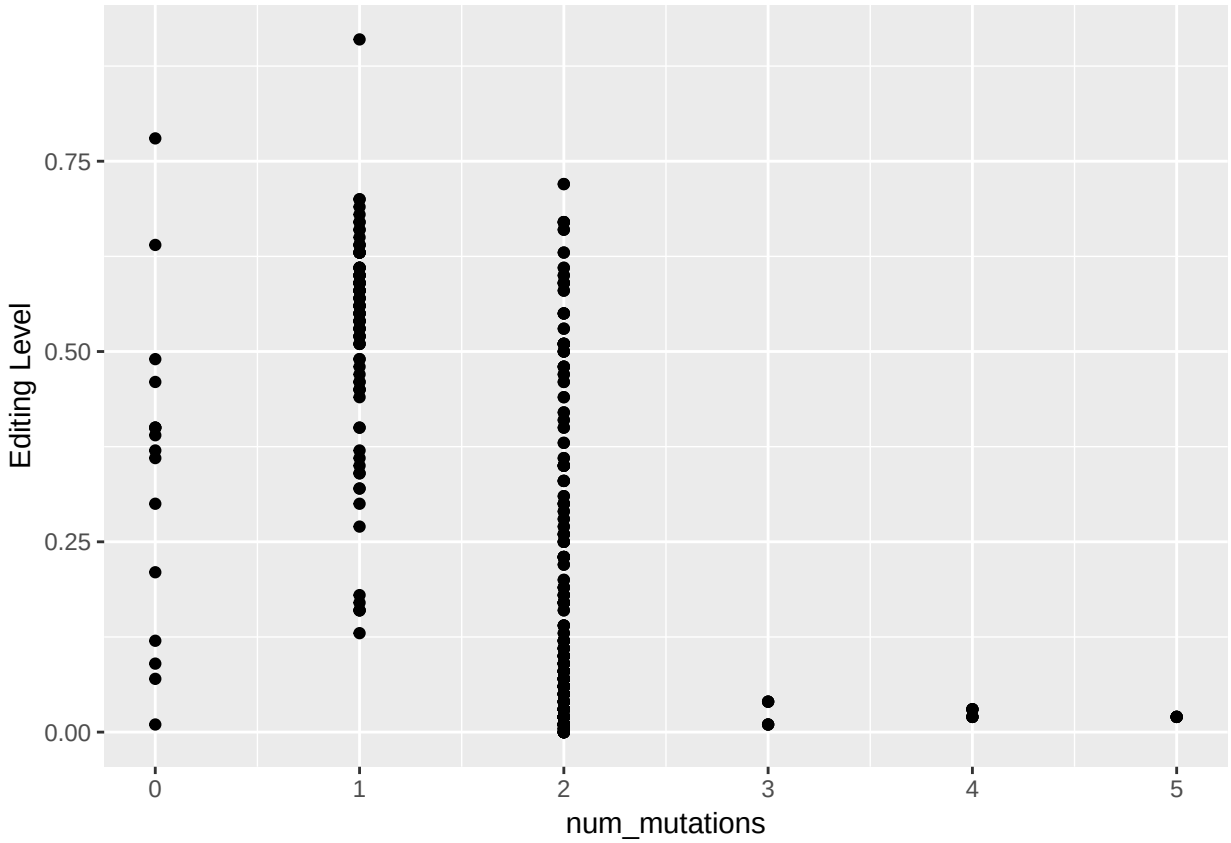
```
## pdf
##   2
```

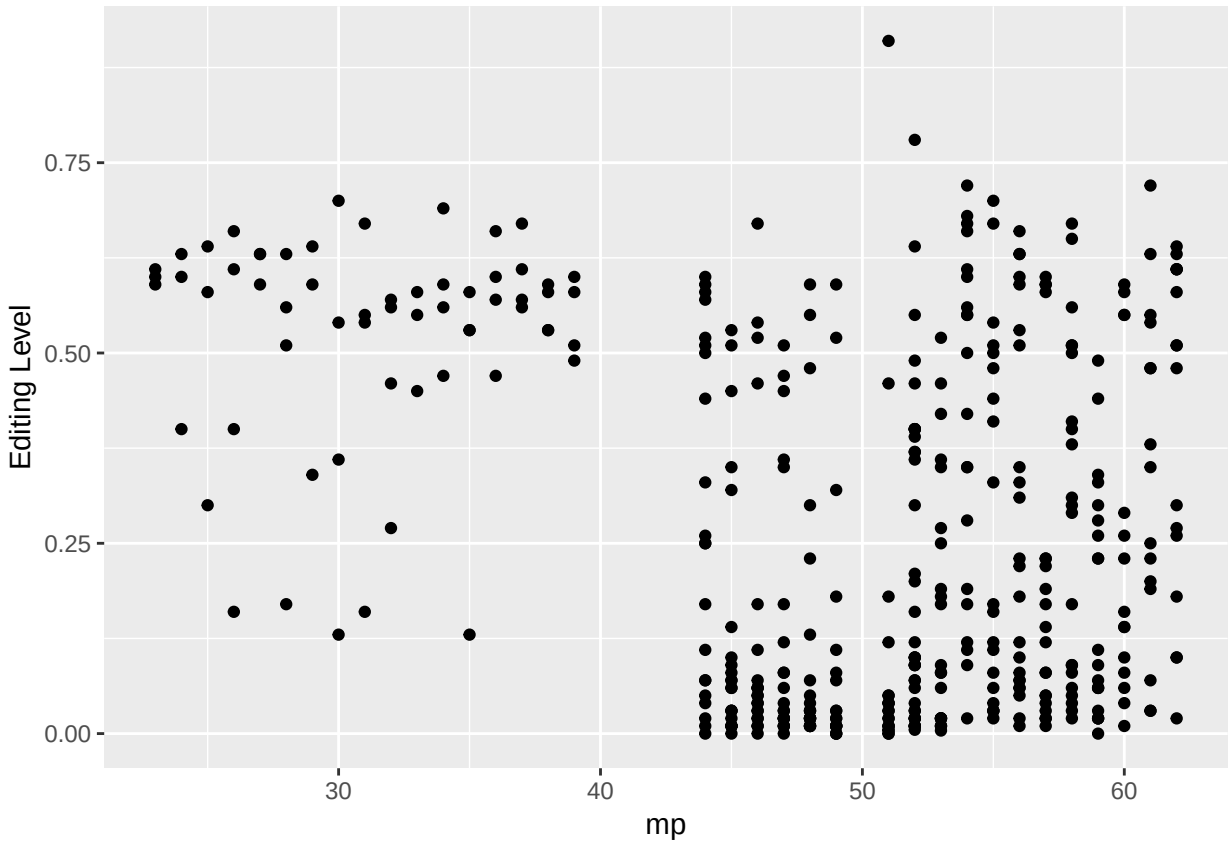
Merge the forests into one ensembl to calculate feature importance

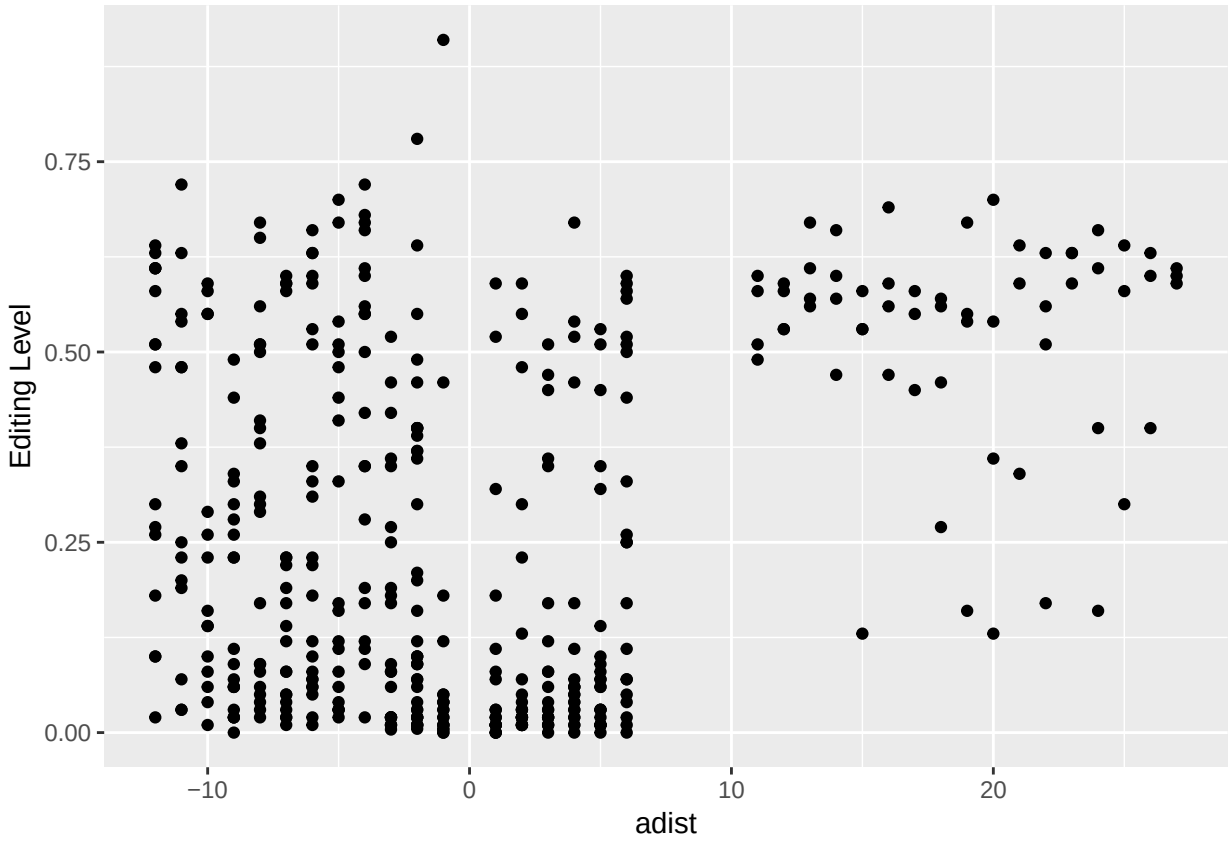
```
##                                                              %IncMSE
## num_mutations                                          0.0268795977
## mp                                                     0.0051791178
## adist                                                  0.0052204917
## mref                                                   0.0022017862
## malt                                                   0.0040954907
## mtype                                                  0.0001030912
## mfeat                                                  0.0010103983
## mfeat_prev                                             0.0006105742
## mfeat_next                                             0.0005266816
## mfeat_same_as_edit                                     0.0002177798
## editing_feature                                        0.0002367811
## stem_length                                            0.0079852250
## hairpin_length                                         0.0016736241
## x1feat_downstream_of_edit_site                         0.0001352330
## x1feat_downstream_of_edit_site_length_editing_strand   0.0134234938
## x1feat_downstream_of_edit_site_length_complementary_strand 0.0017901118
```

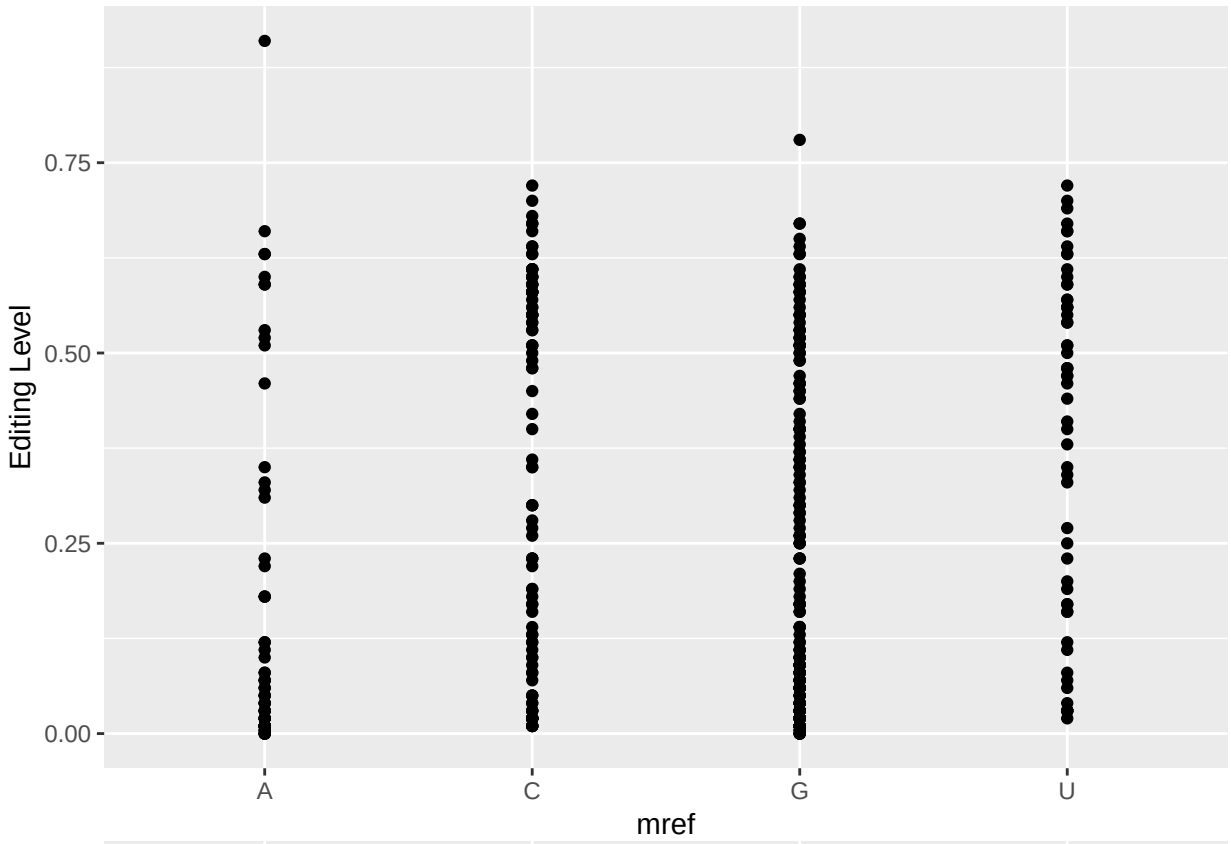
```
## x1feat_downstream_of_edit_site_5prime_cp                   0.0015537742
## x1feat_downstream_of_edit_site_3prime_cp                   0.0001438538
## x2feat_downstream_of_edit_site                             0.0001487270
## x2feat_downstream_of_edit_site_length_editing_strand       0.0135576936
## x2feat_downstream_of_edit_site_length_complementary_strand 0.0018143886
## x2feat_downstream_of_edit_site_5prime_cp                   0.0014999067
## x2feat_downstream_of_edit_site_3prime_cp                   0.0001724570
##                                                            IncNodePurity
## num_mutations                                                 4.50546385
## mp                                                            1.28501564
## adist                                                         1.27019494
## mref                                                          0.57899832
## malt                                                          0.93268091
## mtype                                                         0.08644952
## mfeat                                                         0.24912604
## mfeat_prev                                                    0.19532639
## mfeat_next                                                    0.15619073
## mfeat_same_as_edit                                            0.09011713
## editing_feature                                               0.04970791
## stem_length                                                   1.46786610
## hairpin_length                                                0.44399852
## x1feat_downstream_of_edit_site                                0.02395636
## x1feat_downstream_of_edit_site_length_editing_strand          2.23942561
## x1feat_downstream_of_edit_site_length_complementary_strand    0.31623424
## x1feat_downstream_of_edit_site_5prime_cp                      0.45094820
## x1feat_downstream_of_edit_site_3prime_cp                      0.10136704
## x2feat_downstream_of_edit_site                                0.02948881
## x2feat_downstream_of_edit_site_length_editing_strand          2.21807203
## x2feat_downstream_of_edit_site_length_complementary_strand    0.30931141
## x2feat_downstream_of_edit_site_5prime_cp                      0.41972327
## x2feat_downstream_of_edit_site_3prime_cp                      0.10351837
```

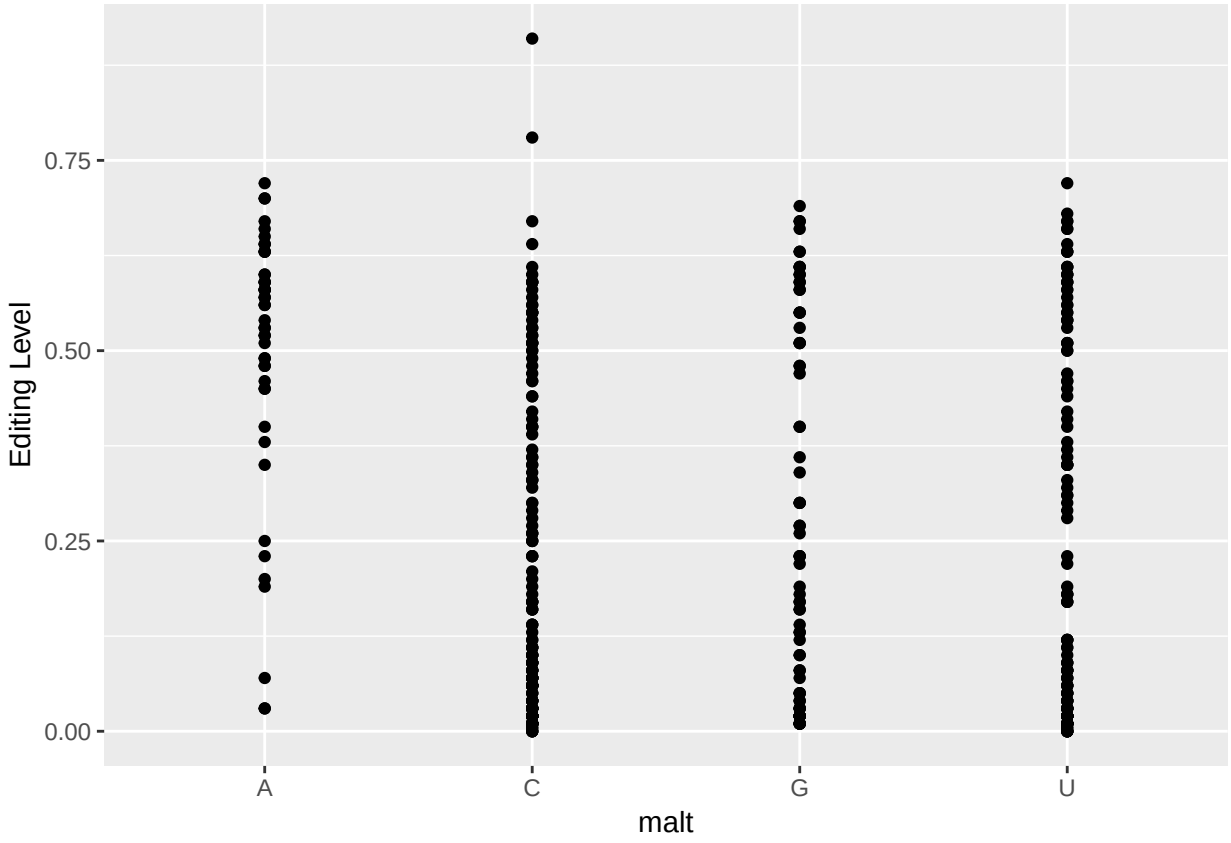
Generate Feature vs Editing Level plots for Top 10 Features

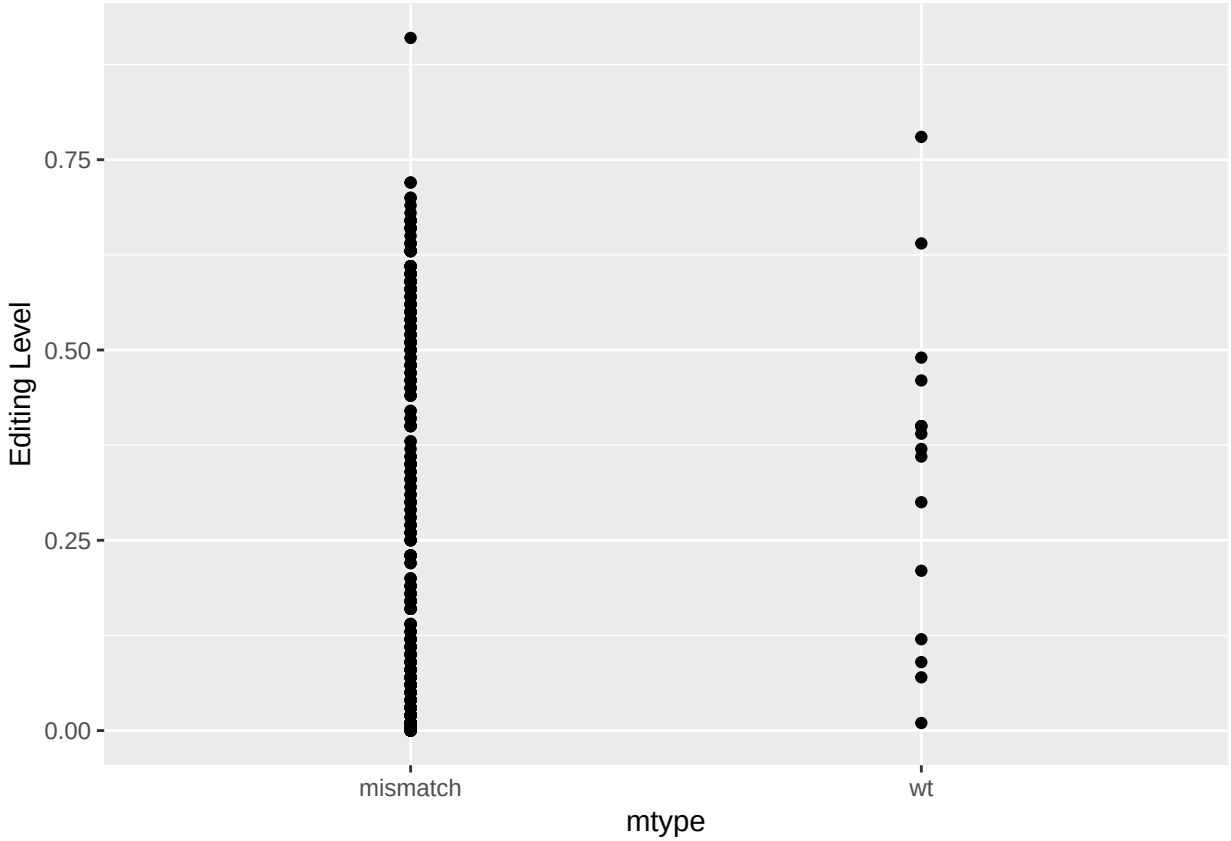


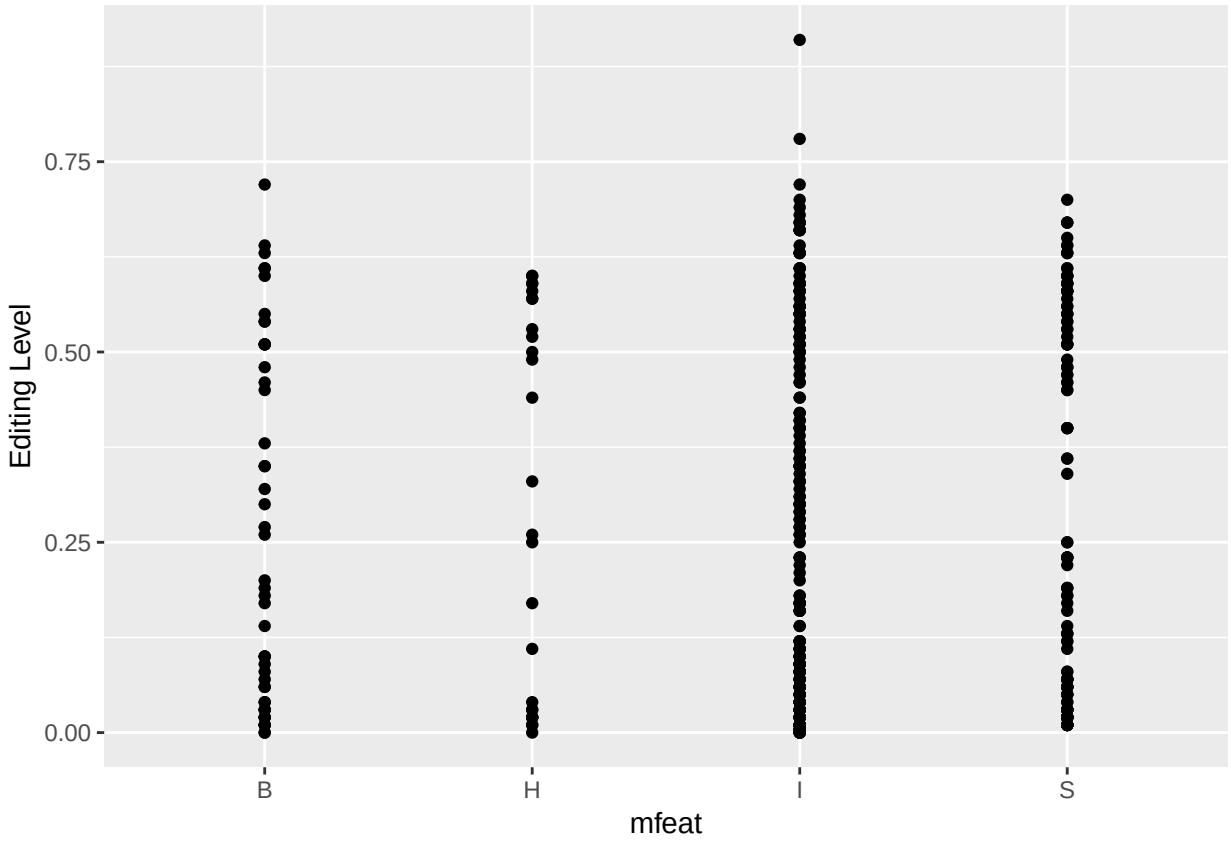

Editing Level
0.75
0.50
0.25
0.00
30
40
50
60
mp

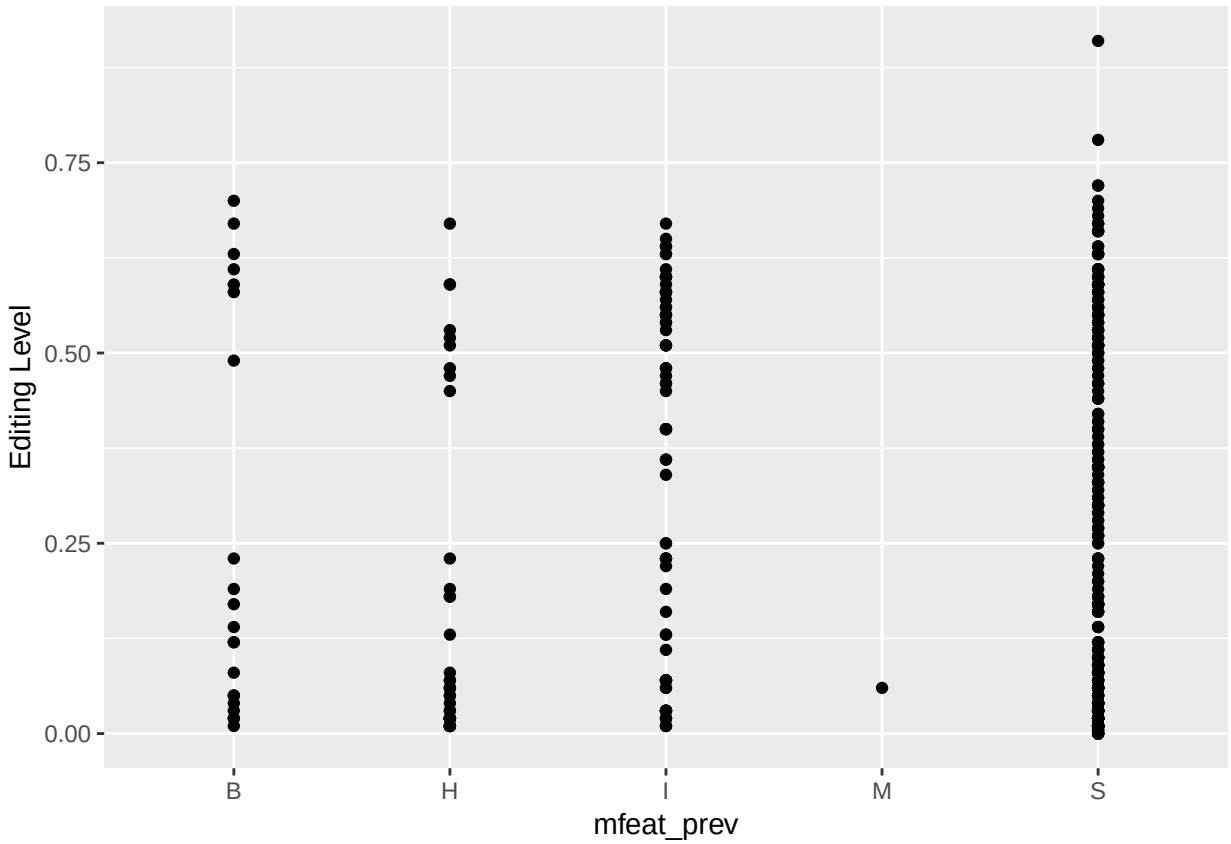

Editing Level
0.75
0.50
0.25
0.00
−10
0
10
20
adist

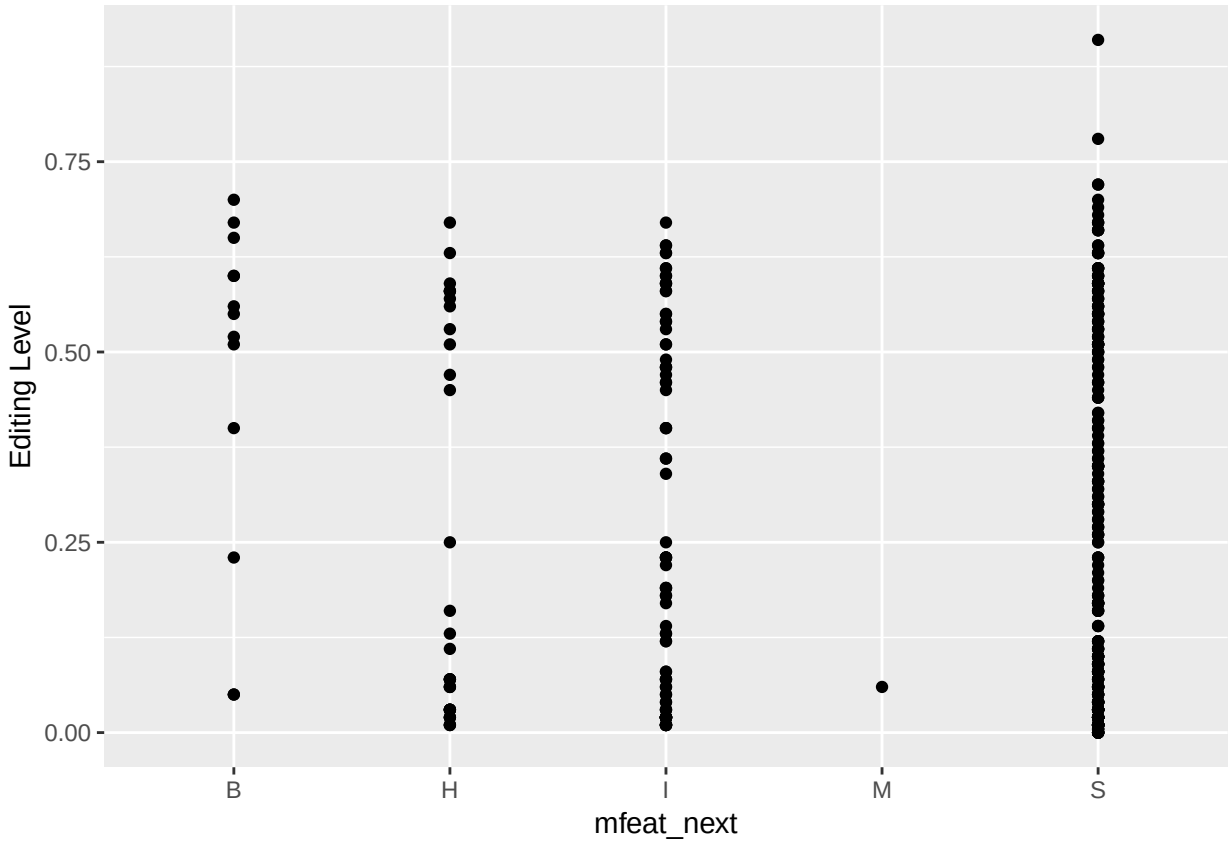

Editing Level
0.75
0.50
0.25
0.00
A
C
G
U
mref

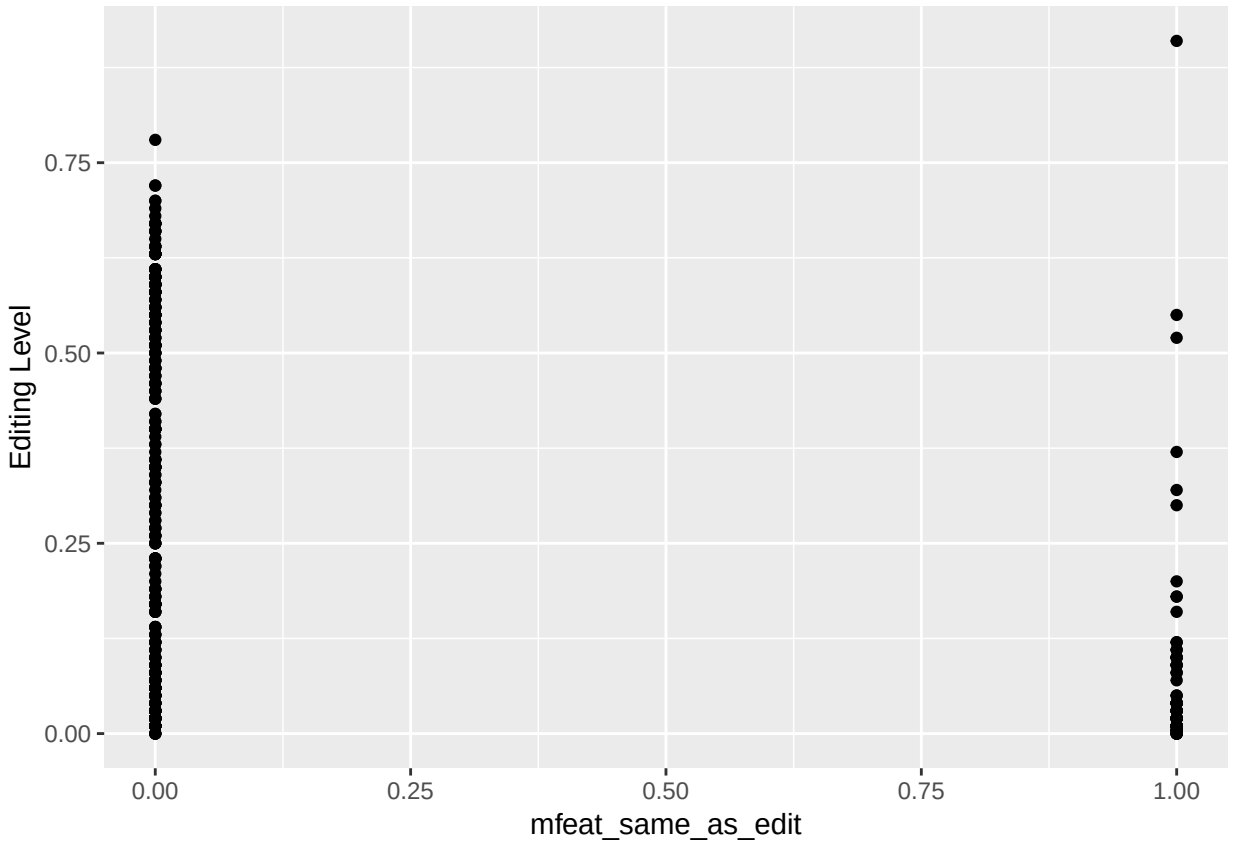

Editing Level
0.75
0.50
0.25
0.00
A
C
G
U
malt


Editing Level
0.75
0.50
0.25
0.00
mismatch
wt
mtype


Editing Level
0.75
0.50
0.25
0.00
B
H
I
S
mfeat


Editing Level
0.75
0.50
0.25
0.00
B
H
I
M
S
mfeat_prev


Editing Level
0.75
0.50
0.25
0.00
B
H
I
M
S
mfeat_next


Editing Level
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
mfeat_same_as_edit