Evaluate testing data (regression) - Lasso

EVE W.

2020-04-13

Contents

```
## user input
project_home <- "~/EVE/examples"
project_name <- "lasso_regression_outCV_test"
```

0. Load Data

```
## Error : $ operator is invalid for atomic vectors
```

```
## 300 of samples were used
```

```
## 100 of full features
```

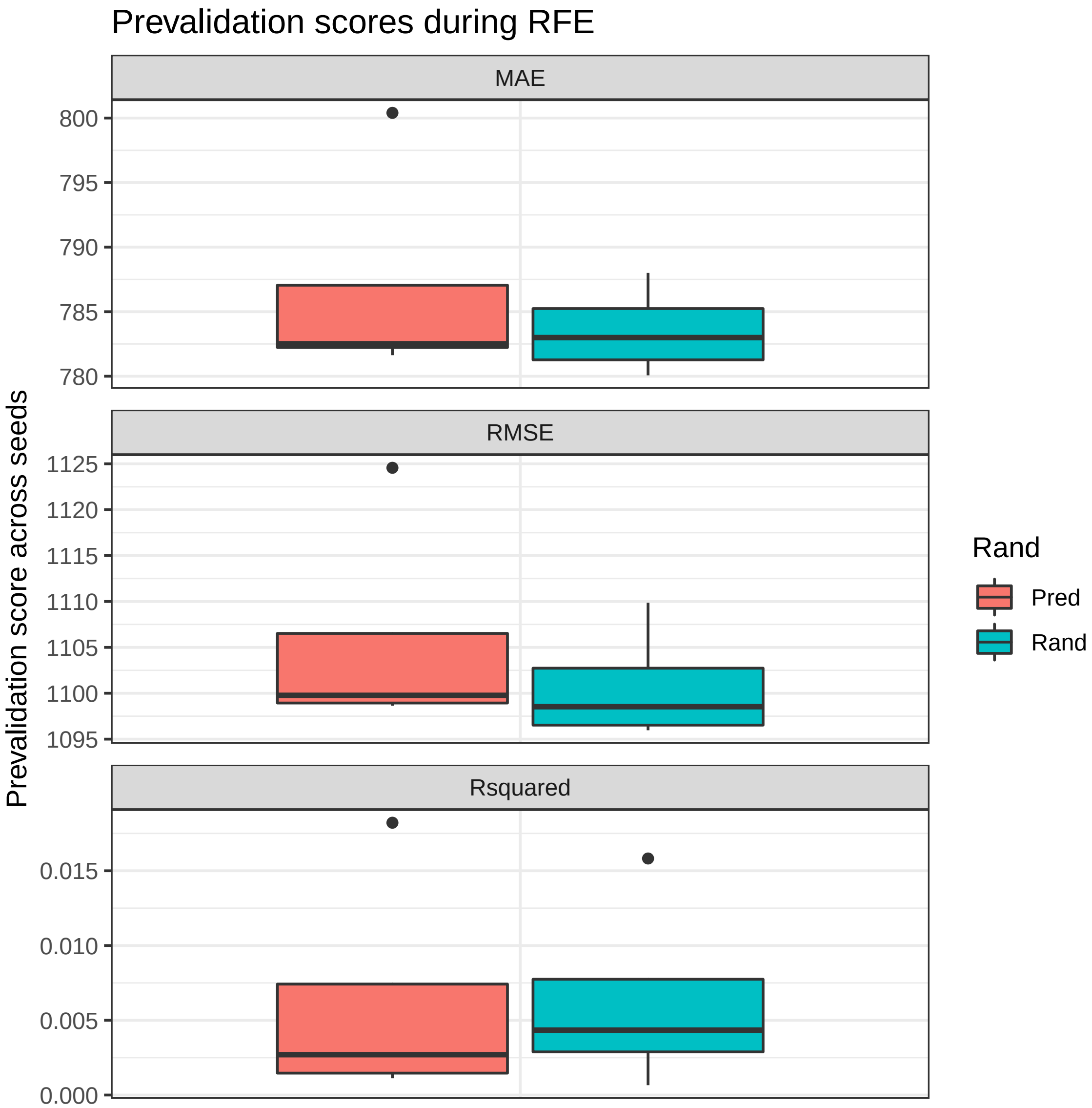
```
## 4 runs, each run contains 3 CVs.
```

```
## os_time :
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   182.8   480.0   889.4  1221.2  7125.0
```

run with lasso.r.

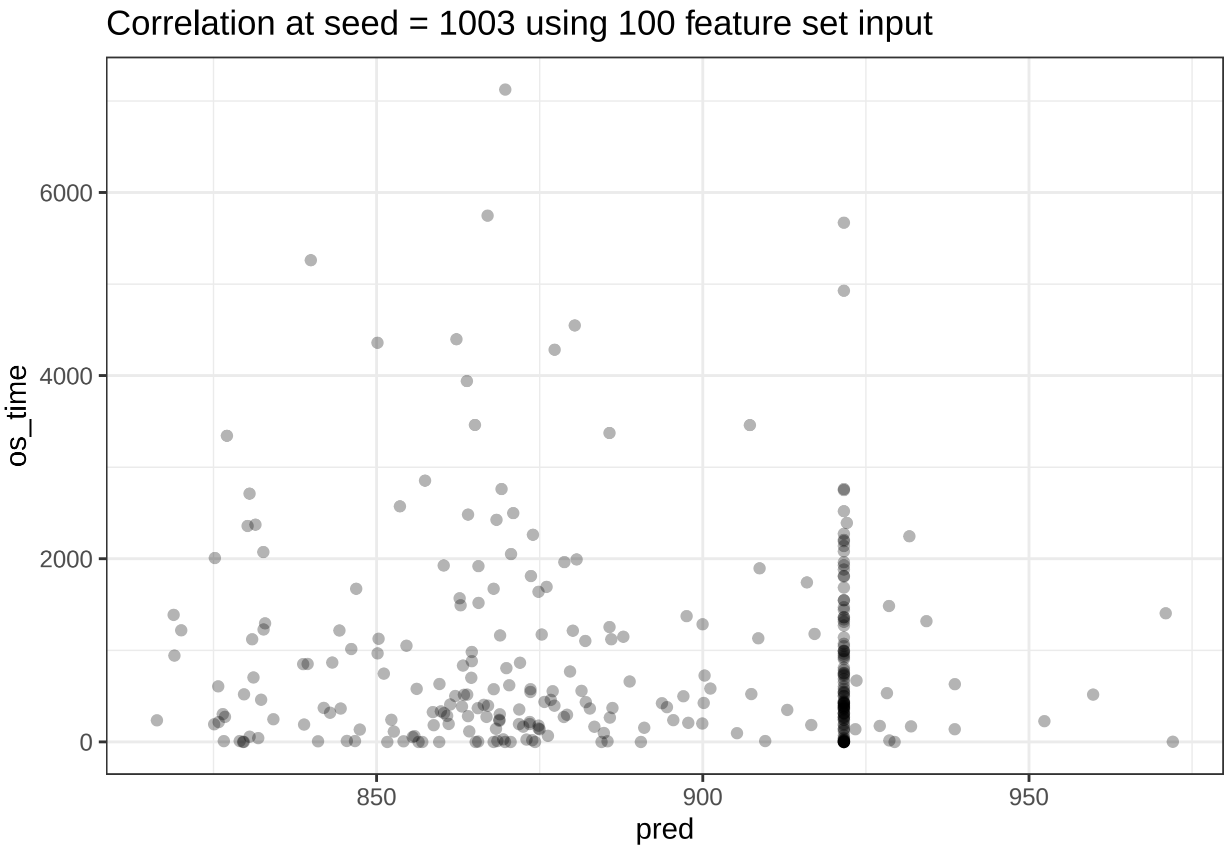
1. Scores



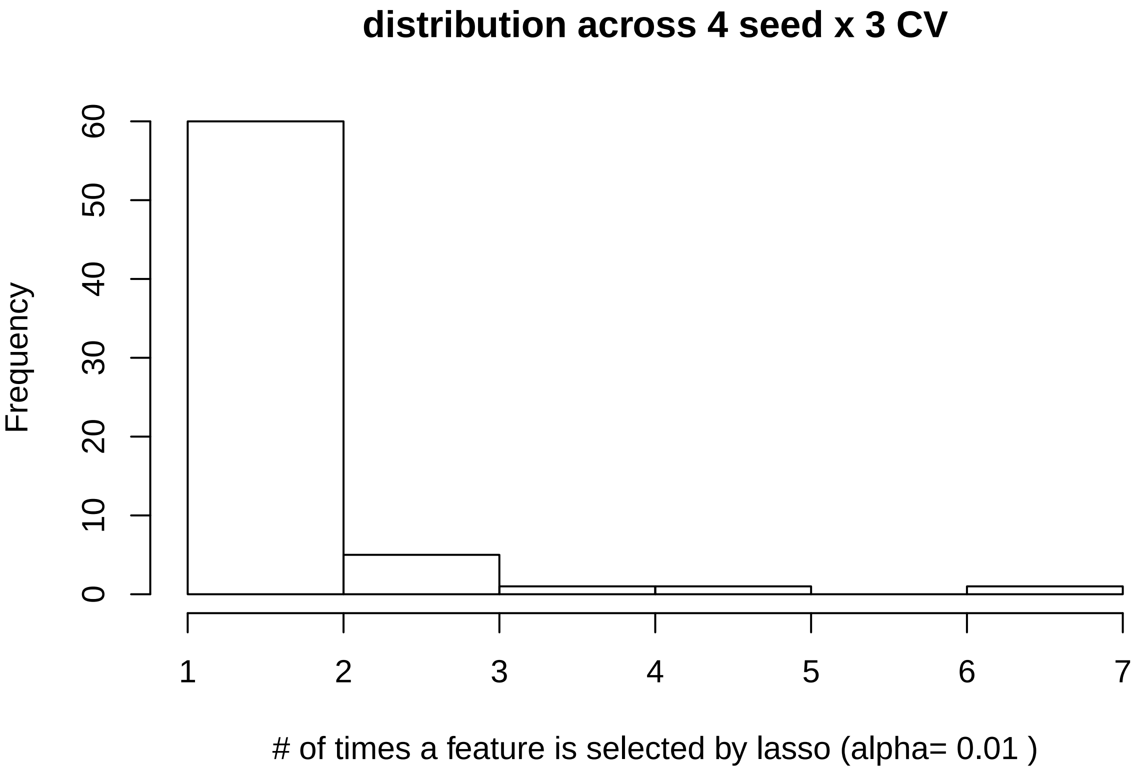
'Pred' compares the actual CV prediction with observed value. 'Rand' compares permuted CV prediction with observed to mimic random prediction.

correlation

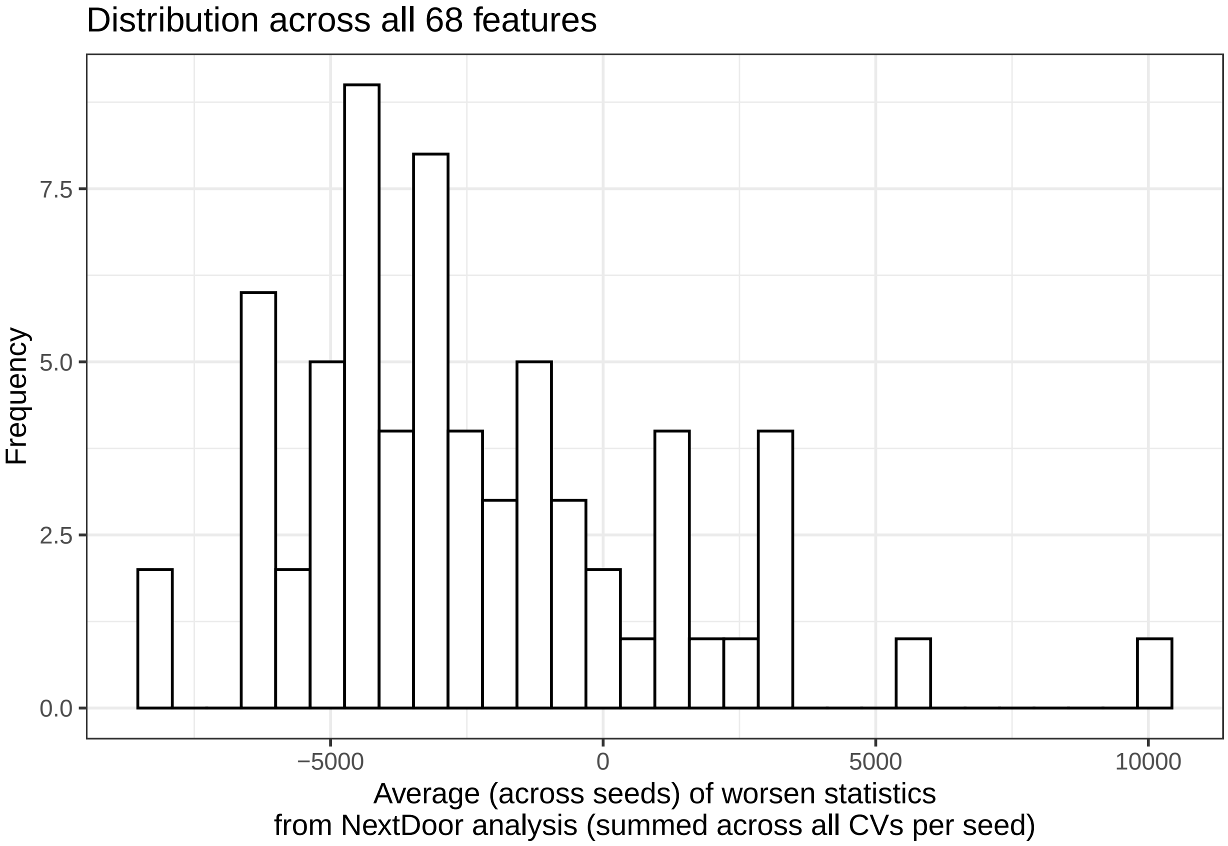
```
##
##
## Table: Averaged pearson correlation across seeds
##
##     cor.avg    cor.sdt
## -----------  ---------
##  -0.0675081   0.046583
```



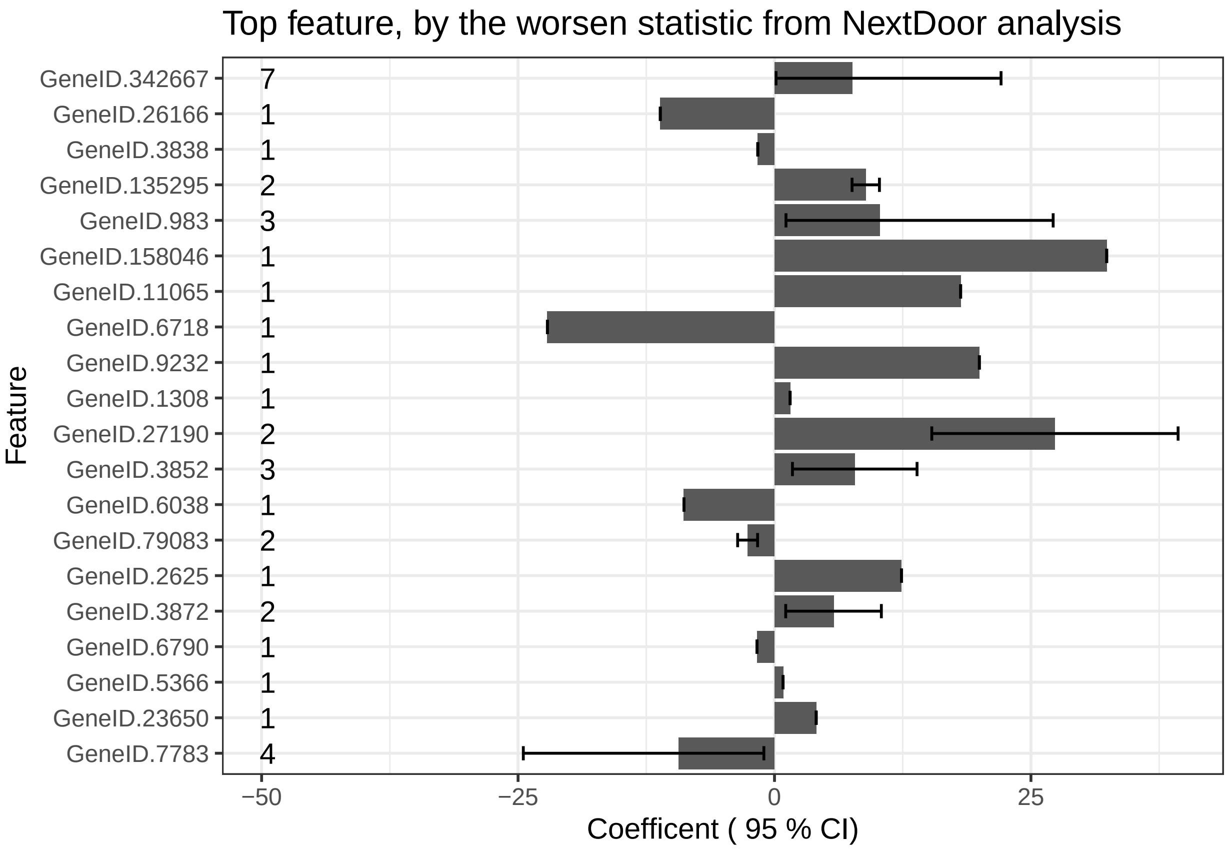
2. Important Features



```
## [1] "there are 68 unique features used from the 100 feature set"
## [1] "summary of number of features used in each run under 4 seeds and 3 CVs"

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.00    4.00    9.00   15.14   14.50   59.00       5

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing non-finite values (stat_bin).
```

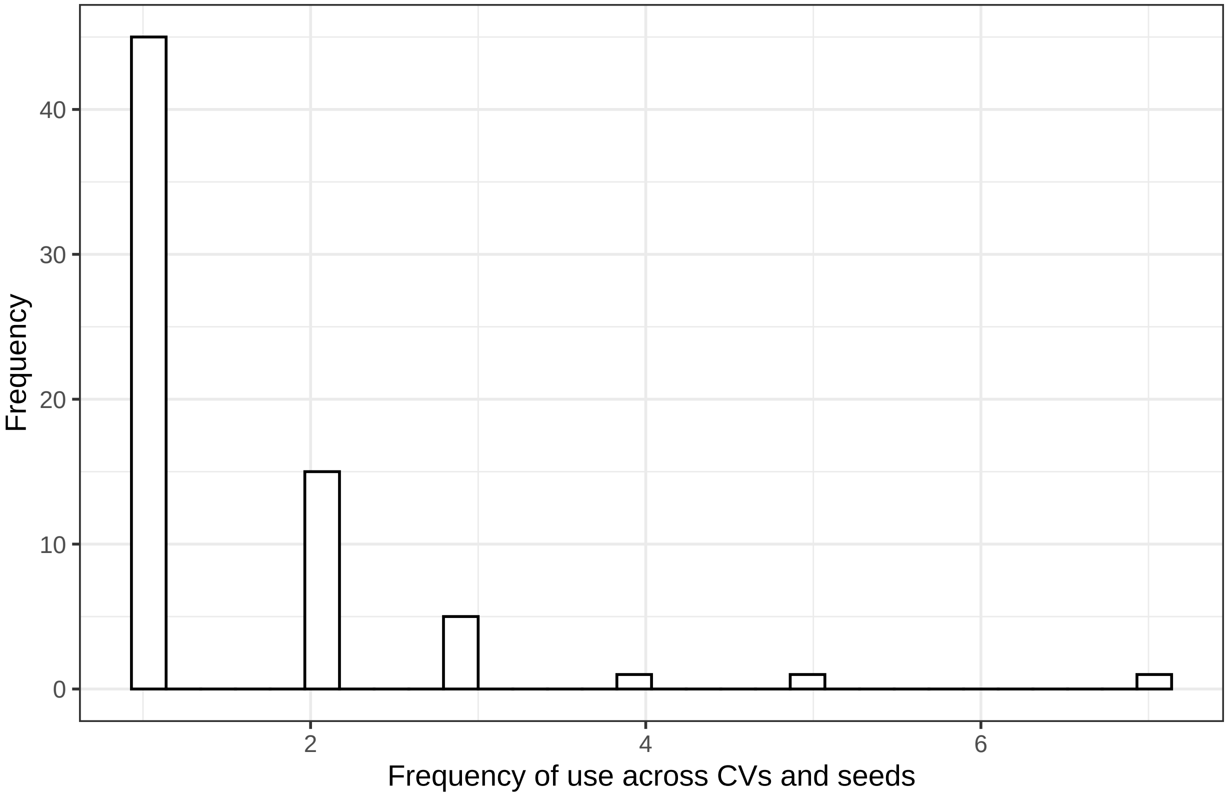


Distribution across all 68 features

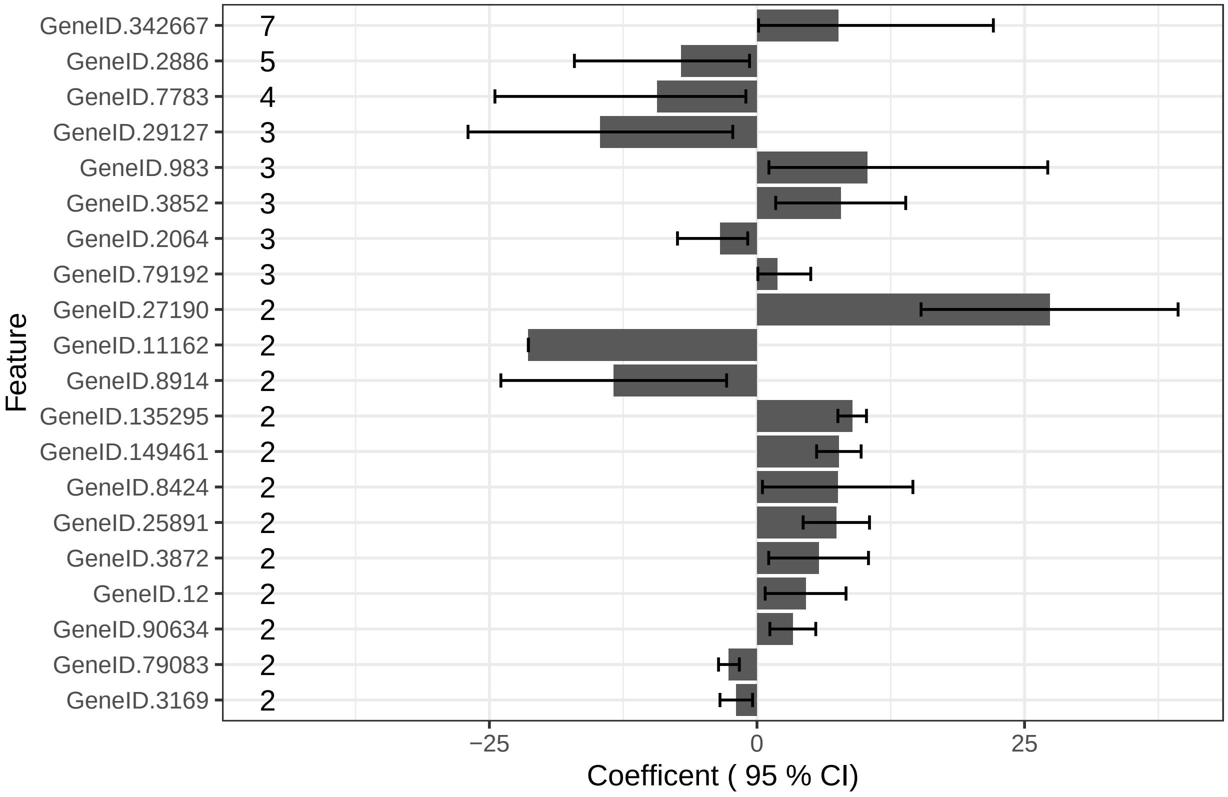


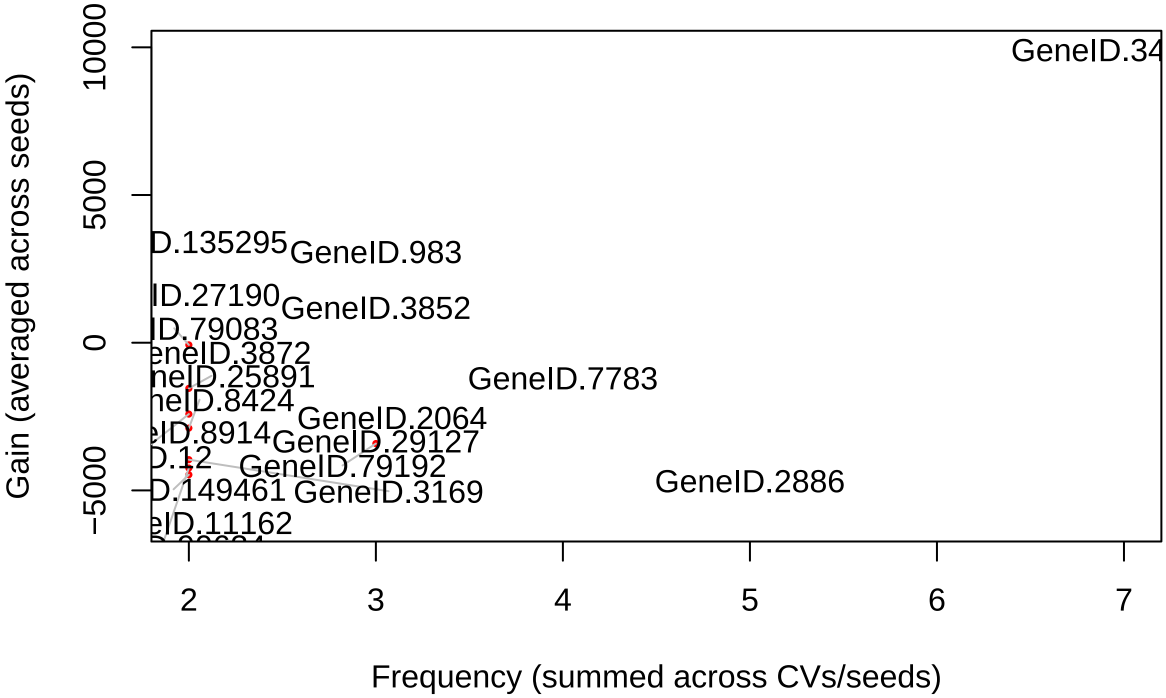
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

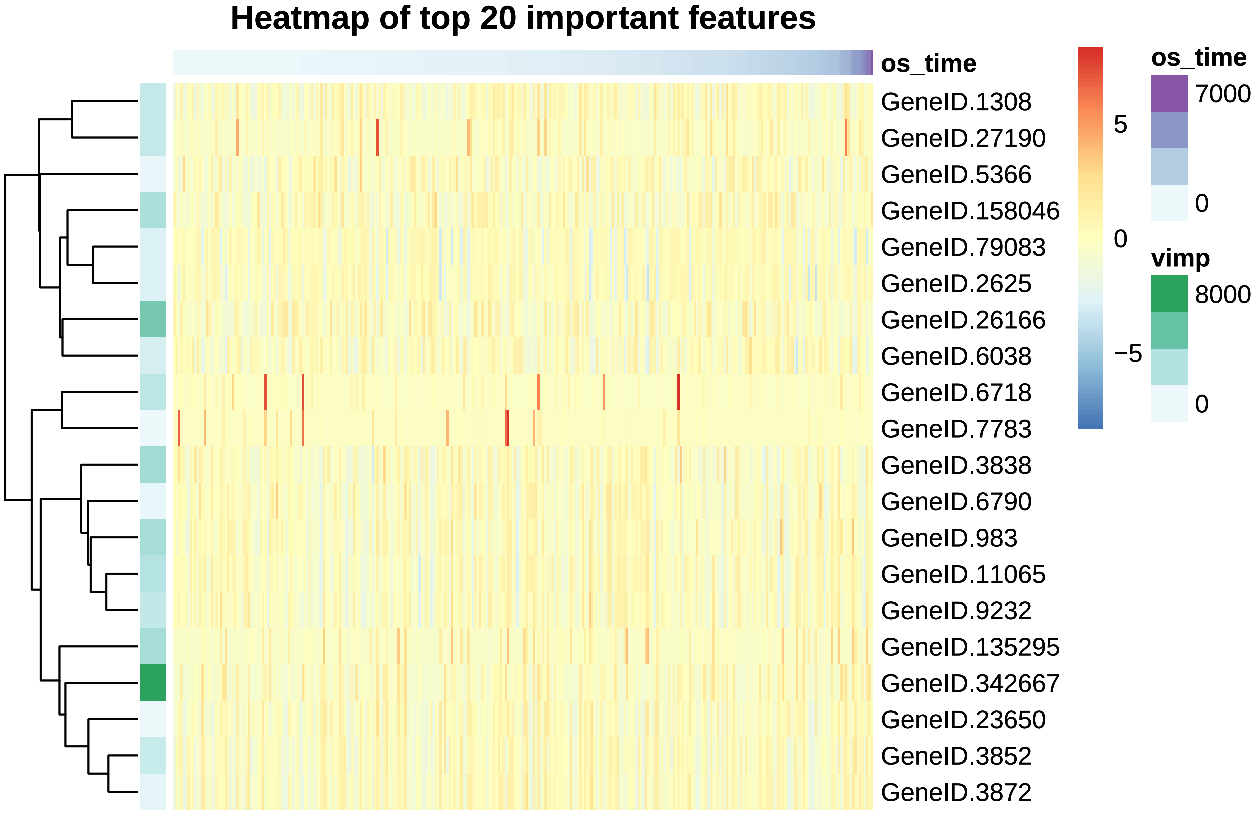
Distribution across all 68 features



Top feature, by usage frequency




GeneID.34
D.135295
GeneID.983
ID.27190
GeneID.3852
ID.79083
eneID.3872
neID.25891
GeneID.7783
neID.8424
GeneID.2064
eID.8914
GeneID.29127
D.12
GeneID.79192
D.149461
GeneID.3169
GeneID.2886
eID.11162
Gain (averaged across seeds)
10000
5000
0
−5000
2
3
4
5
6
7
Frequency (summed across CVs/seeds)


Heatmap of top 20 important features
os_time
GeneID.1308
GeneID.27190
GeneID.5366
GeneID.158046
GeneID.79083
GeneID.2625
GeneID.26166
GeneID.6038
GeneID.6718
GeneID.7783
GeneID.3838
GeneID.6790
GeneID.983
GeneID.11065
GeneID.9232
GeneID.135295
GeneID.342667
GeneID.23650
GeneID.3852
GeneID.3872
5
0
−5
os_time
7000
0
vimp
8000
0