

LDA training on random subsets

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

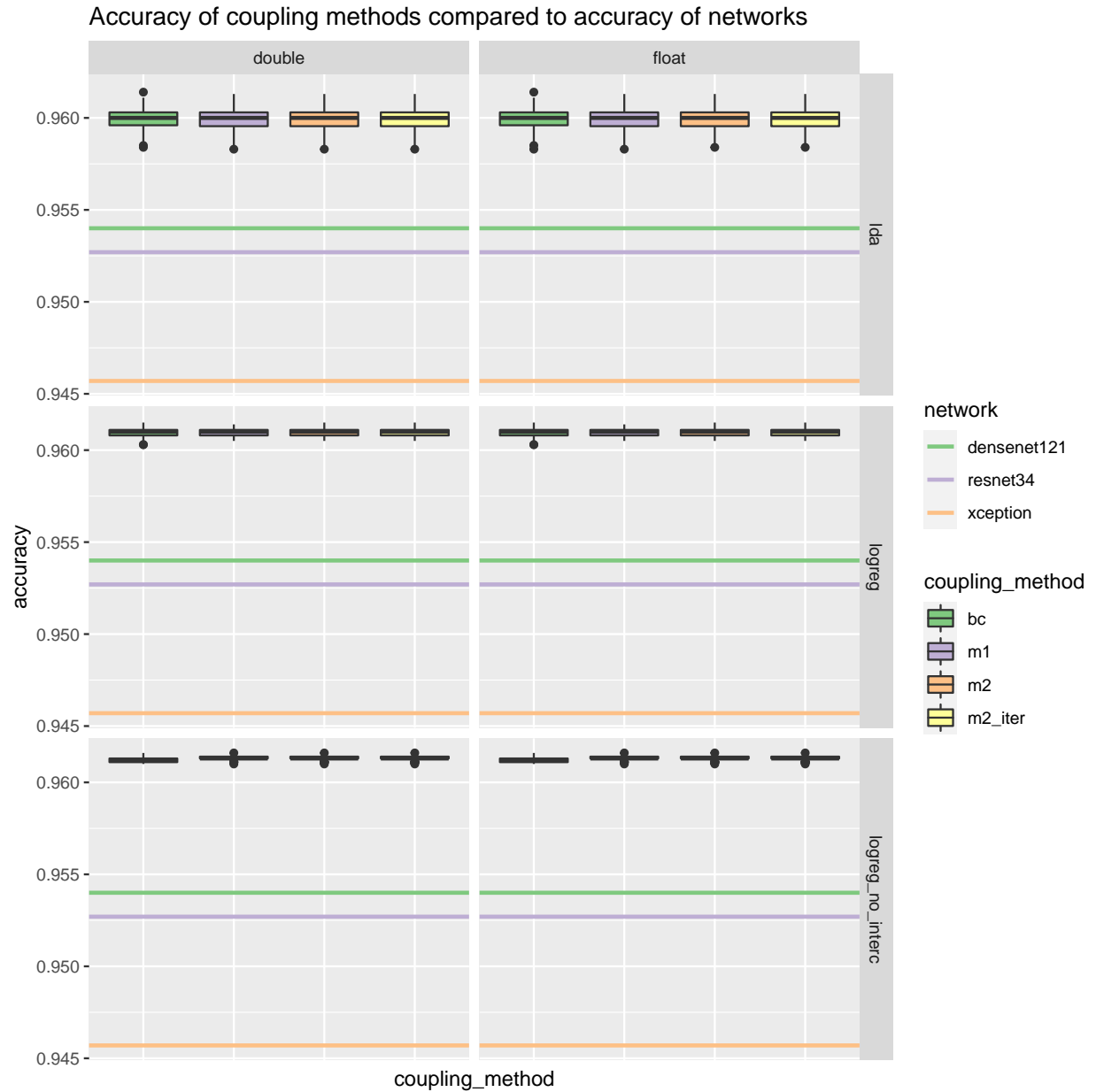
```
##
```

```
##      intersect, setdiff, setequal, union
```

Experiment code is in the file `training_subsets_experiment.py`. Experiment on CIFAR10 dataset. This experiment trains `WeightedLinearEnsemble` on various disjoint subsets of data on which neural networks were trained. Size of these subsets is 500, therefore each class has 50 samples and each LDA is trained on 100 samples. Three different coupling methods are used: method one and two from (Wu, Lin, and Weng 2004) and Bayes covariant method (Such and Barreda 2016). Goal of this experiment is to determine, whether usage of different training sets for LDA has important effects on the performance of the ensemble.

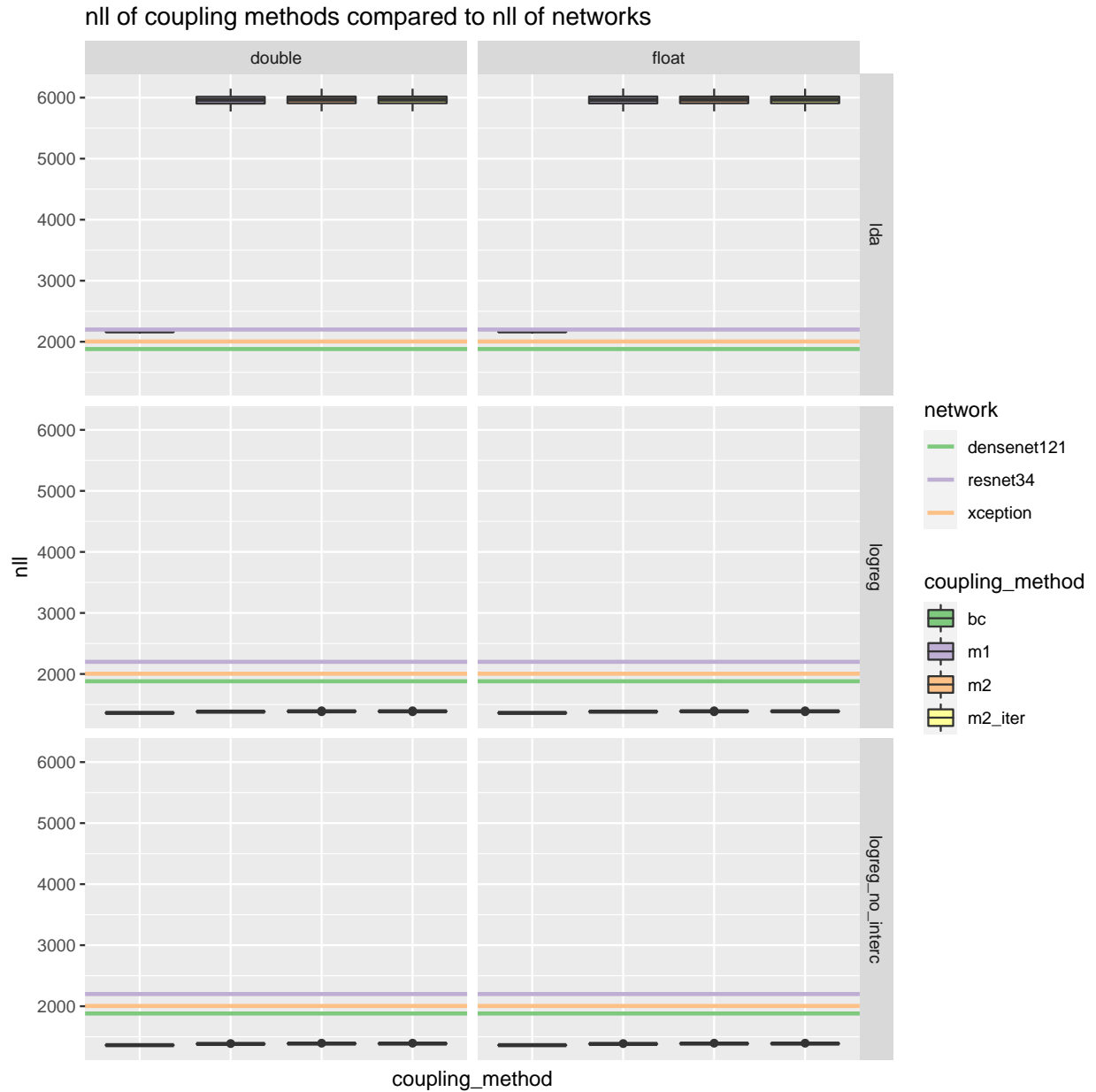
```
acc_ens_subsets <- read.csv("../data/data_train_val_c10/0/exp_subsets_train_outputs/accuracies.csv")
acc_nets <- read.csv("../data/data_train_val_c10/0/exp_subsets_train_outputs/net accuracies.csv")
```

```
box_acc <- ggplot() + geom_boxplot(data=acc_ens_subsets, mapping=aes(x=coupling_method, y=accuracy, fill=
  facet_grid(rows=vars(combining_method), cols=vars(precision)) +
  geom_hline(data=acc_nets, mapping=aes(yintercept=accuracy, color=network), size=1) +
  scale_fill_brewer(type="qual") +
  scale_color_brewer(type="qual") +
  ggtitle("Accuracy of coupling methods compared to accuracy of networks") +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank())
box_acc
```



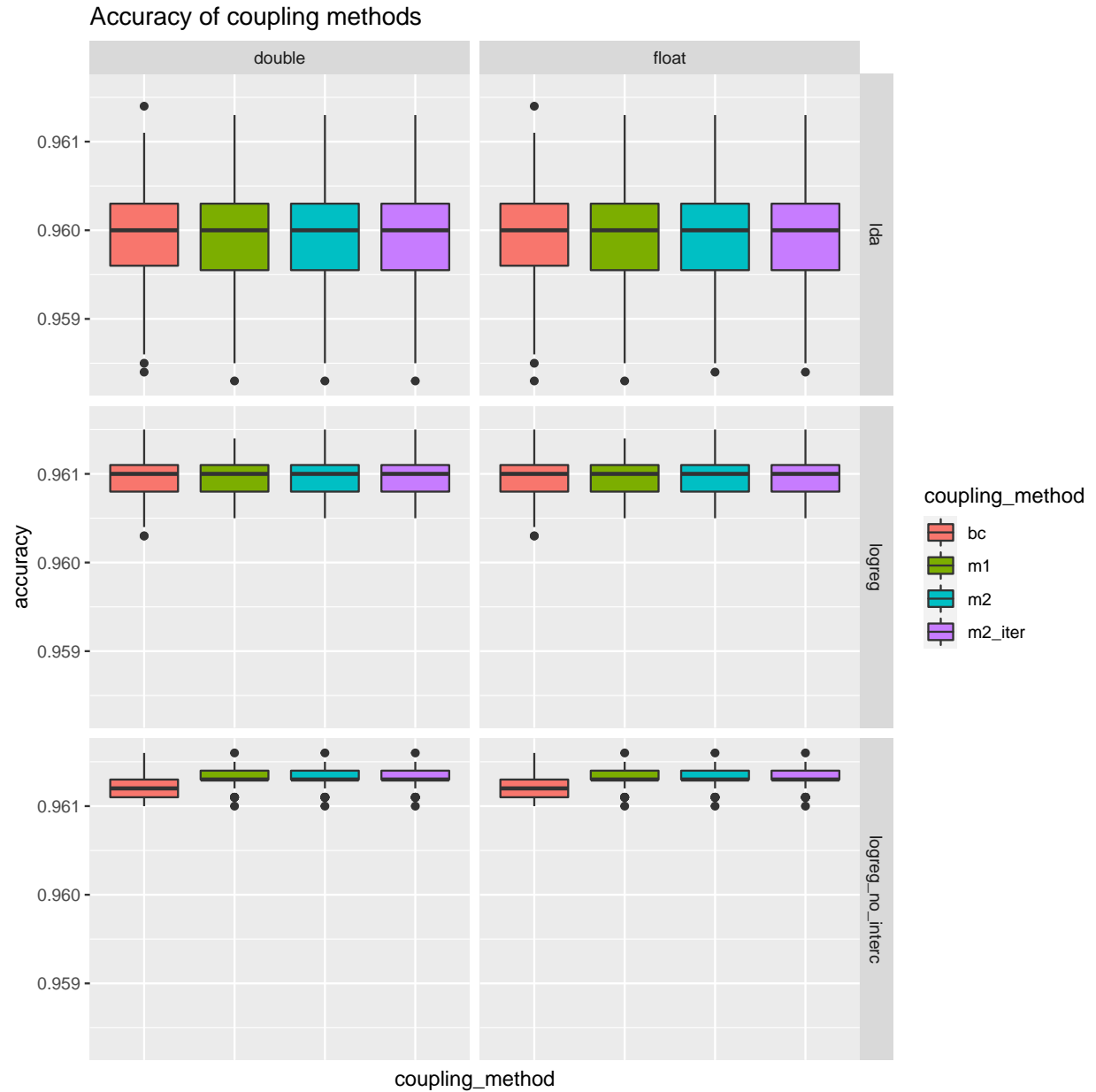
All coupling methods perform similarly. Also we can observe, that the median improvement of the ensemble against the best network is much larger than the 1.5 multiple of IQR. Variance for the method logistic regression is smaller than for lda.

```
box_nll <- ggplot() + geom_boxplot(data=acc_ens_subsets, mapping=aes(x=coupling_method, y=nll, fill=coupling_method)) +
  facet_grid(rows=vars(combining_method), cols=vars(precision)) +
  geom_hline(data=acc_nets, mapping=aes(yintercept=nll, color=network), size=1) +
  scale_fill_brewer(type="qual") +
  scale_color_brewer(type="qual") +
  ggtitle("nll of coupling methods compared to nll of networks") +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank())
box_nll
```



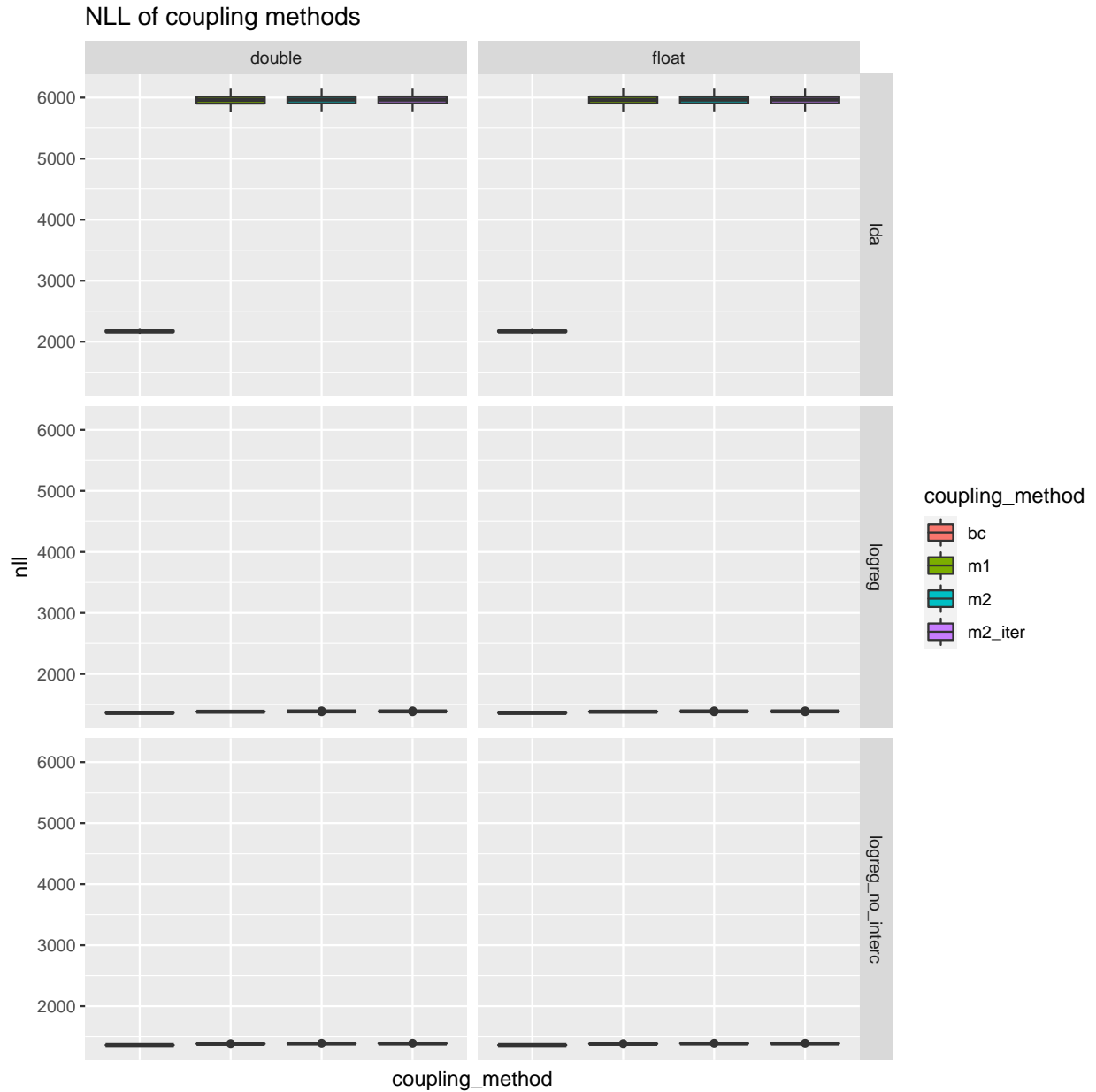
Ensembles using combining method lda have worse nll as the combined networks. This is due to ensemble producing close to zero or zero probability for the correct label in several test samples. However, using combining method logistic regression, we obtained better results than for individual networks.

```
box2_acc <- ggplot() + geom_boxplot(data=acc_ens_subsets, mapping=aes(x=coupling_method, y=accuracy, fill=network)) +
  facet_grid(rows=vars(combining_method), cols=vars(precision)) +
  ggtitle("Accuracy of coupling methods") +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank())
box2_acc
```



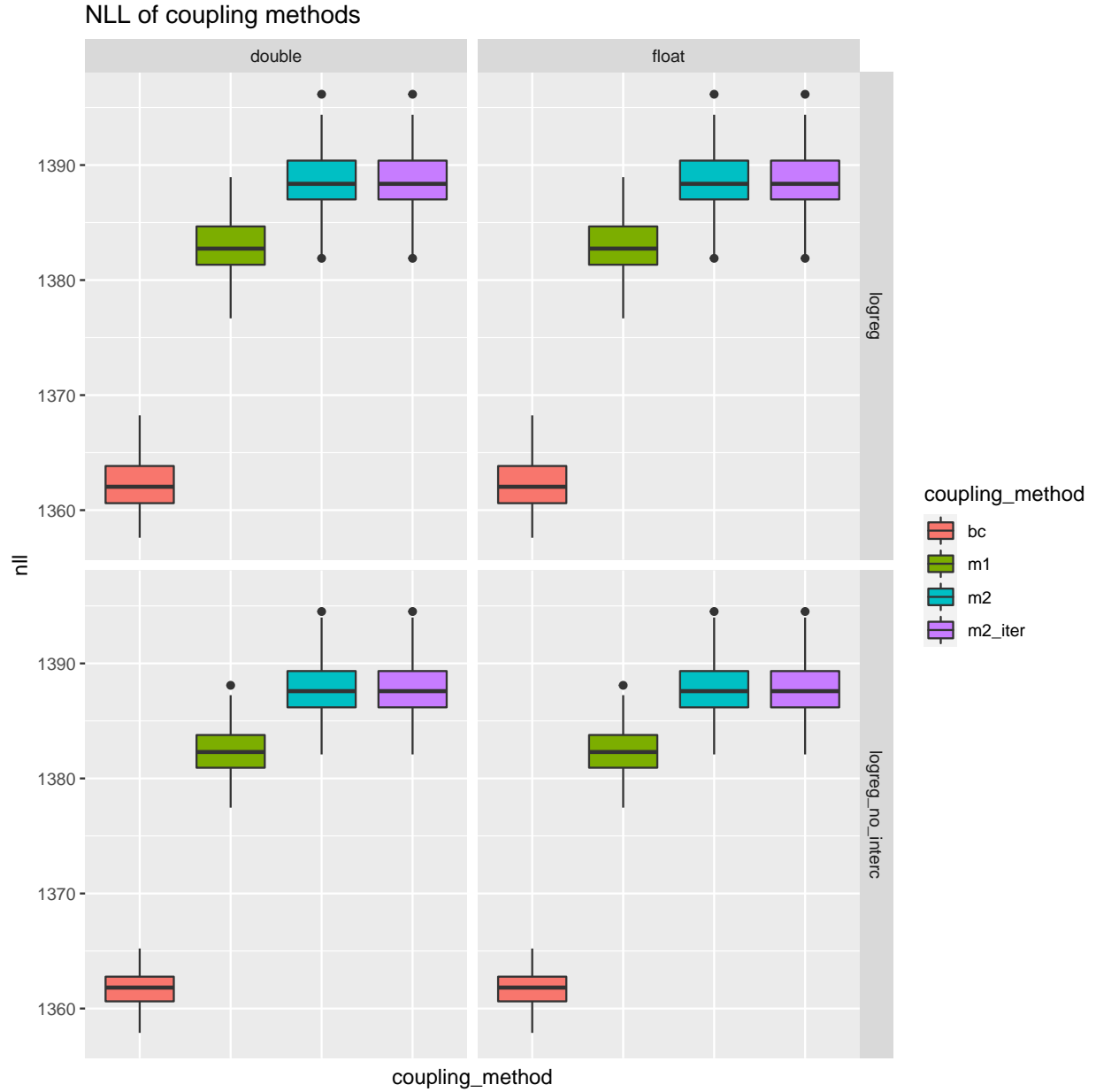
All coupling methods perform similarly, regardless of the single or double precision. A little difference is visible for bc method, where median accuracy for double precision is higher than for float precision. Also, in the case of logistic regression without intercept, bc method performs a bit worse than remaining coupling methods.

```
box2_nll <- ggplot() + geom_boxplot(data=acc_ens_subsets, mapping=aes(x=coupling_method, y=nll, fill=combining_method)) +
  facet_grid(rows=vars(combining_method), cols=vars(precision)) +
  ggtitle("NLL of coupling methods") +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank())
box2_nll
```



Bayes covariant coupling method seems to be doing much better than the remaining coupling methods when combining method lda is used. This may be due to fewer close to zero probabilities for the correct label produced by this method. As can be seen below, this behavior is present to smaller extent also in the case of logistic regression combining method.

```
box2_nll <- acc_ens_subsets %>% filter(combining_method!="lda") %>% ggplot() + geom_boxplot(mapping=aes(
  facet_grid(rows=vars(combining_method), cols=vars(precision)) +
  ggtitle("NLL of coupling methods") +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank())
box2_nll
```



Šuch, Ondrej, and Santiago Barreda. 2016. “Bayes Covariant Multi-Class Classification.” *Pattern Recognition Letters* 84: 99–106.

Wu, Ting-Fan, Chih-Jen Lin, and Ruby C Weng. 2004. “Probability Estimates for Multi-Class Classification by Pairwise Coupling.” *Journal of Machine Learning Research* 5 (Aug): 975–1005.