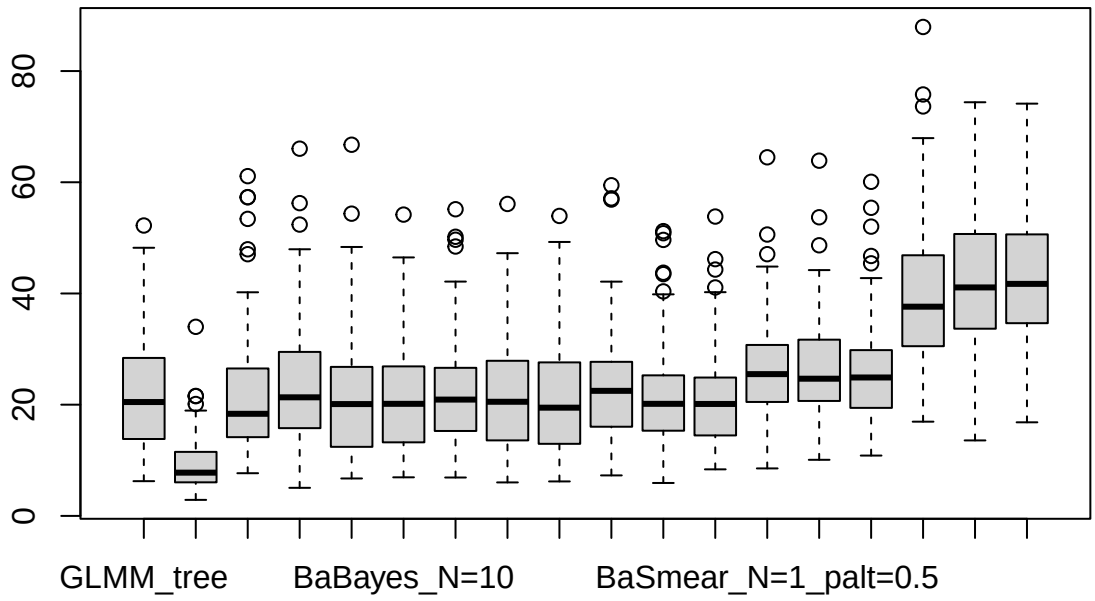
Plots

Marjolein Fokkema

20-7-2022

Boston Housing

```
load(file = "BostonHousing MSE.Rda")
load(file = "BostonHousing tree_size.Rda")
boxplot(MSE)
```




```
colMeans(MSE)
```

```
##              GLMM_tree                   Bart                     Ba
##              22.457318               9.478921              21.566889
##           BaBayes_N=1           BaBayes_N=5          BaBayes_N=10
```

```
##              23.530386              21.786926              21.623075
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##              22.484683              21.574473              21.364545
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##              23.266902              21.820402              21.446998
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##              26.156004              26.711783              26.284915
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##              39.813390              42.392009              42.771864
```

```
which.min(colMeans(MSE[ , -2]))
```

```
## BaSmear_N=10_palt=0
##                   8
```

```
sapply(MSE, sd)
```

```
##              GLMM_tree                   Bart                     Ba
##              10.501581               4.938728              11.125934
##            BaBayes_N=1           BaBayes_N=5           BaBayes_N=10
##              10.669522              11.371690              10.048628
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##               9.738342               9.966675              10.392816
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##               9.832149               9.026467               8.443231
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##               9.042027               9.050548               9.187882
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##              13.099729              12.940042              11.924300
```

```
boxplot(tree_size)
colMeans(tree_size)
```

```
##              GLMM_tree                   Bart                     Ba
##                  11.54                     NA                  13.62
##            BaBayes_N=1           BaBayes_N=5           BaBayes_N=10
##                  13.86                  11.57                  11.44
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##                  17.22                  13.28                  13.45
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##                  14.06                  10.23                  10.16
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##                  13.79                   9.86                   9.58
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##                  13.91                   9.18                   8.79
```

```
sapply(tree_size, sd)
```

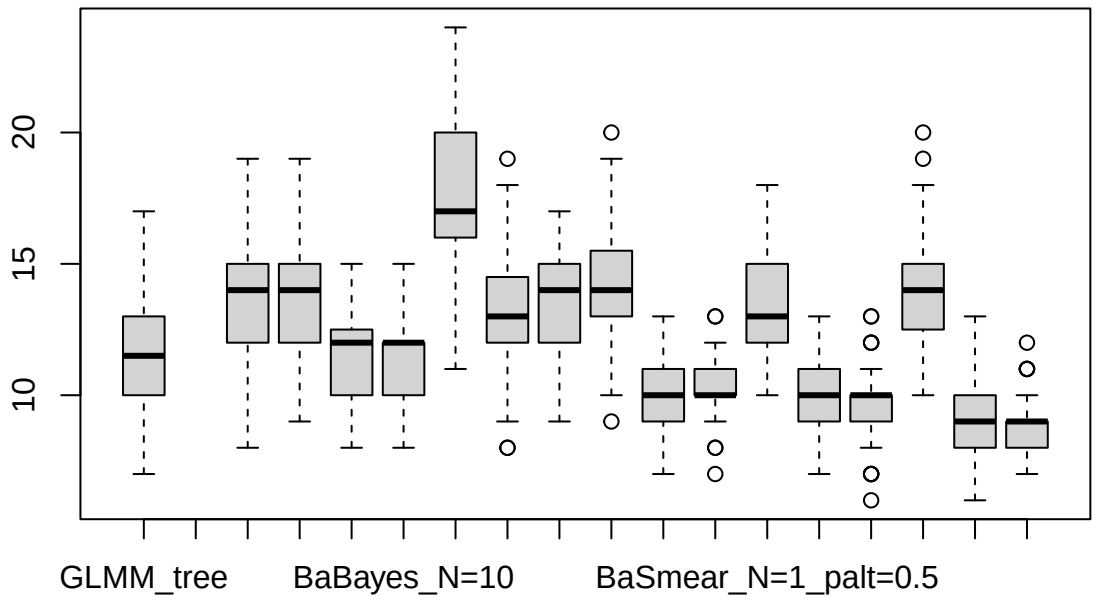
```
##              GLMM_tree                   Bart                     Ba
##               1.8280641                     NA              2.1451378
##            BaBayes_N=1           BaBayes_N=5           BaBayes_N=10
```

```
##              2.2919270              1.6221634              1.7192904
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##              2.8161870              2.0893187              1.6537377
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##              2.0588317              1.2701308              0.9818556
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##              1.9451831              1.3028330              1.3040729
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##              1.9389417              1.3210036              0.9670845
```

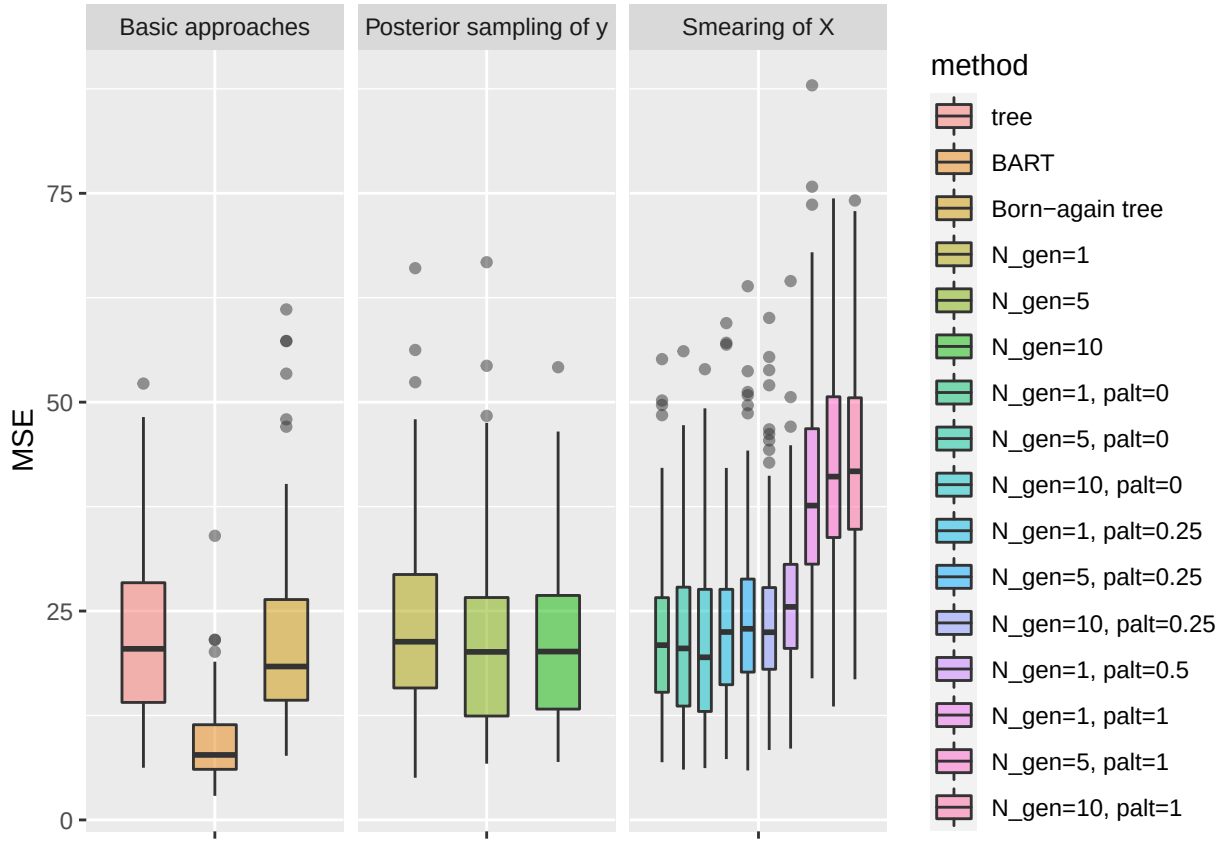
```r
## Create long data
BH_MSE_long <- stack(MSE)
## Create identifier for: original, BA with N, Bayesian BA, Breimand BA
names(BH_MSE_long) <- c("MSE", "method")
BH_MSE_long$panel <- ifelse(grepl("BaBayes", BH_MSE_long$method), "Posterior sampling of y",
                            ifelse(grepl("BaSmear", BH_MSE_long$method), "Smearing of X", "Basic approa

levels(BH_MSE_long$method) <- c("tree", "BART", "Born-again tree", "N_gen=1", "N_gen=5", "N_gen=10",
                                "N_gen=1, palt=0", "N_gen=5, palt=0",  "N_gen=10, palt=0",
                                "N_gen=1, palt=0.25", "N_gen=5, palt=0.25",  "N_gen=10, palt=0.25",
                                "N_gen=1, palt=0.5", "N_gen=5, palt=0.25",  "N_gen=10, palt=0.25",
                                "N_gen=1, palt=1", "N_gen=5, palt=1",  "N_gen=10, palt=1")


library("ggplot2")
```



```
ggplot(BH_MSE_long) +
  geom_boxplot(aes(x=panel, y=MSE, fill = method),
               position=position_dodge(1), alpha = .5, width = .6) +
  facet_grid(~panel, scales = "free", space = "free") +
  labs(x = "", y = "MSE") +
  theme(axis.title.x=element_blank(), axis.text.x=element_blank())
```



Ozone

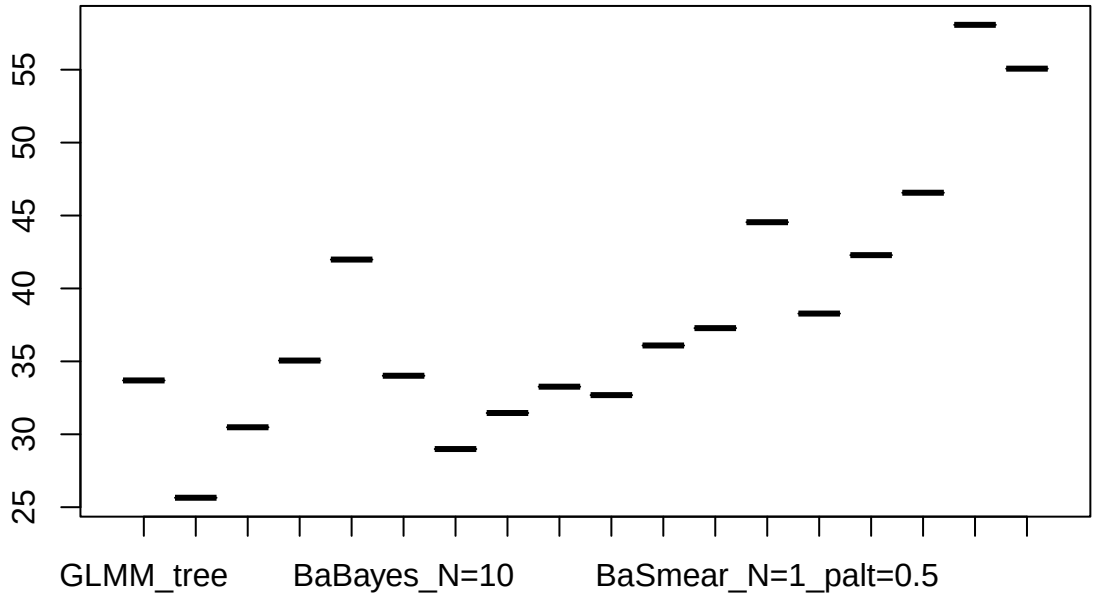
```
load(file = "Ozone MSE.Rda")
load(file = "Ozone tree_size.Rda")

sapply(MSE, function(x) table(is.na(x)))
```

```
##       GLMM_tree Bart Ba BaBayes_N=1 BaBayes_N=5 BaBayes_N=10 BaSmear_N=1_palt=0
## FALSE         1    1  1           1           1            1                   1
## TRUE         99   99 99          99          99           99                  99
##       BaSmear_N=5_palt=0 BaSmear_N=10_palt=0 BaSmear_N=1_palt=0.25
## FALSE                  1                   1                     1
## TRUE                  99                  99                    99
##       BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25 BaSmear_N=1_palt=0.5
## FALSE                     1                      1                    1
```

```
## TRUE                      99                       99                     99
##       BaSmear_N=5_palt=0.5 BaSmear_N=10_palt=0.5 BaSmear_N=1_palt=1
## FALSE                    1                        1                      1
## TRUE                    99                       99                     99
##       BaSmear_N=5_palt=1 BaSmear_N=10_palt=1
## FALSE                  1                    1
## TRUE                  99                   99
```

```
boxplot(MSE)
```




```
colMeans(MSE, na.rm=TRUE)
```

```
##              GLMM_tree                   Bart                     Ba
##               33.69184               25.64857               30.48147
##           BaBayes_N=1           BaBayes_N=5          BaBayes_N=10
##               35.05636               41.98095               34.01445
##    BaSmear_N=1_palt=0    BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##               28.98884               31.45949               33.26468
##  BaSmear_N=1_palt=0.25 BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##               32.68544               36.08909               37.27931
##   BaSmear_N=1_palt=0.5  BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##               44.54252               38.27715               42.28423
##    BaSmear_N=1_palt=1    BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##               46.56337               58.07478               55.07147
```

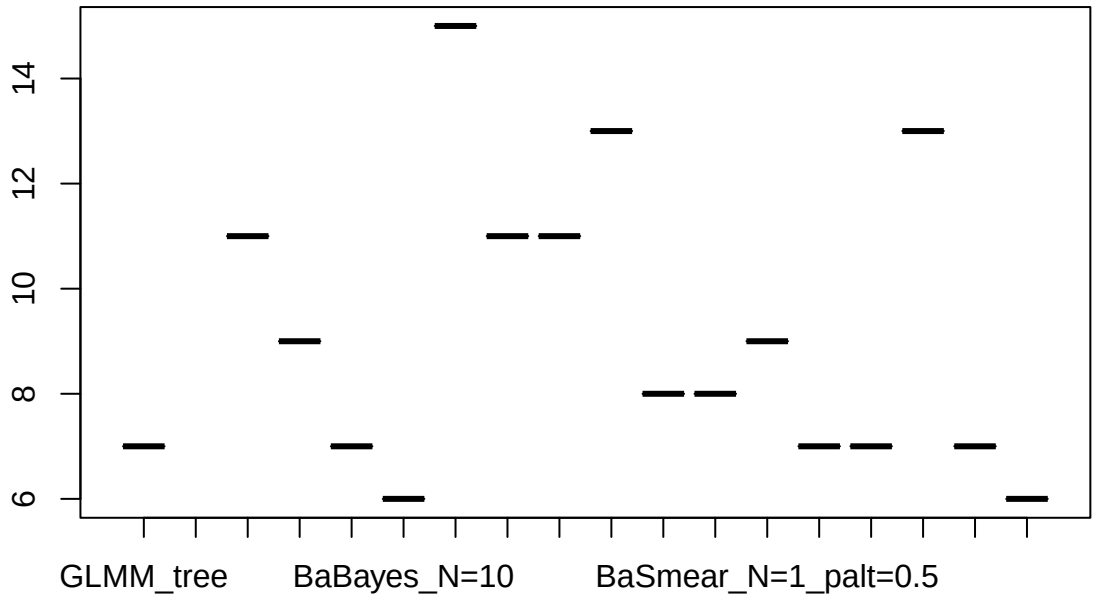
```
which.min(colMeans(MSE[ , -2]))
```

```
## integer(0)
```

```
sapply(MSE, sd)
```

```
##              GLMM_tree                   Bart                     Ba
##                     NA                     NA                     NA
##           BaBayes_N=1           BaBayes_N=5          BaBayes_N=10
##                     NA                     NA                     NA
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##                     NA                     NA                     NA
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##                     NA                     NA                     NA
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##                     NA                     NA                     NA
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##                     NA                     NA                     NA
```

```
boxplot(tree_size)
```




```
colMeans(tree_size)
```

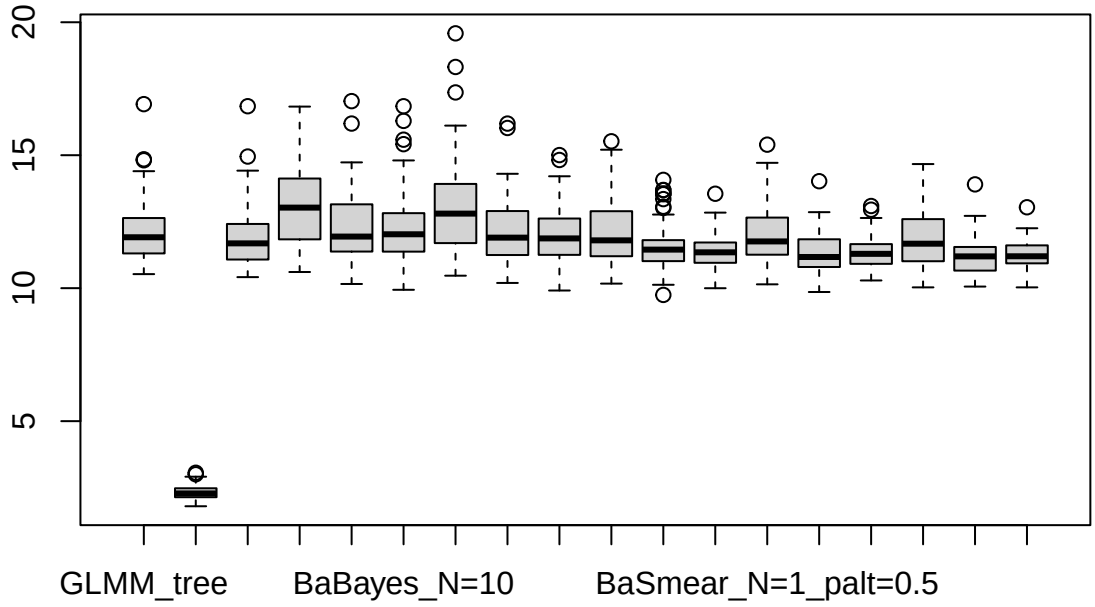
```
##              GLMM_tree                    Bart                      Ba
##                     NA                      NA                      NA
##           BaBayes_N=1           BaBayes_N=5          BaBayes_N=10
##                     NA                      NA                      NA
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##                     NA                      NA                      NA
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##                     NA                      NA                      NA
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##                     NA                      NA                      NA
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##                     NA                      NA                      NA
```

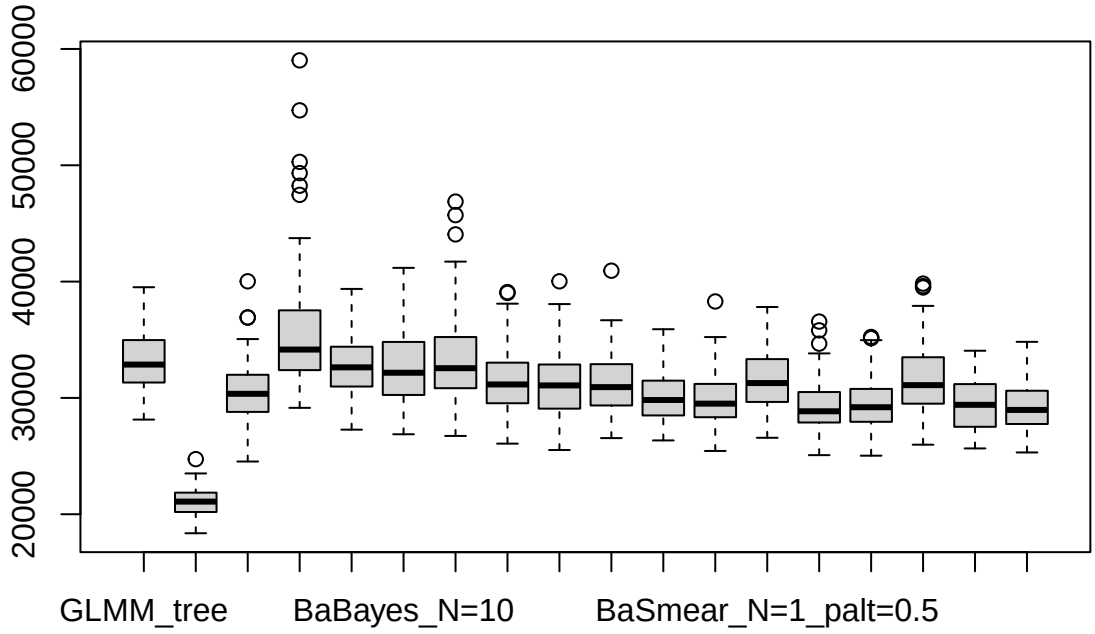
```
sapply(tree_size, sd)
```

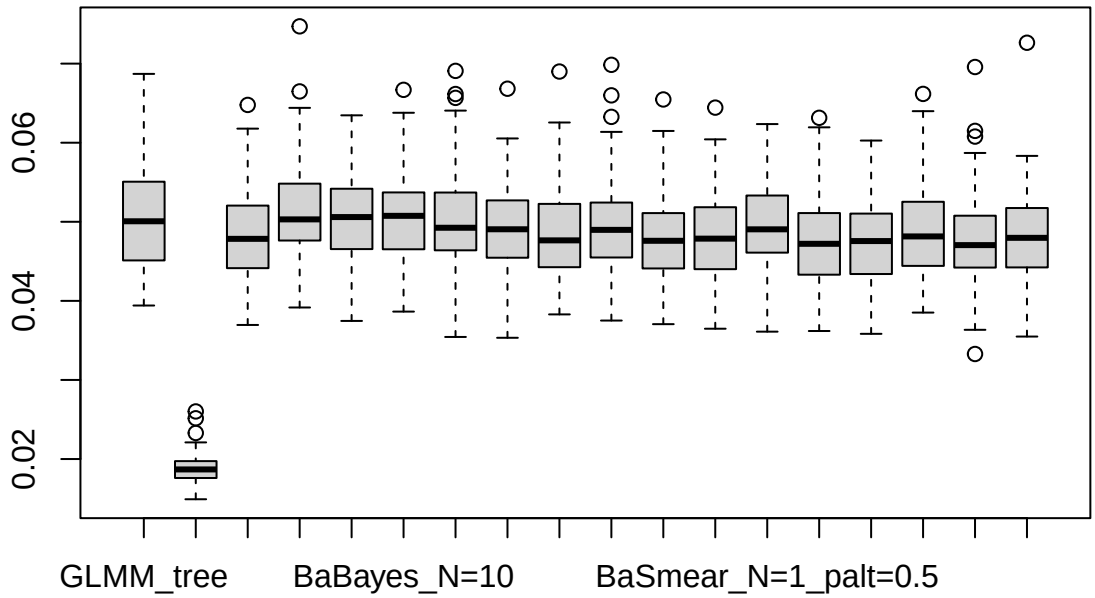
```
##              GLMM_tree                    Bart                      Ba
##                     NA                      NA                      NA
##           BaBayes_N=1           BaBayes_N=5          BaBayes_N=10
##                     NA                      NA                      NA
##     BaSmear_N=1_palt=0     BaSmear_N=5_palt=0    BaSmear_N=10_palt=0
##                     NA                      NA                      NA
##  BaSmear_N=1_palt=0.25  BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
##                     NA                      NA                      NA
##   BaSmear_N=1_palt=0.5   BaSmear_N=5_palt=0.5  BaSmear_N=10_palt=0.5
##                     NA                      NA                      NA
##     BaSmear_N=1_palt=1     BaSmear_N=5_palt=1    BaSmear_N=10_palt=1
##                     NA                      NA                      NA
```

Friedman

```
library("mlbench")
load(file = "Friedman MSE.Rda")
load(file = "Friedman tree_size.Rda")
set.seed(42)
vars <- c(var(mlbench.friedman1(10000)$y),
          var(mlbench.friedman2(210000)$y),
          var(mlbench.friedman3(10000)$y))
sapply(MSE, boxplot)
```


20
15
10
5
GLMM_tree
BaBayes_N=10
BaSmear_N=1_palt=0.5


20000
30000
40000
50000
60000
GLMM_tree
BaBayes_N=10
BaSmear_N=1_palt=0.5




```
##       [,1]         [,2]         [,3]
## stats numeric,90   numeric,90   numeric,90
## n     numeric,18   numeric,18   numeric,18
## conf  numeric,36   numeric,36   numeric,36
## out   numeric,38   numeric,27   numeric,24
## group numeric,38   numeric,27   numeric,24
## names character,18 character,18 character,18
```

```
sapply(MSE, colMeans)
```

```
##                            [,1]     [,2]       [,3]
## GLMM_tree              12.076007 33187.34 0.05033159
## Bart                    2.304523 21054.85 0.01880145
## Ba                     11.899889 30487.48 0.04839876
## BaBayes_N=1            13.166062 35566.36 0.05146380
## BaBayes_N=5            12.247483 32711.49 0.05013863
## BaBayes_N=10           12.207703 32719.61 0.05031103
## BaSmear_N=1_palt=0     13.042185 33394.36 0.05041427
## BaSmear_N=5_palt=0     12.073354 31498.65 0.04928297
## BaSmear_N=10_palt=0    11.976251 31029.08 0.04846355
## BaSmear_N=1_palt=0.25  12.075405 31172.79 0.04935003
## BaSmear_N=5_palt=0.25  11.507193 30023.96 0.04778332
## BaSmear_N=10_palt=0.25 11.339046 29879.85 0.04781966
## BaSmear_N=1_palt=0.5   12.034039 31381.05 0.04956370
## BaSmear_N=5_palt=0.5   11.288992 29343.29 0.04745073
```

```
## BaSmear_N=10_palt=0.5  11.329361 29413.42 0.04769477
## BaSmear_N=1_palt=1     11.819585 31570.06 0.04928029
## BaSmear_N=5_palt=1     11.198944 29506.09 0.04784267
## BaSmear_N=10_palt=1    11.247026 29137.52 0.04811013
```

```
round((1/vars)*(t(sapply(MSE, colMeans))), digits = 3)
```

```
##      GLMM_tree  Bart    Ba BaBayes_N=1 BaBayes_N=5 BaBayes_N=10
## [1,]     0.486 0.093 0.479       0.530       0.493        0.491
## [2,]     0.208 0.132 0.191       0.223       0.205        0.205
## [3,]     0.442 0.165 0.425       0.451       0.440        0.441
##      BaSmear_N=1_palt=0 BaSmear_N=5_palt=0 BaSmear_N=10_palt=0
## [1,]              0.525              0.486               0.482
## [2,]              0.210              0.198               0.195
## [3,]              0.442              0.432               0.425
##      BaSmear_N=1_palt=0.25 BaSmear_N=5_palt=0.25 BaSmear_N=10_palt=0.25
## [1,]                 0.486                 0.463                  0.456
## [2,]                 0.196                 0.188                  0.187
## [3,]                 0.433                 0.419                  0.420
##      BaSmear_N=1_palt=0.5 BaSmear_N=5_palt=0.5 BaSmear_N=10_palt=0.5
## [1,]                0.484                0.454                 0.456
## [2,]                0.197                0.184                 0.185
## [3,]                0.435                0.416                 0.418
##      BaSmear_N=1_palt=1 BaSmear_N=5_palt=1 BaSmear_N=10_palt=1
## [1,]              0.476              0.451               0.453
## [2,]              0.198              0.185               0.183
## [3,]              0.432              0.420               0.422
```

```
sapply(MSE, function(x) which.min(colMeans(x[ , -2])))
```

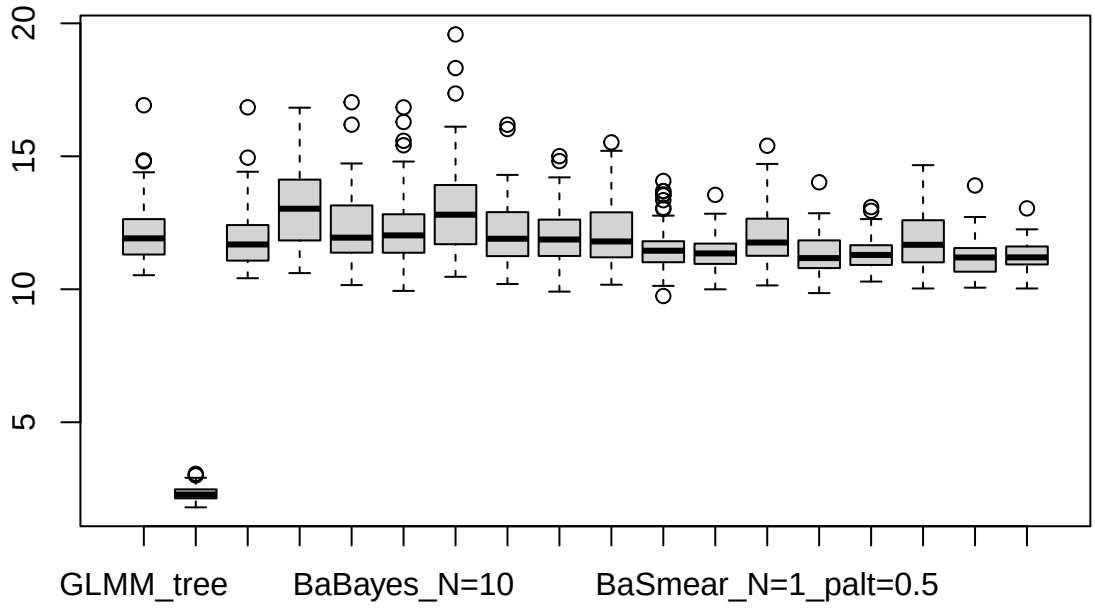
```
##   BaSmear_N=5_palt=1  BaSmear_N=10_palt=1 BaSmear_N=5_palt=0.5
##                   16                   17                   13
```

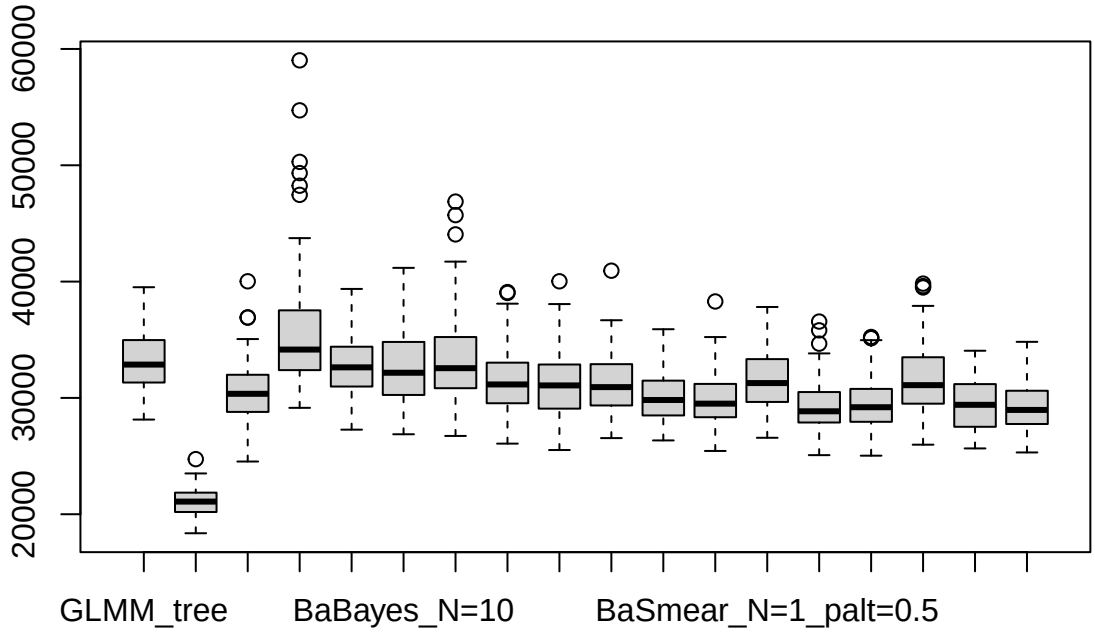
```
sapply(MSE, function(x) sapply(x, sd))
```

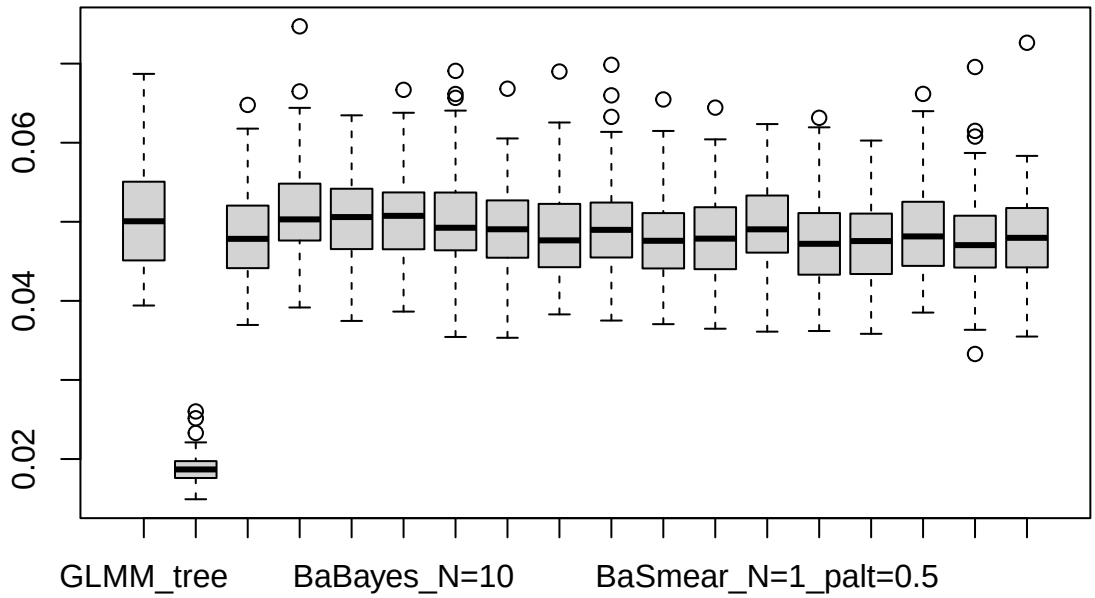
```
##                             [,1]     [,2]        [,3]
## GLMM_tree              1.0830838 2599.800 0.006148888
## Bart                   0.2321861 1206.724 0.001907856
## Ba                     1.1238855 2565.068 0.005855916
## BaBayes_N=1            1.5623625 5189.825 0.006404210
## BaBayes_N=5            1.2824767 2559.969 0.005485072
## BaBayes_N=10           1.3374984 3283.079 0.005854500
## BaSmear_N=1_palt=0     1.6714595 3839.703 0.006336858
## BaSmear_N=5_palt=0     1.2009723 2653.883 0.005337941
## BaSmear_N=10_palt=0    1.0352730 2696.107 0.005994306
## BaSmear_N=1_palt=0.25  1.1761710 2496.272 0.005894570
## BaSmear_N=5_palt=0.25  0.8533598 2113.327 0.005512098
## BaSmear_N=10_palt=0.25 0.5798938 2351.263 0.005618629
## BaSmear_N=1_palt=0.5   1.1020613 2541.695 0.005726027
## BaSmear_N=5_palt=0.5   0.7202745 2222.965 0.005854793
## BaSmear_N=10_palt=0.5  0.5822913 2295.298 0.005720670
## BaSmear_N=1_palt=1     1.0387107 2930.827 0.006163060
```

```
## BaSmear_N=5_palt=1      0.6842474 2190.938 0.005722531
## BaSmear_N=10_palt=1     0.5267453 1979.481 0.005788078
```

```
sapply(MSE, boxplot)
```




20000
30000
40000
50000
60000
GLMM_tree
BaBayes_N=10
BaSmear_N=1_palt=0.5




```
##       [,1]         [,2]         [,3]
## stats numeric,90   numeric,90   numeric,90
## n     numeric,18   numeric,18   numeric,18
## conf  numeric,36   numeric,36   numeric,36
## out   numeric,38   numeric,27   numeric,24
## group numeric,38   numeric,27   numeric,24
## names character,18 character,18 character,18
```

```
sapply(tree_size, colMeans)
```

```
##                         [,1]  [,2]  [,3]
## GLMM_tree               6.85  9.02  6.76
## Bart                      NA    NA    NA
## Ba                      7.16 11.88  9.48
## BaBayes_N=1             9.02  9.97  8.08
## BaBayes_N=5             6.95  9.56  7.25
## BaBayes_N=10            6.83  9.61  6.94
## BaSmear_N=1_palt=0      9.01 12.04 10.27
## BaSmear_N=5_palt=0      7.46 11.31  9.33
## BaSmear_N=10_palt=0     7.32 11.48  9.53
## BaSmear_N=1_palt=0.25   7.86 11.95  9.96
## BaSmear_N=5_palt=0.25   5.89 11.67  9.02
## BaSmear_N=10_palt=0.25 5.77 11.32  8.99
## BaSmear_N=1_palt=0.5    7.59 12.02  9.90
## BaSmear_N=5_palt=0.5    5.74 11.85  9.51
```

```
## BaSmear_N=10_palt=0.5  5.41 11.83  8.99
## BaSmear_N=1_palt=1     7.63 12.16 10.45
## BaSmear_N=5_palt=1     5.80 11.96  9.47
## BaSmear_N=10_palt=1    5.36 12.13  9.01
```

```
sapply(tree_size, function(x) sapply(x, sd))
```

```
##                              [,1]     [,2]     [,3]
## GLMM_tree               1.1752928 1.189007 1.064771
## Bart                          NA       NA       NA
## Ba                      1.2449494 1.423327 1.359441
## BaBayes_N=1             1.4769611 1.175980 1.186456
## BaBayes_N=5             1.3361712 1.066667 1.057680
## BaBayes_N=10            1.2231356 1.278375 1.052366
## BaSmear_N=1_palt=0      1.5731009 1.392041 1.398809
## BaSmear_N=5_palt=0      1.1670562 1.088925 1.271720
## BaSmear_N=10_palt=0     1.1536503 1.039425 1.266946
## BaSmear_N=1_palt=0.25   1.3028330 1.274260 1.340285
## BaSmear_N=5_palt=0.25   0.8515583 1.082972 1.497338
## BaSmear_N=10_palt=0.25 0.7365631 1.127077 1.290955
## BaSmear_N=1_palt=0.5    1.3111411 1.247057 1.403459
## BaSmear_N=5_palt=0.5    0.7052473 1.209182 1.210184
## BaSmear_N=10_palt=0.5   0.5876679 1.271720 1.329502
## BaSmear_N=1_palt=1      1.3154482 1.276991 1.373450
## BaSmear_N=5_palt=1      0.6816498 1.213726 1.452653
## BaSmear_N=10_palt=1     0.5599423 1.211519 1.424852
```

Breast cancer

```
load(file = "BreastCancer MSE.Rda")
load(file = "BreastCancer acc.Rda")
load(file = "BreastCancer tree_size.Rda")
boxplot(MSE)
boxplot(acc)
colMeans(MSE)
which.min(colMeans(MSE[ , -2]))
which.min(colMeans(acc[ , -2]))
sapply(MSE, sd)
boxplot(tree_size)
sapply(tree_size, colMeans(tree_size, na.rm=TRUE)
sapply(tree_size, sd)
```

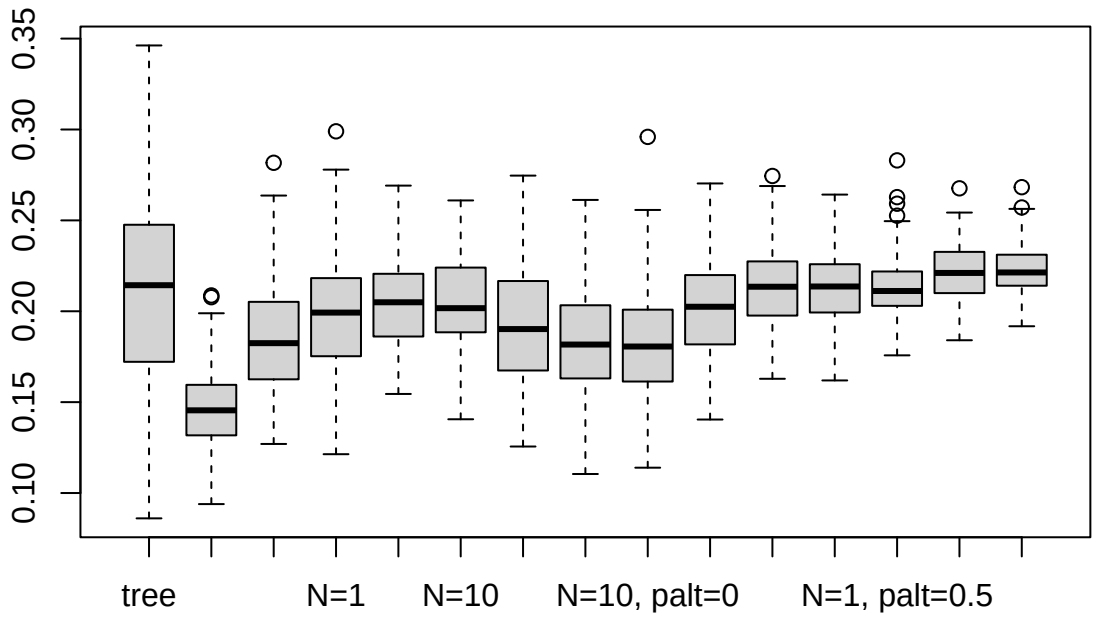
Ionosphere

```
load(file = "Ionosphere MSE.Rda")
load(file = "Ionosphere acc.Rda")
load(file = "Ionosphere tree_size.Rda")
boxplot(MSE)
```

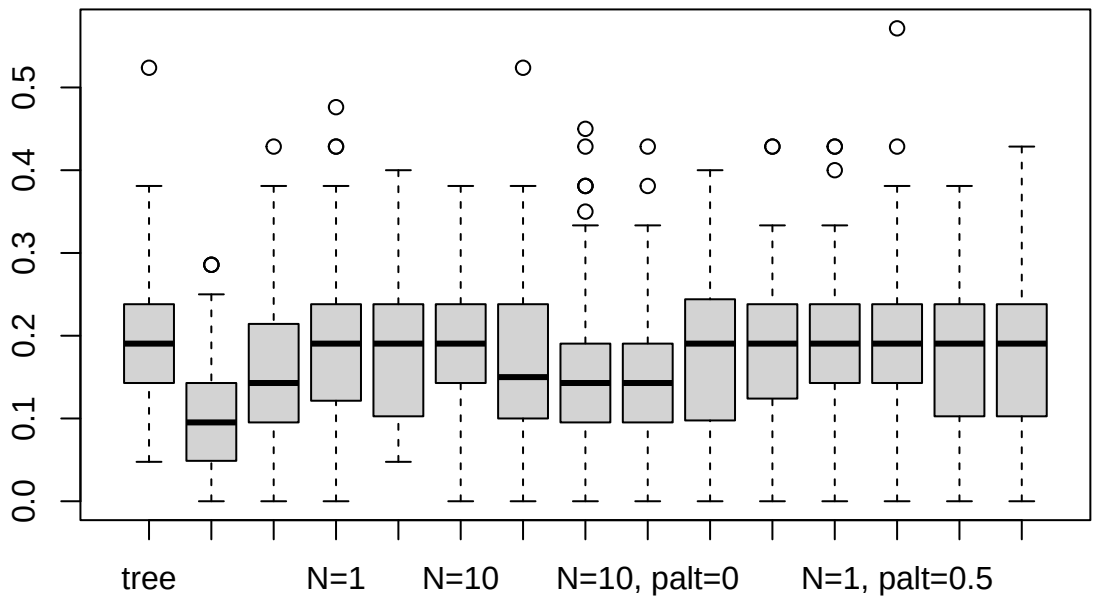
```
boxplot(acc)
colMeans(MSE)
which.min(colMeans(MSE[ , -2]))
which.min(colMeans(acc[ , -2]))
sapply(MSE, sd)
boxplot(tree_size)
colMeans(tree_size, na.rm=TRUE)
sapply(tree_size, sd)
```

Sonar

```
load(file = "Sonar MSE.Rda")
load(file = "Sonar acc.Rda")
load(file = "Sonar tree_size.Rda")
boxplot(MSE)
```




```
boxplot(acc)
```




```
colMeans(MSE)
```

```
##            tree            BART Born-again tree             N=1             N=5
##       0.2141351       0.1468447       0.1847915       0.1998829       0.2053185
##            N=10     N=1, palt=0     N=5, palt=0    N=10, palt=0  N=1, palt=0.25
##       0.2060179       0.1911827       0.1832554       0.1832785       0.2019839
##  N=5, palt=0.25 N=10, palt=0.25   N=1, palt=0.5   N=5, palt=0.5  N=10, palt=0.5
##       0.2142539       0.2139775       0.2144150       0.2215831       0.2230961
```

```
which.min(colMeans(MSE[ , -2]))
```

```
## N=5, palt=0
##           7
```

```
which.min(colMeans(acc[ , -2]))
```

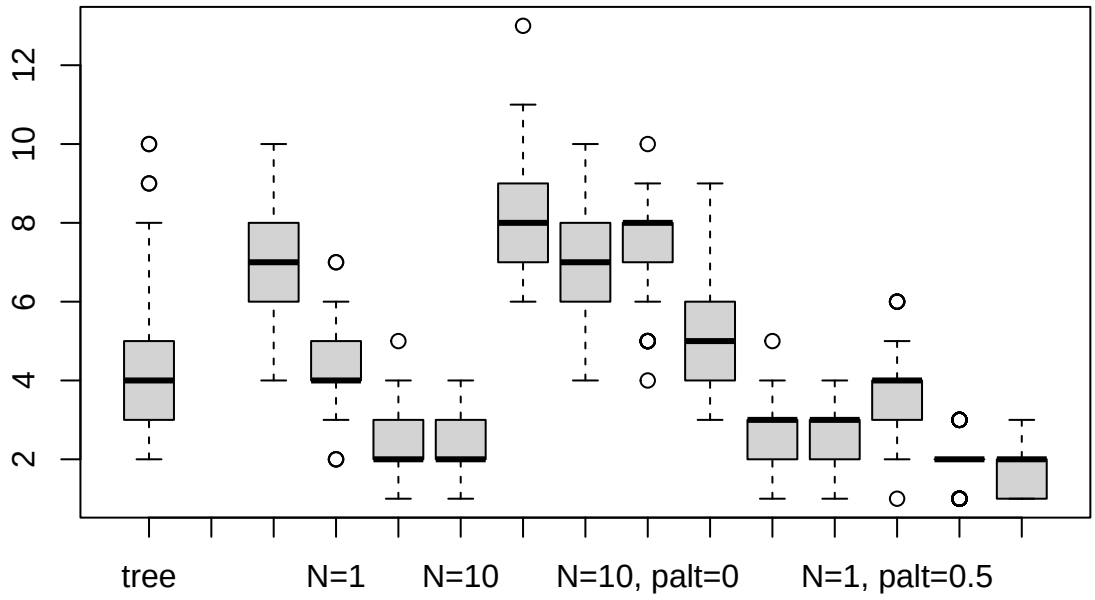
```
## N=10, palt=0
##            8
```

```
sapply(MSE, sd)
```

```
##            tree            BART Born-again tree             N=1             N=5
##      0.05597754      0.02335226      0.03209856      0.03057963      0.02687019
```

```
##           N=10     N=1, palt=0     N=5, palt=0    N=10, palt=0  N=1, palt=0.25
##      0.02553597      0.03340737      0.03102662      0.03316581      0.02696557
##  N=5, palt=0.25 N=10, palt=0.25   N=1, palt=0.5   N=5, palt=0.5  N=10, palt=0.5
##      0.02292906      0.02112953      0.01913660      0.01589785      0.01430024
```

```
boxplot(tree_size)
```




```
colMeans(tree_size, na.rm=TRUE)
```

```
##           tree            BART Born-again tree            N=1            N=5
##           4.32             NaN            7.17           4.29           2.48
##           N=10     N=1, palt=0     N=5, palt=0    N=10, palt=0  N=1, palt=0.25
##           2.34            8.30            7.06           7.44           5.31
##  N=5, palt=0.25 N=10, palt=0.25   N=1, palt=0.5   N=5, palt=0.5  N=10, palt=0.5
##           2.82            2.54            3.64           1.91           1.69
```

```
sapply(tree_size, sd)
```

```
##           tree            BART Born-again tree            N=1            N=5
##      1.7516515              NA       1.3857500      0.9877533      0.8584694
##           N=10     N=1, palt=0     N=5, palt=0    N=10, palt=0  N=1, palt=0.25
##      0.8787043       1.3521401       1.2698525       1.1833867       1.1780398
##  N=5, palt=0.25 N=10, palt=0.25   N=1, palt=0.5   N=5, palt=0.5  N=10, palt=0.5
##      0.6416519       0.6878454       1.0873004       0.6210939       0.6145541
```