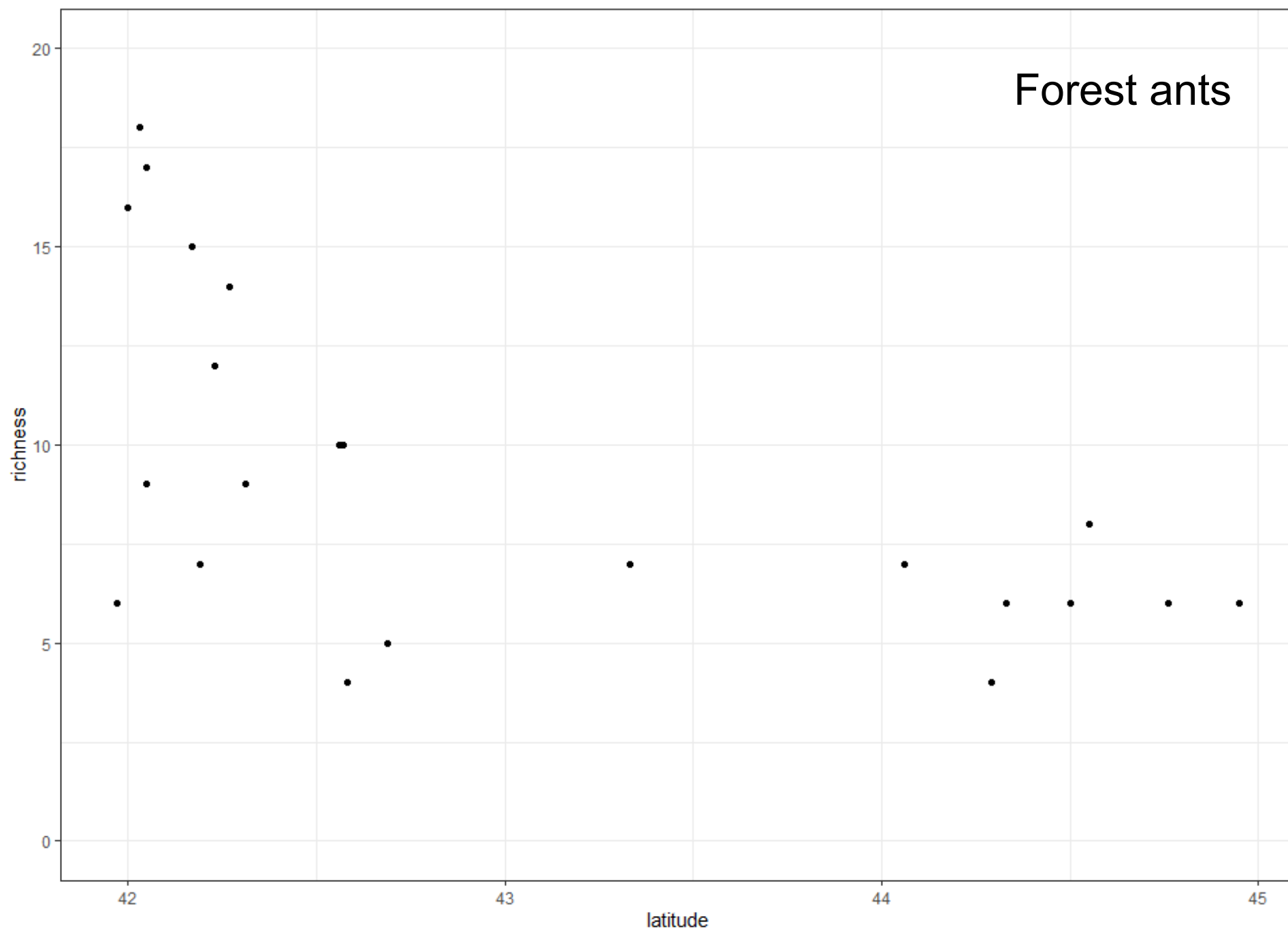
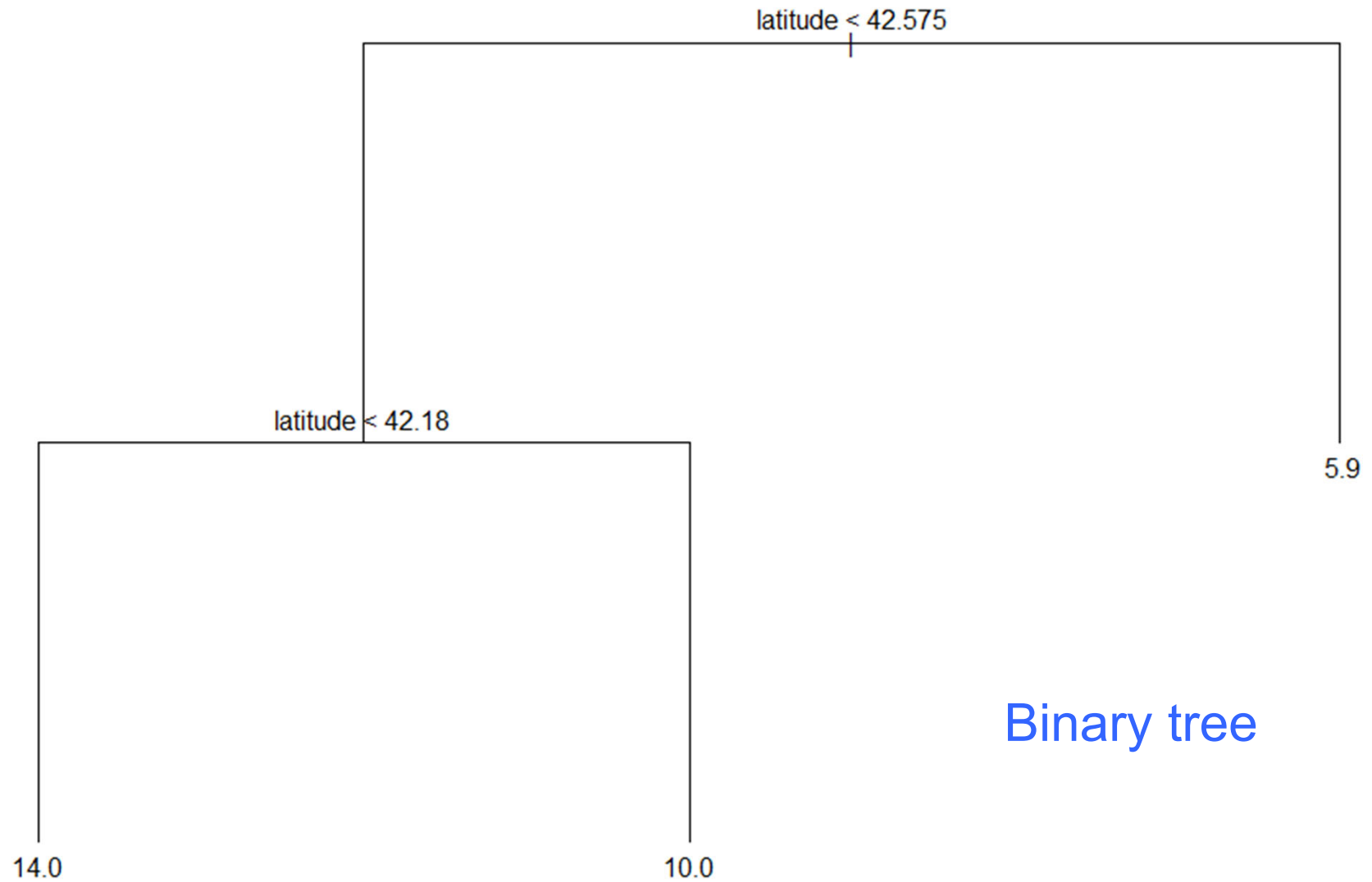


Today

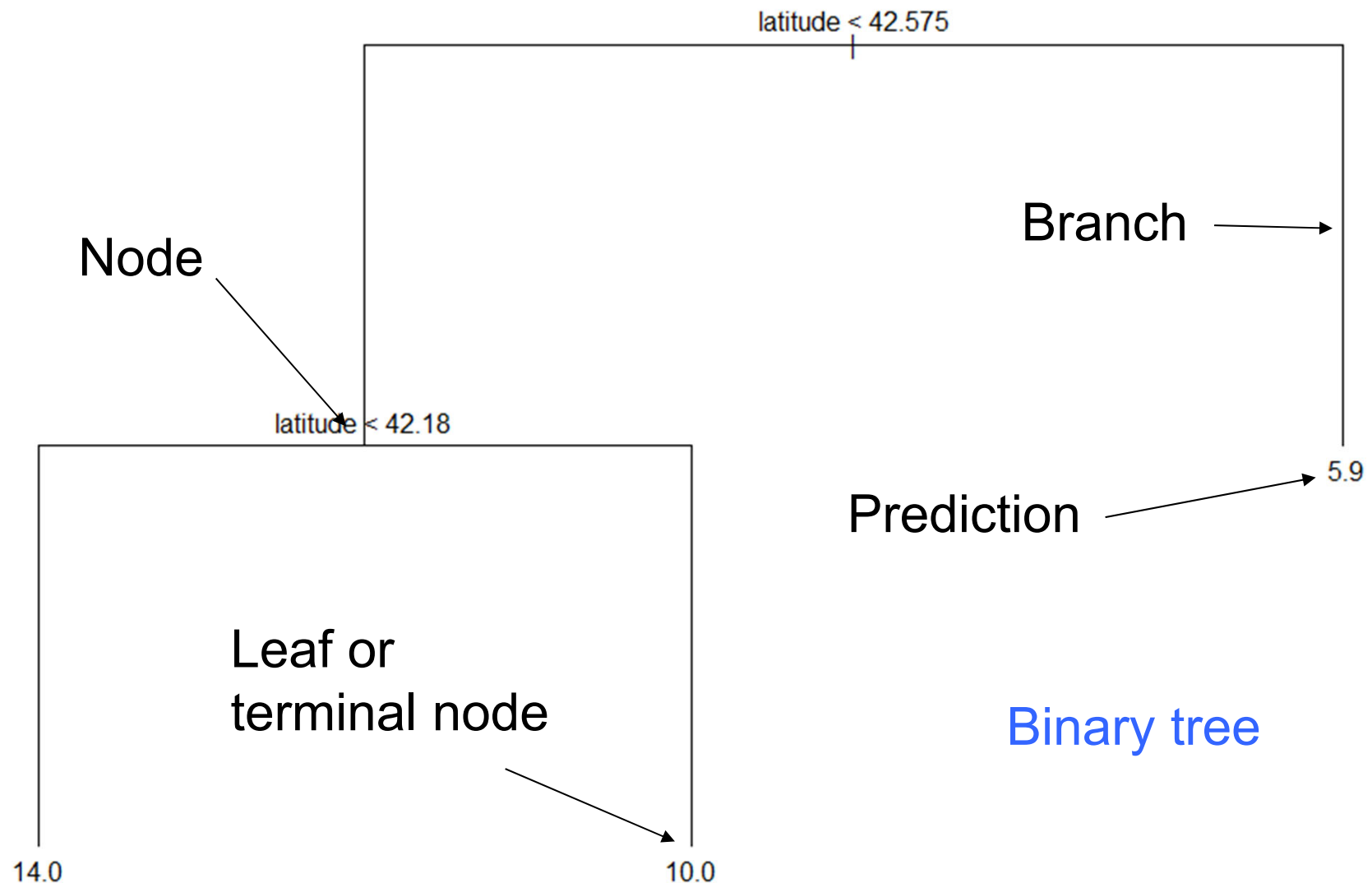
- Decision tree models
- + training and inference



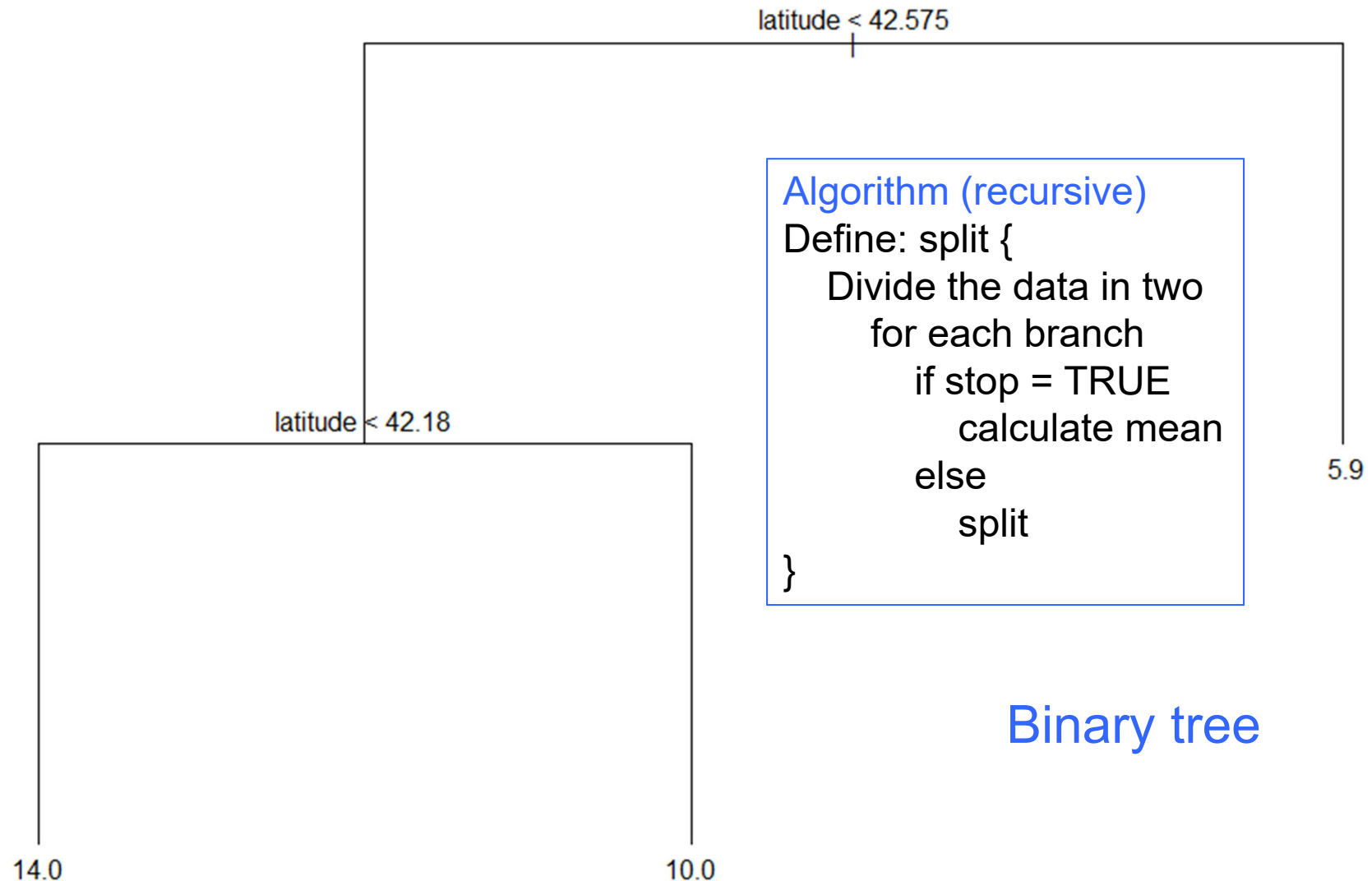
A regression tree model

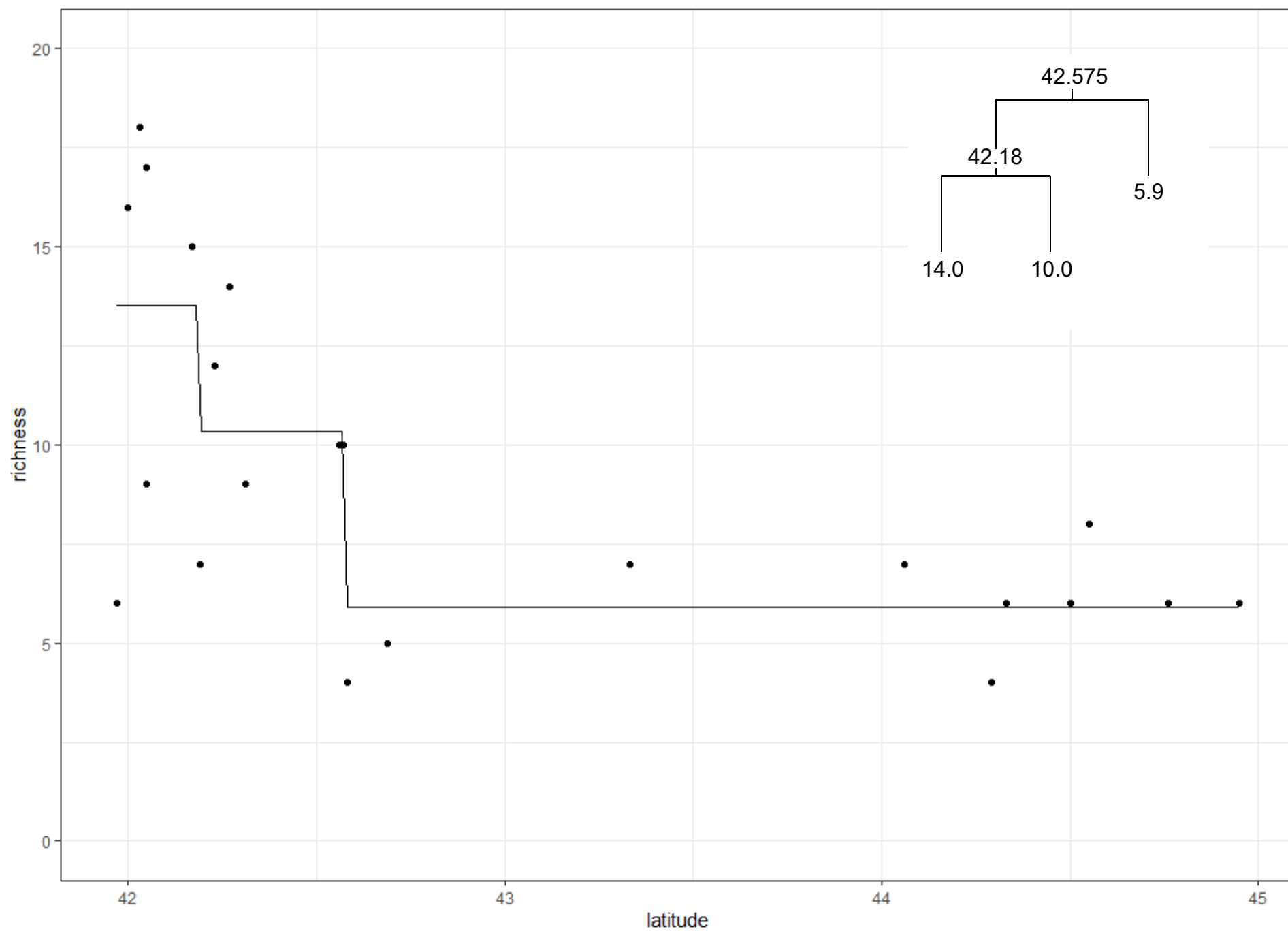


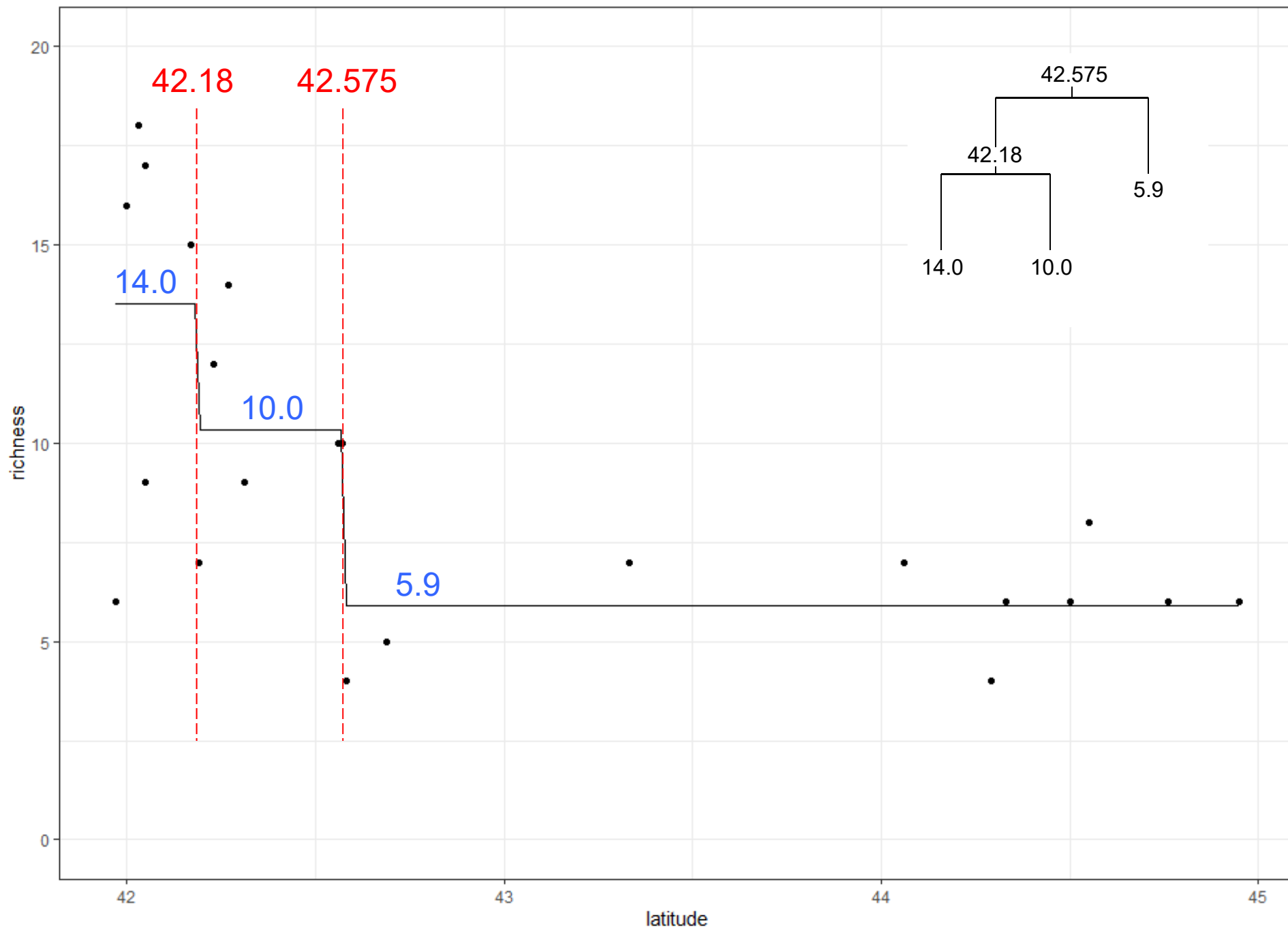
A regression tree model



A regression tree model



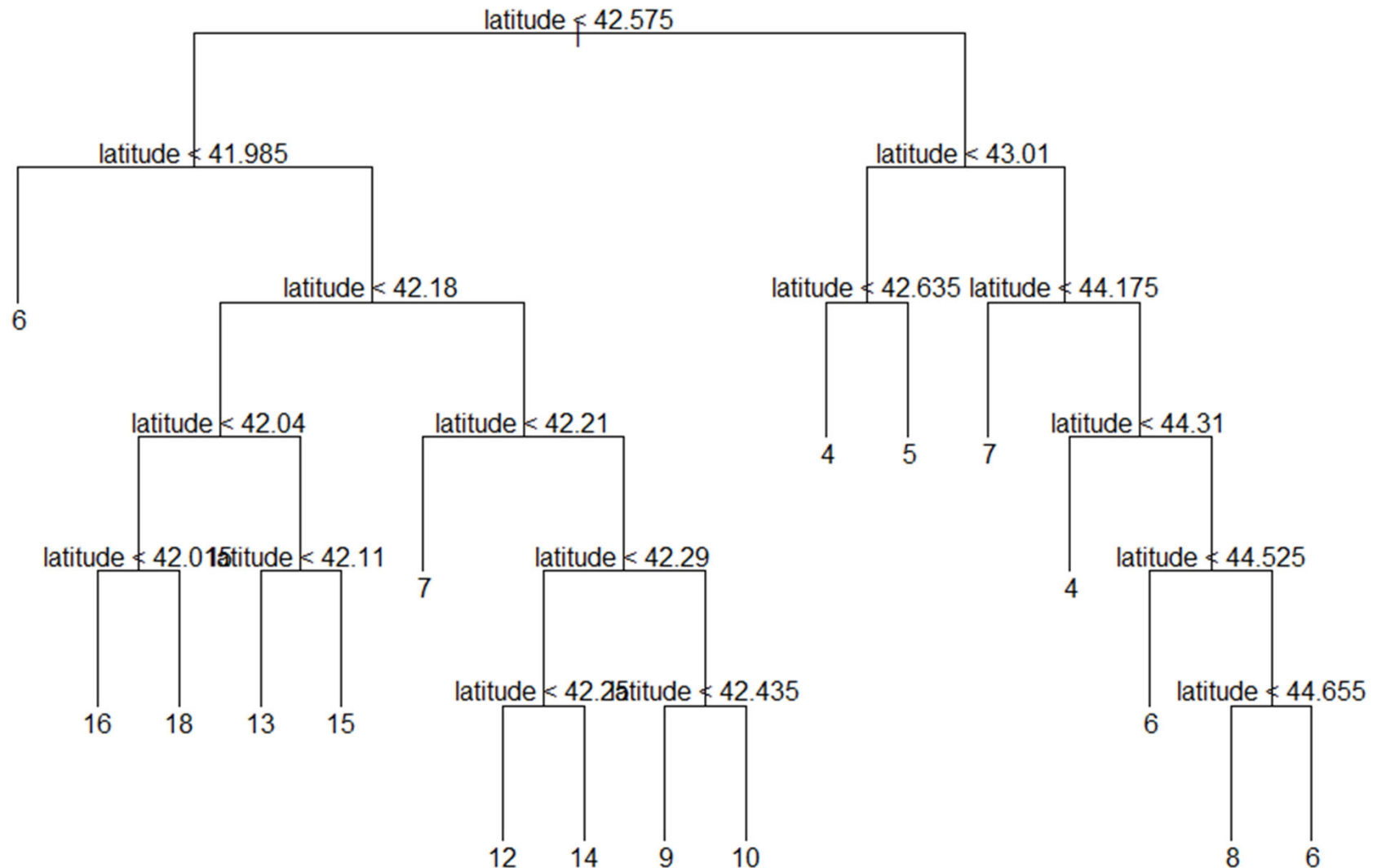


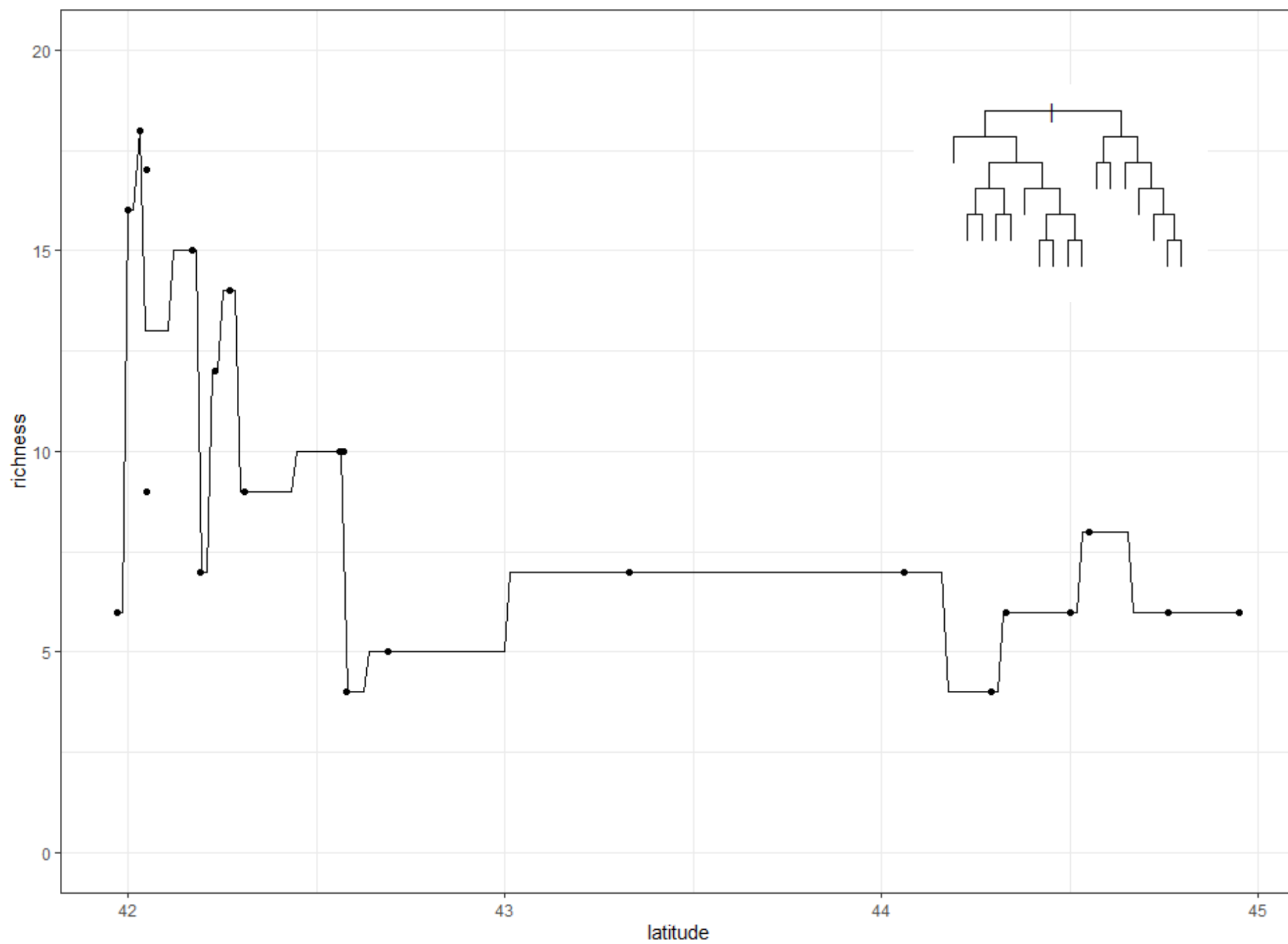


Training

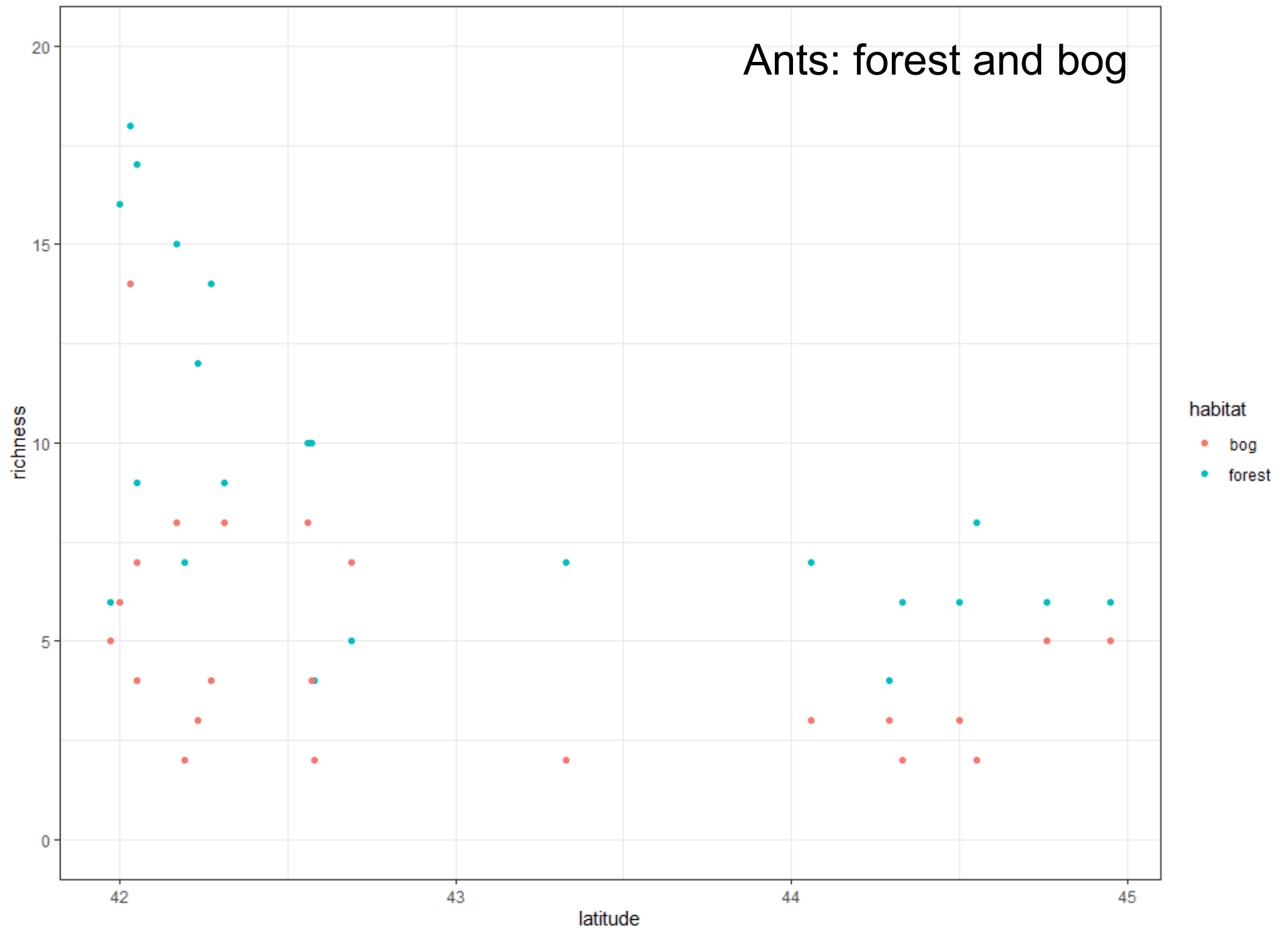
- Find splits that minimize training error
 - regression: SSQ
 - classification: Gini index or entropy
- Algorithm: binary recursive partitioning
 - split (minimize error)
 - if stop = FALSE, split again,
- Stopping rules
 - e.g. degree of error improvement, data per node

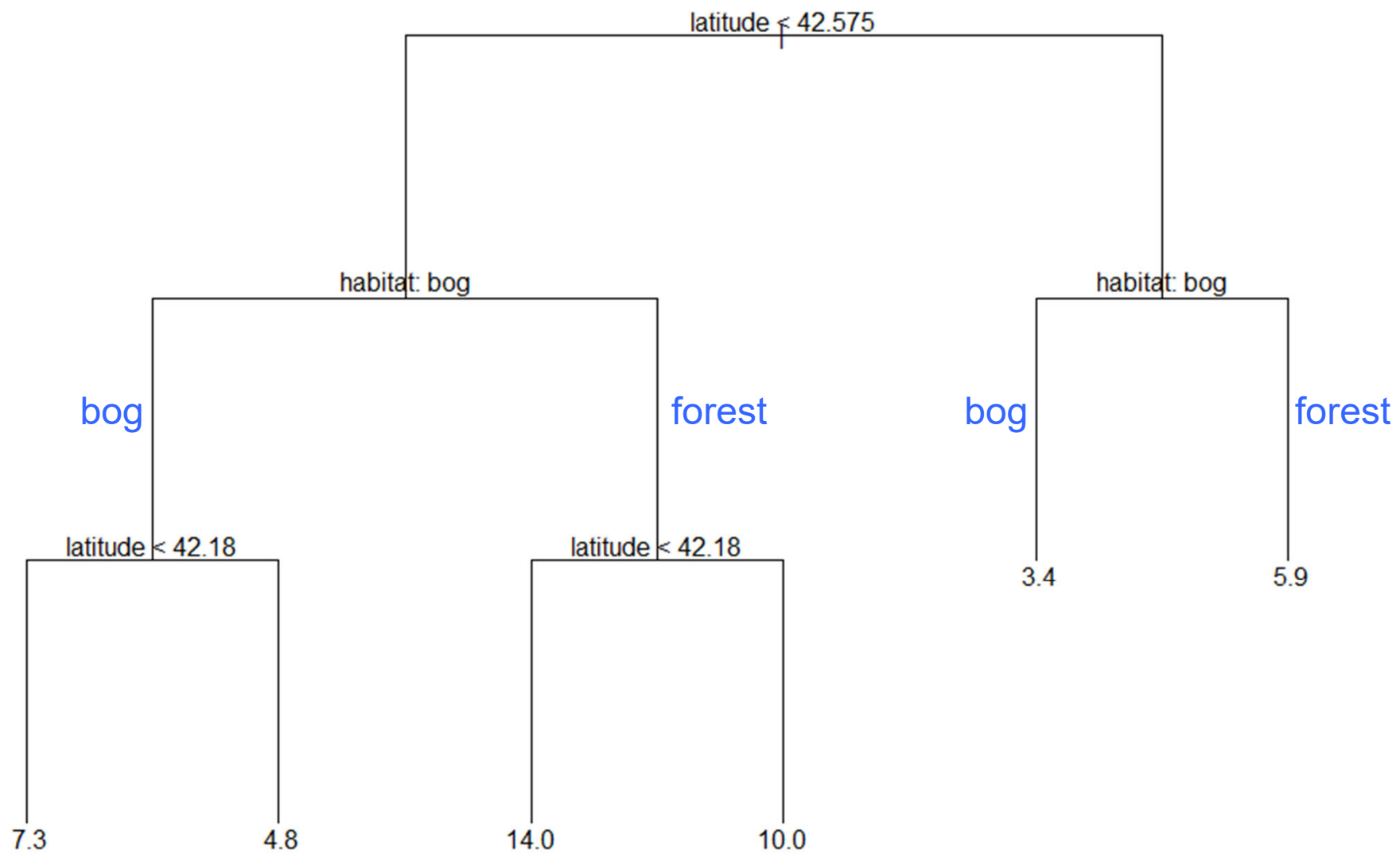
Same data, deeper tree



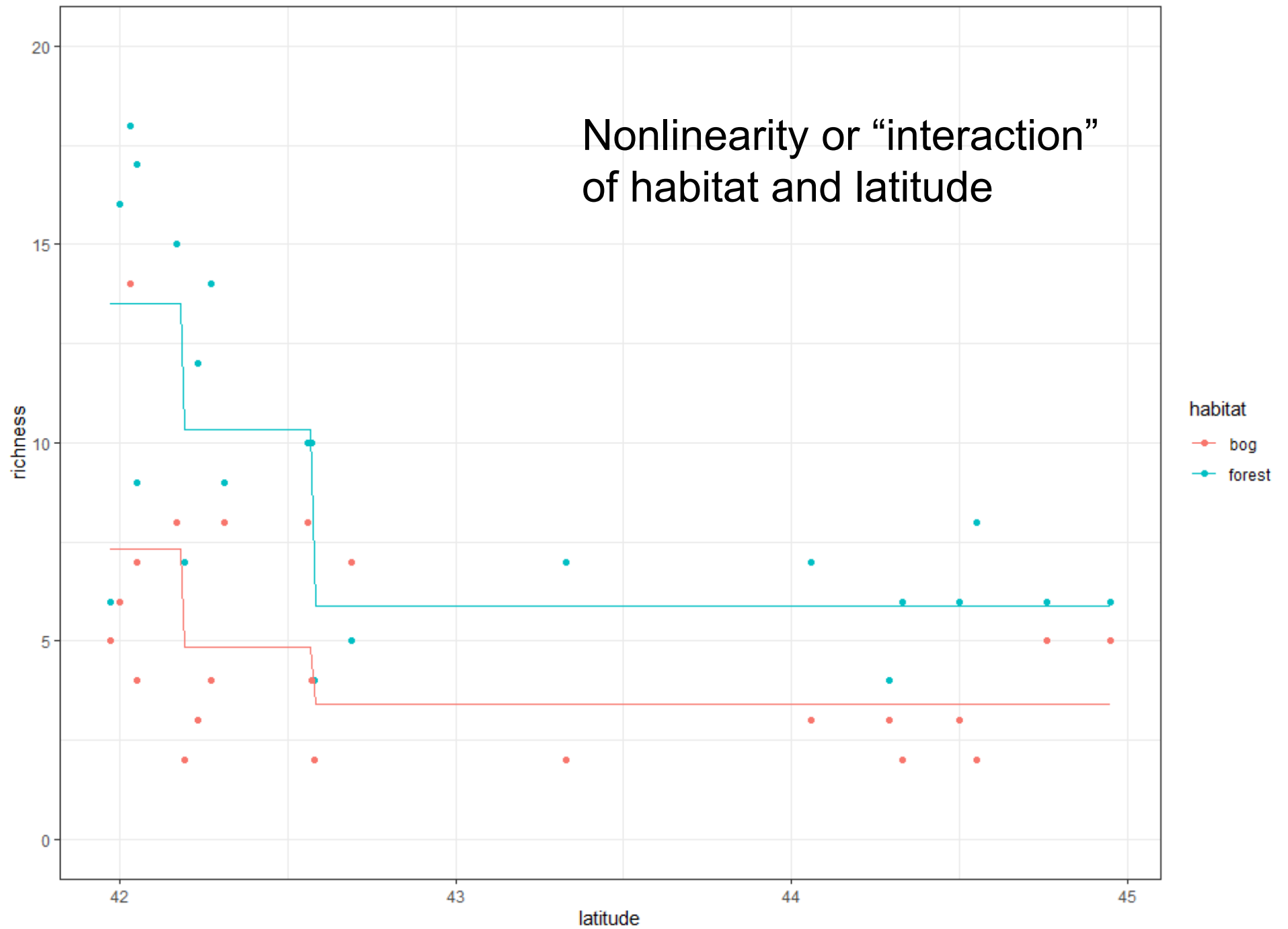


Ants: forest and bog



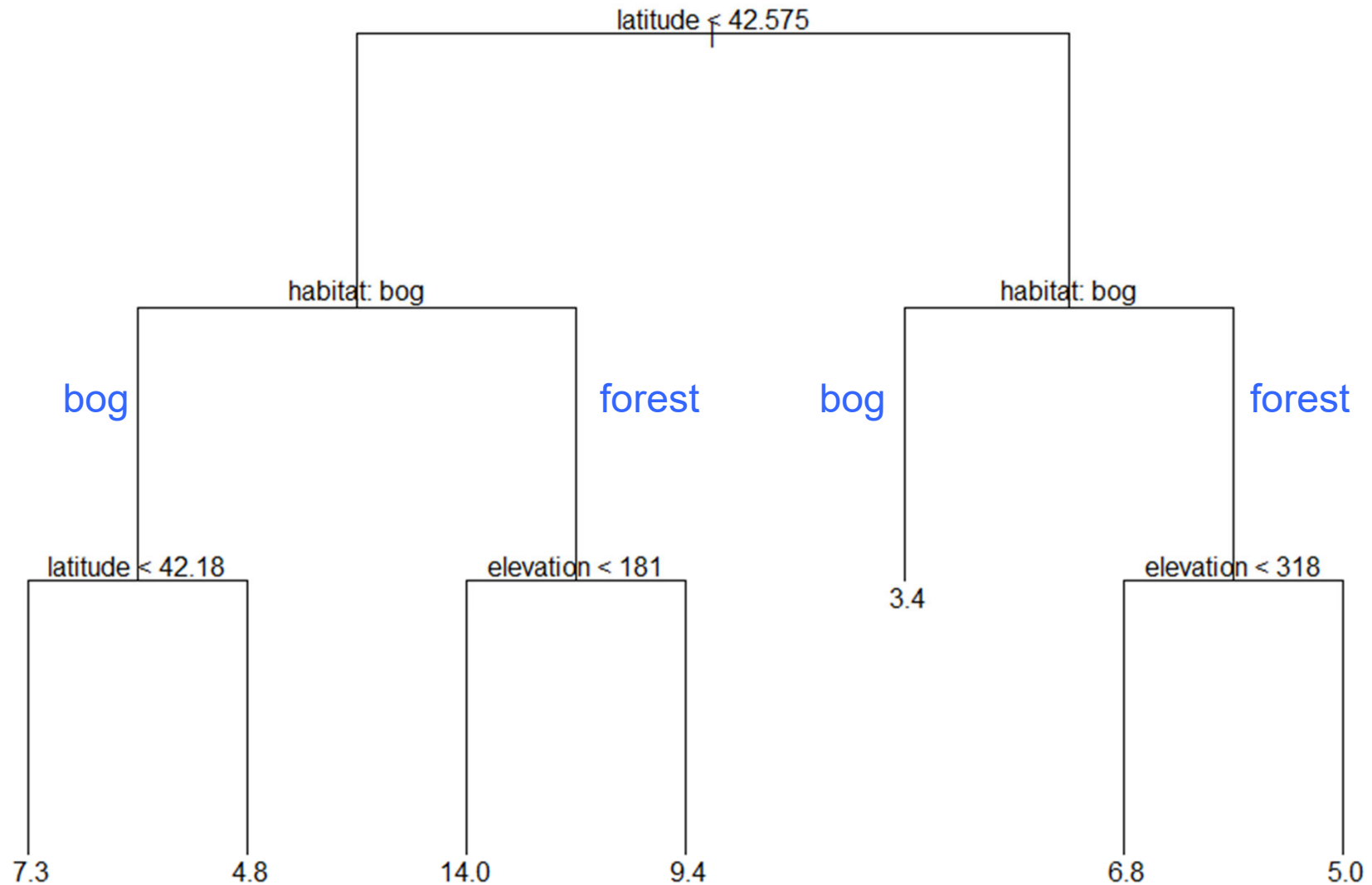


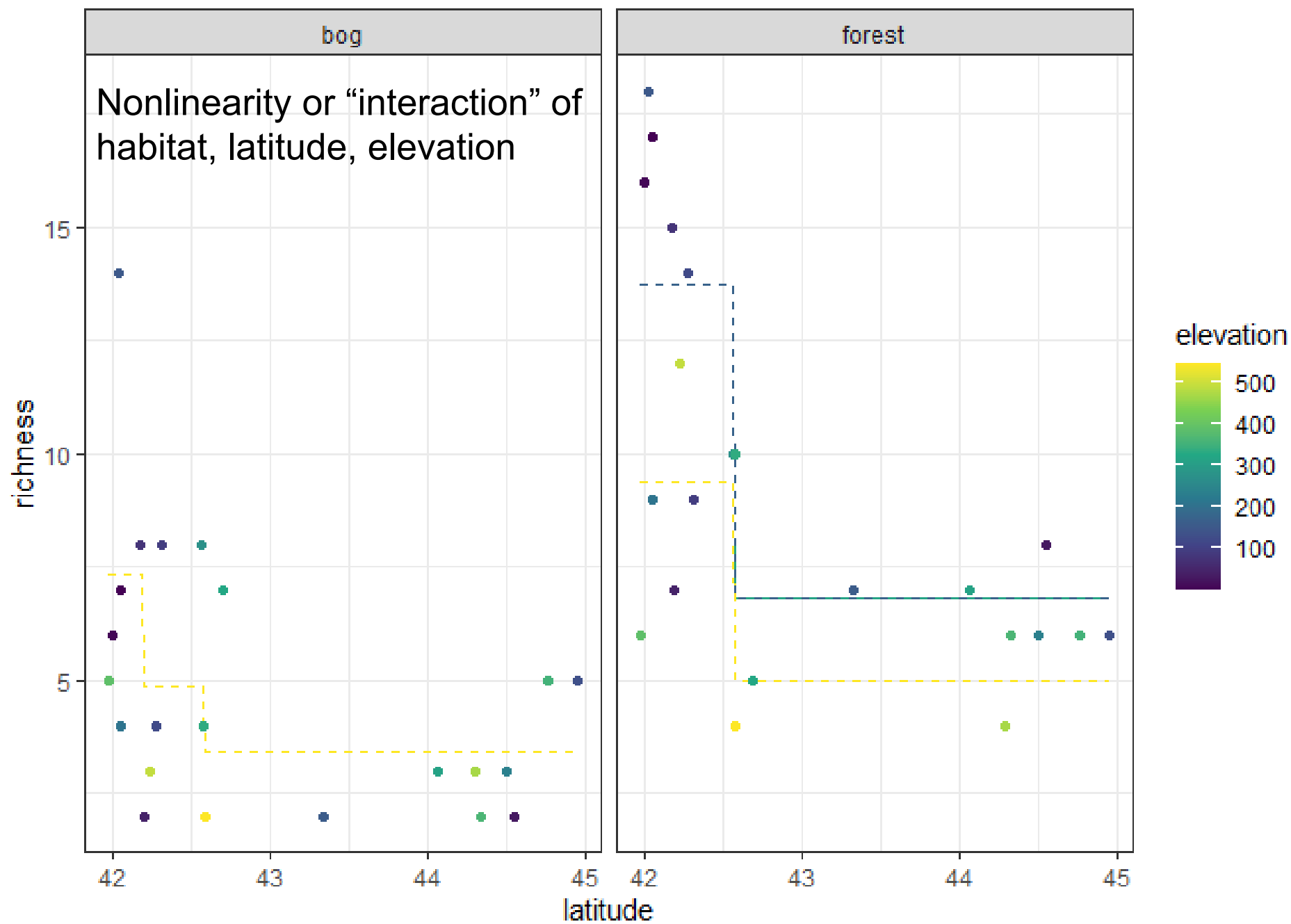
Nonlinearity or “interaction” of habitat and latitude



```
> head(ants)
  habitat latitude elevation richness
1 forest    41.97      389         6
2 forest    42.00         8        16
3 forest    42.03     152        18
4 forest    42.05         1        17
5 forest    42.05     210         9
6 forest    42.17         78        15
```

All 3 predictors





Inference

- k-fold CV
- Can also use for tree complexity
 - training: complexity penalty
 - e.g. $\text{loss} = \text{SSQ} + \alpha T$
 - where α is a tuning parameter, T is number of leaves
 - “pruning”

