Using BART with a virtual species!

Getting Started

So you're interested in using `embarcadero` to do species distribution modeling with Bayesian additive regression trees! That's great. BARTs are a powerful way to do machine learning and, while not a new method per se, they are very new for SDMs.

Most of the core functionality of `embarcadero` is actually a wrapper for `dbarts`, which runs the actual BART fitting process. This vignete will show you

1. How to run BARTs
2. Variable importance measures
3. Automated variable selection
4. Partial dependence plots
5. Visualizing the posterior distribution

There's also just going to be some general comments on the process of using BARTs, the challenges to working with them, and some things that are hopefully coming next.

If you want to install, do it using devtools for now:

```
#> Loading required package: raster
#> Loading required package: sp
#>
#> Attaching package: 'raster'
#> The following object is masked from 'package:dplyr':
#>
#>     select
#> The following object is masked from 'package:tidyr':
#>
#>     extract
#> Loading required package: dbarts
#>
#> Attaching package: 'dbarts'
#> The following object is masked from 'package:raster':
#>
#>     extract
#> The following object is masked from 'package:tidyr':
#>
#>     extract
#> Loading required package: Metrics
#> Loading required package: dismo
#> Loading required package: ROCR
#> Loading required package: gplots
#>
#> Attaching package: 'gplots'
#> The following object is masked from 'package:stats':
#>
#>     lowess
#> Loading required package: patchwork
#>
#> Attaching package: 'patchwork'
#> The following object is masked from 'package:raster':
```

```
#>
#>     area
#> Loading required package: cowplot
#>
#> Attaching package: 'cowplot'
#> The following object is masked from 'package:patchwork':
#>
#>     align_plots
#> The following object is masked from 'package:ggplot2':
#>
#>     ggsave
#> Loading required package: velox
#> Loading required package: ggpubr
#> Loading required package: magrittr
#>
#> Attaching package: 'magrittr'
#> The following object is masked from 'package:dbarts':
#>
#>     extract
#> The following object is masked from 'package:raster':
#>
#>     extract
#> The following object is masked from 'package:purrr':
#>
#>     set_names
#> The following object is masked from 'package:tidyr':
#>
#>     extract
#>
#> Attaching package: 'ggpubr'
#> The following object is masked from 'package:cowplot':
#>
#>     get_legend
#> The following object is masked from 'package:raster':
#>
#>     rotate
#> Loading required package: matrixStats
#>
#> Attaching package: 'matrixStats'
#> The following object is masked from 'package:dplyr':
#>
#>     count
#> Loading required package: data.table
#>
#> Attaching package: 'data.table'
#> The following object is masked from 'package:raster':
#>
#>     shift
#> The following objects are masked from 'package:dplyr':
#>
#>     between, first, last
#> The following object is masked from 'package:purrr':
#>
#>     transpose
```

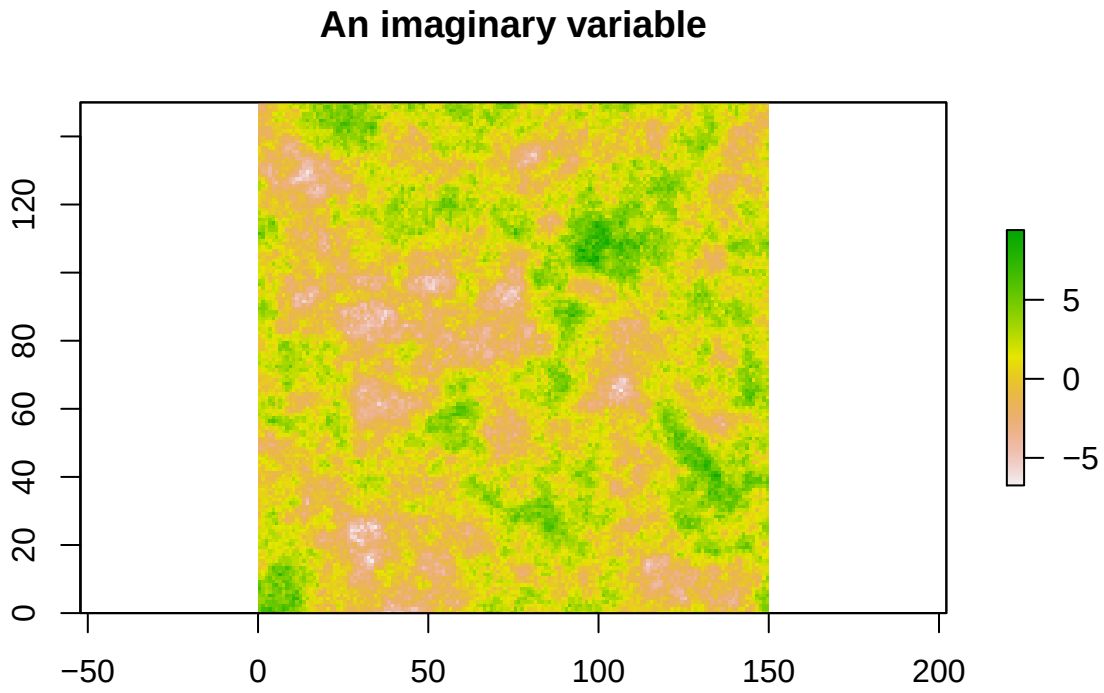
```
#>
#> Attaching package: 'embarcadero'
#> The following object is masked from 'package:purrr':
#>
#>     partial
```

Doors are closing; please stand clear of the doors.

Creating the virtual species

First, let's create an imaginary landscape. We do this using the NLMR package:

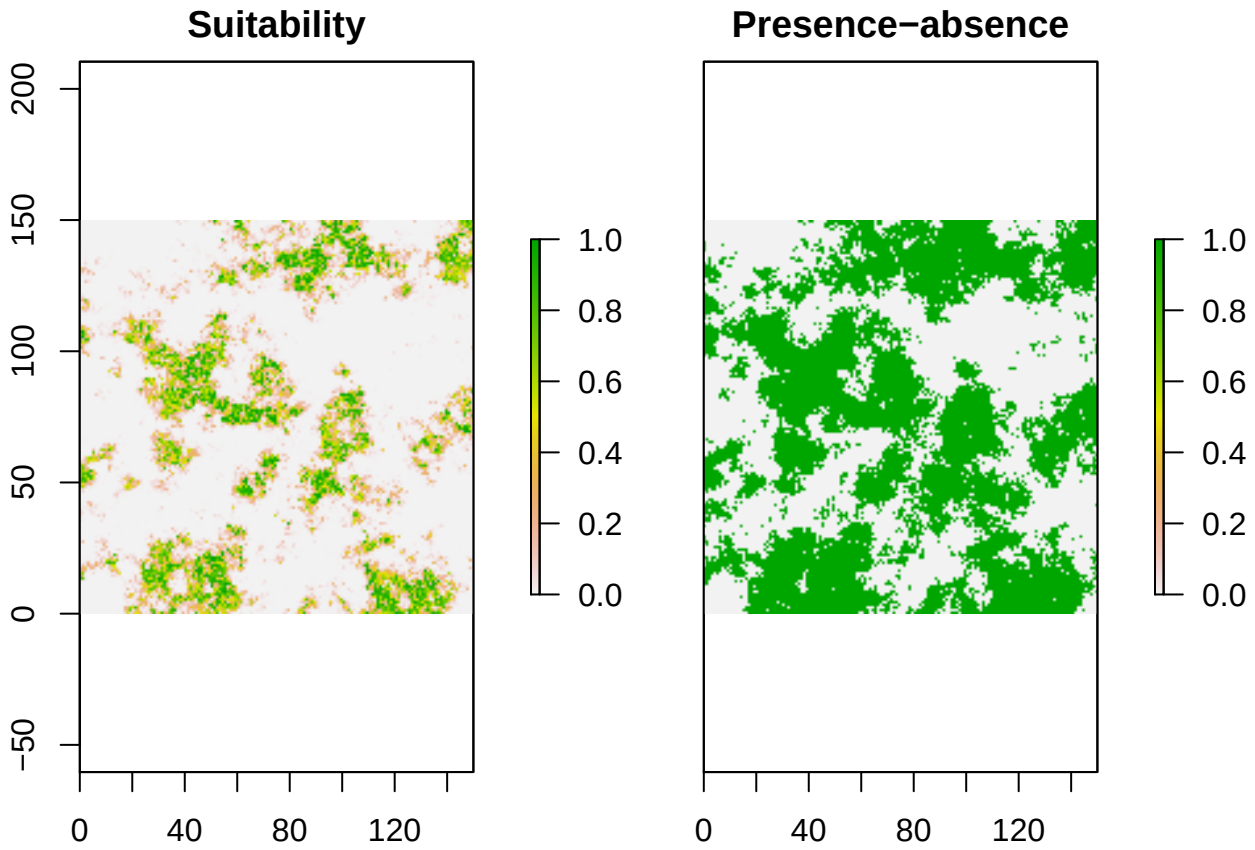
```
onelandscape <- function(x) {NLMR::nlm_gaussianfield(nrow = 150,
                                                     ncol = 150,
                                                     rescale = FALSE)}
climate <- stack(lapply(c(1:8), onelandscape))
names(climate) <- c('x1','x2','x3','x4','x5','x6','x7','x8')

plot(climate[[1]],main='An imaginary variable')
```



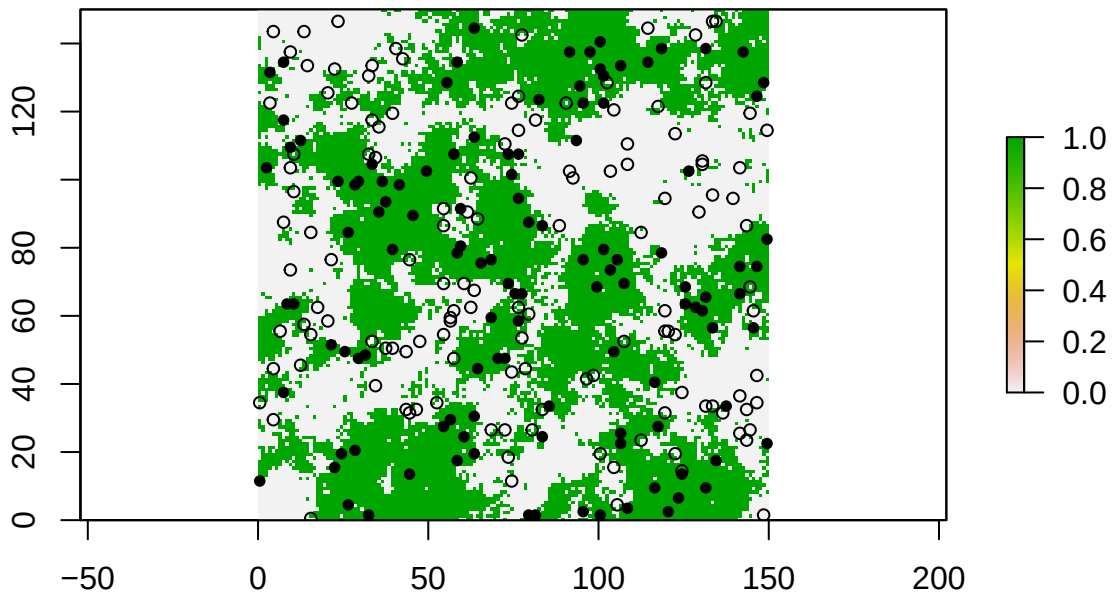
Next, let's make a species using the 'virtualspecies' package. Our imaginary species will only responds to variables 1-4, making variables 5-8 uninformative predictors (hopefully our model will drop them):

```
# Generate the species' climatic niche

random.sp <- generateRandomSp(climate[[1:4]],
```

```
                              # ^ These are the informative predictors
                              approach="pca",
                              relations='gaussian',
                              species.prevalence=0.5,
                              realistic.sp = TRUE,
                              PA.method='threshold')
#>  - Perfoming the pca
#>  - Defining the response of the species along PCA axes
#>  - Calculating suitability values
#>    The final environmental suitability was rescaled between 0 and 1.
#>                    To disable, set argument rescale = FALSE
#>  - Converting into Presence - Absence
#>    Threshold conversion finished:
#>
#> - cutoff = 0.00103681027010919
#> - species prevalence =0.5
```




```
# Generate some presences, and some absences, with imperfect detection

sp.points <- sampleOccurrences(random.sp,
                               n=250,
                               type = 'presence-absence')
```




```
# Extract the associated climate values

occ <- SpatialPoints(sp.points$sample.points[,c('x','y')])
occ.df <- cbind(sp.points$sample.points,
                raster::extract(climate, occ))

# Finally, let's drop the long-lats and the "Real" ground truthed presence-absence values, and just lea

occ.df <- occ.df[,-c(1:3)]
```

Alright. Now that we have the dataset, let's get to modeling!

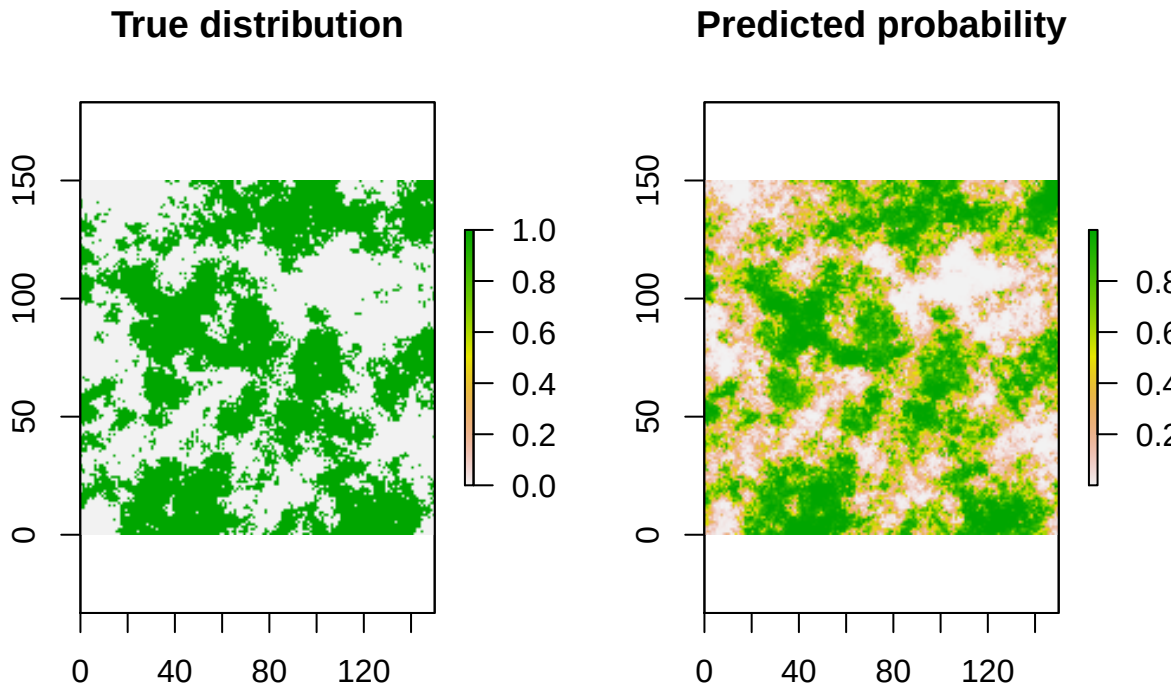
We could easily throw all our data in one model, run it on defaults, make a map, and never think about it again. There's no laws against it.

```
# Train the model
sdm <- bart(y.train=occ.df[,'Observed'],
            x.train=occ.df[,c(2:9)],
            keeptrees = TRUE) # It's very important this is set to TRUE
#>
#> Running BART with binary y
#>
#> number of trees: 200
#> number of chains: 1, number of threads 1
#> Prior:
#>  k: 2.000000
#>  power and base for tree prior: 2.000000 0.950000
```

```
#>  use quantiles for rule cut points: false
#> data:
#>  number of training observations: 250
#>  number of test observations: 0
#>  number of explanatory variables: 8
#>
#> Cutoff rules c in x<=c vs x>c
#> Number of cutoffs: (var: number of possible c):
#> (1: 100) (2: 100) (3: 100) (4: 100) (5: 100)
#> (6: 100) (7: 100) (8: 100)
#>
#> offsets:
#>  reg : 0.00 0.00 0.00 0.00 0.00
#> Running mcmc loop:
#> iteration: 100 (of 1000)
#> iteration: 200 (of 1000)
#> iteration: 300 (of 1000)
#> iteration: 400 (of 1000)
#> iteration: 500 (of 1000)
#> iteration: 600 (of 1000)
#> iteration: 700 (of 1000)
#> iteration: 800 (of 1000)
#> iteration: 900 (of 1000)
#> iteration: 1000 (of 1000)
#> total seconds in loop: 3.233851
#>
#> Tree sizes, last iteration:
#> [1] 3 3 2 1 3 3 3 1 2 3 1 2 2 3 2 2 3 2
#> 2 1 2 3 2 3 2 3 2 2 2 1 2 2 3 4 1 2 2 3
#> 2 2 2 2 2 3 2 2 2 2 3 2 3 3 2 3 2 2 2 3
#> 3 2 3 3 3 2 2 4 3 2 3 2 2 2 2 3 2 3 2 3
#> 3 3 2 1 2 2 4 2 3 3 3 2 3 2 3 4 2 2 2 3
#> 1 1 3 2 2 2 3 3 2 3 2 2 1 2 3 2 2 3 2 2
#> 1 2 2 2 2 2 2 2 3 2 3 2 3 4 2 2 2 3 3 2
#> 3 2 3 2 2 3 2 1 2 1 2 3 3 2 2 3 2 2 2 2
#> 2 2 2 2 1 3 2 4 3 2 3 3 2 2 3 3 1 2 2 2
#> 2 3 2 2 2 4 2 2 2 2 3 3 3 2 2 3 2 2 2 2
#> 2 2
#>
#> Variable Usage, last iteration (var:count):
#> (1: 33) (2: 27) (3: 29) (4: 42) (5: 22)
#> (6: 40) (7: 37) (8: 33)
#> DONE BART

# Predict the species distribution!
map <- predict(sdm, climate, splitby=5)
#> Estimated time to total prediction (mins):
#> 1.111144
#>
  |
  |                                                                      |   0%
  |
  |=============                                                         |  20%
```

```
  |
  |=========================                                             |  40%
  |
  |=================================================                     |  60%
  |
  |=========================================================             |  80%
  |
  |======================================================================| 100%
```

```
# How's it look?
par(mfrow=c(1,2))
plot(random.sp$pa.raster, main='True distribution')
plot(map, main='Predicted probability')
```



This is.... alright? We could probably do better. One of the easiest ways is to cut some variables we don't think are performing well. In BART, pushing the models towards a smaller number of trees forces the variables to compete a bit, and preferentially upweights the better ones. We can plot that!

```
# A variable importance diagnostic. What's behaving well?
# This takes a while to run normally! Drop the iter if you want a plot faster with more variance.
varimp.diag(occ.df[,2:9], occ.df[,'Observed'], iter=50)
#>
#>  10 tree models: 50 iterations
#>
  |
  |                                                                       |   0%
```

```
|
|=                                                                |   2%
|
|===                                                              |   4%
|
|====                                                             |   6%
|
|=====                                                            |   8%
|
|=======                                                          |  10%
|
|========                                                         |  12%
|
|=========                                                        |  14%
|
|==========                                                       |  16%
|
|============                                                     |  18%
|
|=============                                                    |  20%
|
|==============                                                   |  22%
|
|================                                                 |  24%
|
|=================                                                |  26%
|
|===================                                              |  28%
|
|====================                                             |  30%
|
|======================                                           |  32%
|
|=======================                                          |  34%
|
|========================                                         |  36%
|
|========================                                         |  38%
|
|==========================                                       |  40%
|
|===========================                                      |  42%
|
|============================                                     |  44%
|
|==============================                                   |  46%
|
|===============================                                  |  48%
|
|================================                                 |  50%
|
|==================================                               |  52%
|
```

```
  |=================================                                |  54%
  |
  |==================================                               |  56%
  |
  |===================================                              |  58%
  |
  |====================================                             |  60%
  |
  |=====================================                            |  62%
  |
  |======================================                           |  64%
  |
  |=======================================                          |  66%
  |
  |========================================                         |  68%
  |
  |=========================================                        |  70%
  |
  |==========================================                       |  72%
  |
  |===========================================                      |  74%
  |
  |============================================                     |  76%
  |
  |=============================================                    |  78%
  |
  |==============================================                   |  80%
  |
  |===============================================                  |  82%
  |
  |================================================                 |  84%
  |
  |=================================================                |  86%
  |
  |==================================================               |  88%
  |
  |===================================================              |  90%
  |
  |====================================================             |  92%
  |
  |=====================================================            |  94%
  |
  |======================================================           |  96%
  |
  |=======================================================          |  98%
  |
  |========================================================================| 100%
#>  20 tree models: 50 iterations
#>
  |
  |                                                                  |   0%
  |
  |=                                                                 |   2%
```

```
|
|===                                                    |   4%
|
|====                                                   |   6%
|
|=====                                                  |   8%
|
|======                                                 |  10%
|
|========                                               |  12%
|
|=========                                              |  14%
|
|==========                                             |  16%
|
|============                                           |  18%
|
|=============                                          |  20%
|
|==============                                         |  22%
|
|================                                       |  24%
|
|=================                                      |  26%
|
|==================                                     |  28%
|
|====================                                   |  30%
|
|=====================                                  |  32%
|
|======================                                 |  34%
|
|======================                                 |  36%
|
|========================                               |  38%
|
|=========================                              |  40%
|
|==========================                             |  42%
|
|============================                           |  44%
|
|============================                           |  46%
|
|==============================                         |  48%
|
|==============================                         |  50%
|
|================================                       |  52%
|
|================================                       |  54%
|
```

```
  |==================================                                    |  56%
  |
  |====================================                                  |  58%
  |
  |======================================                                |  60%
  |
  |========================================                              |  62%
  |
  |==========================================                            |  64%
  |
  |============================================                          |  66%
  |
  |==============================================                        |  68%
  |
  |================================================                      |  70%
  |
  |==================================================                    |  72%
  |
  |====================================================                  |  74%
  |
  |======================================================                |  76%
  |
  |========================================================              |  78%
  |
  |==========================================================            |  80%
  |
  |============================================================          |  82%
  |
  |==============================================================        |  84%
  |
  |================================================================      |  86%
  |
  |==================================================================    |  88%
  |
  |====================================================================  |  90%
  |
  |======================================================================|  92%
  |
  |========================================================================|  94%
  |
  |==========================================================================|  96%
  |
  |============================================================================| 98%
  |
  |==============================================================================| 100%
#>  50 tree models: 50 iterations
#>
  |
  |                                                                      |   0%
  |
  |=                                                                     |   2%
  |
  |===                                                                   |   4%
```

```
|
|====                                                              |   6%
|
|=====                                                             |   8%
|
|======                                                            |  10%
|
|========                                                          |  12%
|
|=========                                                         |  14%
|
|==========                                                        |  16%
|
|============                                                      |  18%
|
|=============                                                     |  20%
|
|==============                                                    |  22%
|
|================                                                  |  24%
|
|==================                                                |  26%
|
|===================                                               |  28%
|
|====================                                              |  30%
|
|======================                                            |  32%
|
|=======================                                           |  34%
|
|========================                                          |  36%
|
|========================                                          |  38%
|
|=========================                                         |  40%
|
|==========================                                        |  42%
|
|============================                                      |  44%
|
|=============================                                     |  46%
|
|==============================                                    |  48%
|
|================================                                  |  50%
|
|=================================                                 |  52%
|
|==================================                                |  54%
|
|===================================                               |  56%
|
```

```
  |==================================                                    |  58%
  |
  |====================================                                  |  60%
  |
  |======================================                                |  62%
  |
  |========================================                              |  64%
  |
  |==========================================                            |  66%
  |
  |============================================                          |  68%
  |
  |==============================================                        |  70%
  |
  |================================================                      |  72%
  |
  |==================================================                    |  74%
  |
  |====================================================                  |  76%
  |
  |======================================================                |  78%
  |
  |========================================================              |  80%
  |
  |==========================================================            |  82%
  |
  |============================================================          |  84%
  |
  |==============================================================        |  86%
  |
  |================================================================      |  88%
  |
  |==================================================================    |  90%
  |
  |====================================================================  |  92%
  |
  |======================================================================|  94%
  |
  |========================================================================|  96%
  |
  |==========================================================================|  98%
  |
  |============================================================================| 100%
#>  100 tree models: 50 iterations
#>
  |
  |                                                                      |   0%
  |
  |=                                                                     |   2%
  |
  |===                                                                   |   4%
  |
  |====                                                                  |   6%
```

```
|
|=====                                                                 |   8%
|
|======                                                                |  10%
|
|========                                                              |  12%
|
|=========                                                             |  14%
|
|==========                                                            |  16%
|
|============                                                          |  18%
|
|==============                                                        |  20%
|
|==============                                                        |  22%
|
|================                                                      |  24%
|
|==================                                                    |  26%
|
|====================                                                  |  28%
|
|======================                                                |  30%
|
|======================                                                |  32%
|
|========================                                              |  34%
|
|========================                                              |  36%
|
|==========================                                            |  38%
|
|==========================                                            |  40%
|
|============================                                          |  42%
|
|============================                                          |  44%
|
|==============================                                        |  46%
|
|================================                                      |  48%
|
|================================                                      |  50%
|
|==================================                                    |  52%
|
|==================================                                    |  54%
|
|====================================                                  |  56%
|
|====================================                                  |  58%
|
```

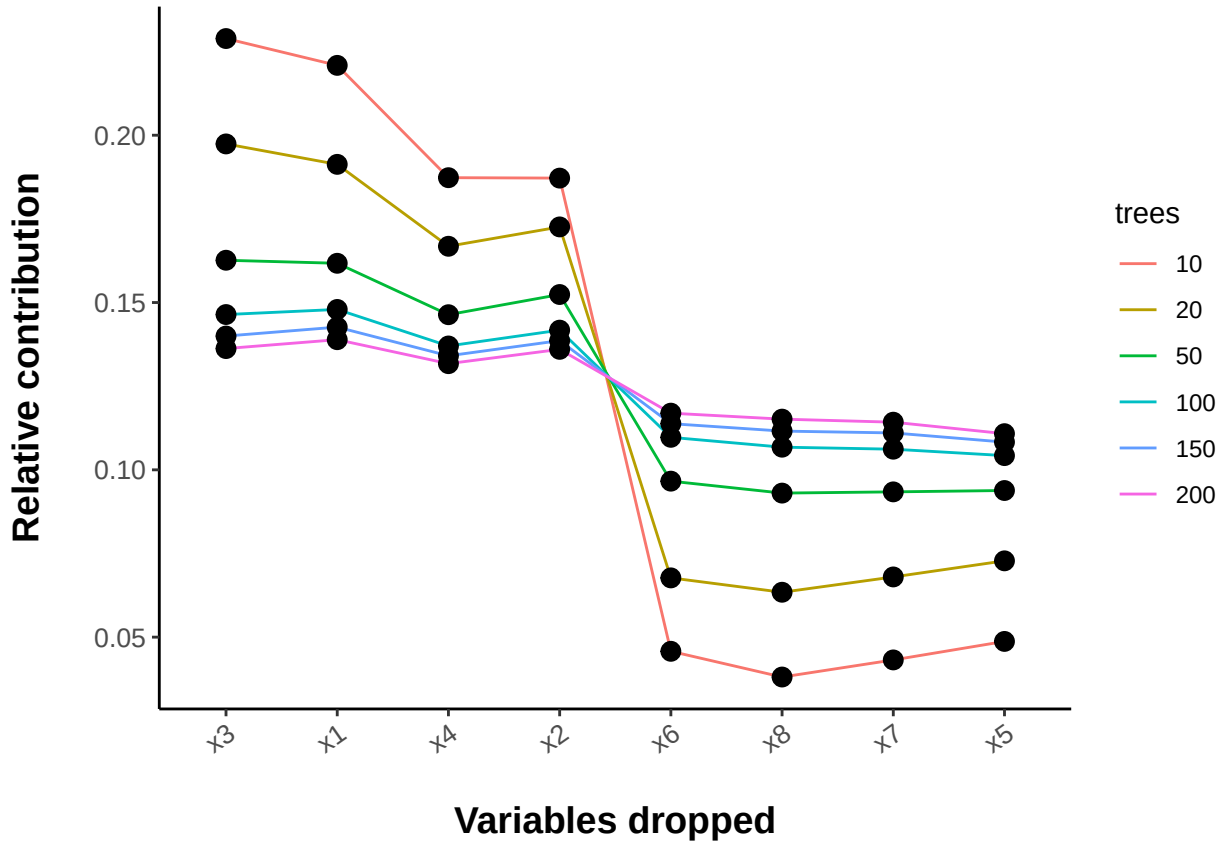
```
  |=================================                     |  60%
  |
  |==================================                    |  62%
  |
  |===================================                   |  64%
  |
  |====================================                  |  66%
  |
  |=====================================                 |  68%
  |
  |======================================                |  70%
  |
  |=======================================               |  72%
  |
  |========================================              |  74%
  |
  |=========================================             |  76%
  |
  |==========================================            |  78%
  |
  |===========================================           |  80%
  |
  |============================================          |  82%
  |
  |=============================================         |  84%
  |
  |==============================================        |  86%
  |
  |===============================================       |  88%
  |
  |================================================      |  90%
  |
  |=================================================     |  92%
  |
  |==================================================    |  94%
  |
  |===================================================   |  96%
  |
  |====================================================  |  98%
  |
  |======================================================| 100%
#>  150 tree models: 50 iterations
#>
  |
  |                                                      |   0%
  |
  |=                                                     |   2%
  |
  |===                                                   |   4%
  |
  |====                                                  |   6%
  |
  |=====                                                 |   8%
```

```
|
|======                                                 |  10%
|
|========                                               |  12%
|
|=========                                              |  14%
|
|==========                                             |  16%
|
|============                                           |  18%
|
|=============                                          |  20%
|
|===============                                        |  22%
|
|================                                       |  24%
|
|==================                                     |  26%
|
|===================                                    |  28%
|
|====================                                   |  30%
|
|======================                                 |  32%
|
|======================                                 |  34%
|
|========================                               |  36%
|
|=========================                              |  38%
|
|===========================                            |  40%
|
|============================                           |  42%
|
|==============================                         |  44%
|
|===============================                        |  46%
|
|================================                       |  48%
|
|==================================                     |  50%
|
|===================================                    |  52%
|
|====================================                   |  54%
|
|======================================                 |  56%
|
|=======================================                |  58%
|
|========================================               |  60%
|
```

```
  |=====================================                                |  62%
  |
  |======================================                               |  64%
  |
  |=======================================                              |  66%
  |
  |========================================                             |  68%
  |
  |=========================================                            |  70%
  |
  |==========================================                           |  72%
  |
  |===========================================                          |  74%
  |
  |============================================                         |  76%
  |
  |=============================================                        |  78%
  |
  |==============================================                       |  80%
  |
  |===============================================                      |  82%
  |
  |================================================                     |  84%
  |
  |=================================================                    |  86%
  |
  |==================================================                   |  88%
  |
  |===================================================                  |  90%
  |
  |====================================================                 |  92%
  |
  |=====================================================                |  94%
  |
  |======================================================               |  96%
  |
  |======================================================================|  98%
  |
  |======================================================================| 100%
#>  200 tree models: 50 iterations
#>
  |
  |                                                                      |   0%
  |
  |=                                                                     |   2%
  |
  |===                                                                   |   4%
  |
  |====                                                                  |   6%
  |
  |=====                                                                 |   8%
  |
  |======                                                                |  10%
```

```
|
|========                                                 |  12%
|
|=========                                                |  14%
|
|==========                                               |  16%
|
|============                                             |  18%
|
|==============                                           |  20%
|
|==============                                           |  22%
|
|================                                         |  24%
|
|=================                                        |  26%
|
|===================                                      |  28%
|
|====================                                     |  30%
|
|======================                                   |  32%
|
|======================                                   |  34%
|
|========================                                 |  36%
|
|=========================                                |  38%
|
|=========================                                |  40%
|
|============================                             |  42%
|
|============================                             |  44%
|
|==============================                           |  46%
|
|==============================                           |  48%
|
|================================                         |  50%
|
|================================                         |  52%
|
|==================================                       |  54%
|
|==================================                       |  56%
|
|====================================                     |  58%
|
|====================================                     |  60%
|
|======================================                   |  62%
|
```

```
|=======================================                          |  64%
|
|========================================                         |  66%
|
|=========================================                        |  68%
|
|==========================================                       |  70%
|
|===========================================                      |  72%
|
|============================================                     |  74%
|
|=============================================                    |  76%
|
|==============================================                   |  78%
|
|===============================================                  |  80%
|
|================================================                 |  82%
|
|=================================================                |  84%
|
|==================================================               |  86%
|
|===================================================              |  88%
|
|====================================================             |  90%
|
|=====================================================            |  92%
|
|======================================================           |  94%
|
|=======================================================          |  96%
|
|======================================================== |  98%
|
|=========================================================|  100%
```



OK. This would suggest that some of the variables, like x5 and x7, are less important while x3 and x4 are more important. But that doesn't give us a concrete list of what to drop. Luckily, embarcadero has a stepwise variable set reduction function. Let's run that, and then retrain the model.

```
# Stepwise variable set reduction
step.model <- variable.step(x.data=occ.df[,2:9],
                            y.data=occ.df['Observed'])
#> [1] Number of variables included: 8
#> [1] Dropped:
#> [1]
#>
  |
  |                                                                      |   0%
  |
  |=                                                                     |   2%
  |
  |===                                                                   |   4%
  |
  |====                                                                  |   6%
  |
  |=====                                                                 |   8%
  |
  |======                                                                |  10%
  |
  |========                                                              |  12%
  |
  |=========                                                             |  14%
```

```
|
|==========                                                     |  16%
|
|============                                                   |  18%
|
|==============                                                 |  20%
|
|==============                                                 |  22%
|
|================                                               |  24%
|
|==================                                             |  26%
|
|====================                                           |  28%
|
|======================                                         |  30%
|
|========================                                       |  32%
|
|========================                                       |  34%
|
|==========================                                     |  36%
|
|============================                                   |  38%
|
|============================                                   |  40%
|
|==============================                                 |  42%
|
|================================                               |  44%
|
|================================                               |  46%
|
|==================================                             |  48%
|
|==================================                             |  50%
|
|====================================                           |  52%
|
|======================================                         |  54%
|
|======================================                         |  56%
|
|========================================                       |  58%
|
|==========================================                     |  60%
|
|==========================================                     |  62%
|
|============================================                   |  64%
|
|============================================                   |  66%
|
```

```
  |=========================================                        |  68%
  |
  |==========================================                       |  70%
  |
  |===========================================                      |  72%
  |
  |============================================                     |  74%
  |
  |=============================================                    |  76%
  |
  |==============================================                   |  78%
  |
  |===============================================                  |  80%
  |
  |================================================                 |  82%
  |
  |=================================================                |  84%
  |
  |==================================================               |  86%
  |
  |===================================================              |  88%
  |
  |====================================================             |  90%
  |
  |=====================================================            |  92%
  |
  |======================================================           |  94%
  |
  |=======================================================          |  96%
  |
  |========================================================         |  98%
  |
  |==========================================================| 100%[1] ------------------------
#> [1] Number of variables included: 7
#> [1] Dropped:
#> [1] x8
#>
  |
  |                                                                  |   0%
  |
  |=                                                                 |   2%
  |
  |===                                                               |   4%
  |
  |====                                                              |   6%
  |
  |=====                                                             |   8%
  |
  |======                                                            |  10%
  |
  |========                                                          |  12%
  |
  |=========                                                         |  14%
```

```
|
|==========                                                    |  16%
|
|============                                                  |  18%
|
|==============                                                |  20%
|
|==============                                                |  22%
|
|================                                              |  24%
|
|==================                                            |  26%
|
|====================                                          |  28%
|
|======================                                        |  30%
|
|========================                                      |  32%
|
|========================                                      |  34%
|
|==========================                                    |  36%
|
|============================                                  |  38%
|
|============================                                  |  40%
|
|==============================                                |  42%
|
|================================                              |  44%
|
|================================                              |  46%
|
|==================================                            |  48%
|
|==================================                            |  50%
|
|====================================                          |  52%
|
|======================================                        |  54%
|
|======================================                        |  56%
|
|========================================                      |  58%
|
|========================================                      |  60%
|
|==========================================                    |  62%
|
|==========================================                    |  64%
|
|============================================                  |  66%
|
```

```
  |=========================================                        |  68%
  |
  |==========================================                       |  70%
  |
  |===========================================                      |  72%
  |
  |============================================                     |  74%
  |
  |=============================================                    |  76%
  |
  |==============================================                   |  78%
  |
  |===============================================                  |  80%
  |
  |================================================                 |  82%
  |
  |=================================================                |  84%
  |
  |==================================================               |  86%
  |
  |===================================================              |  88%
  |
  |====================================================             |  90%
  |
  |=====================================================            |  92%
  |
  |======================================================           |  94%
  |
  |=======================================================          |  96%
  |
  |========================================================         |  98%
  |
  |==========================================================| 100%[1] ------------------------
#> [1] Number of variables included: 6
#> [1] Dropped:
#> [1] x8 x6
#>
  |
  |                                                                  |   0%
  |
  |=                                                                 |   2%
  |
  |===                                                               |   4%
  |
  |====                                                              |   6%
  |
  |=====                                                             |   8%
  |
  |======                                                            |  10%
  |
  |========                                                          |  12%
  |
  |=========                                                         |  14%
```

```
|
|==========                                                    |  16%
|
|============                                                  |  18%
|
|==============                                                |  20%
|
|==============                                                |  22%
|
|================                                              |  24%
|
|================                                              |  26%
|
|==================                                            |  28%
|
|====================                                          |  30%
|
|======================                                        |  32%
|
|======================                                        |  34%
|
|========================                                      |  36%
|
|========================                                      |  38%
|
|==========================                                    |  40%
|
|============================                                  |  42%
|
|============================                                  |  44%
|
|==============================                                |  46%
|
|==============================                                |  48%
|
|================================                              |  50%
|
|================================                              |  52%
|
|==================================                            |  54%
|
|==================================                            |  56%
|
|====================================                          |  58%
|
|====================================                          |  60%
|
|======================================                        |  62%
|
|======================================                        |  64%
|
|========================================                      |  66%
|
```

```
  |=========================================                        |  68%
  |
  |==========================================                       |  70%
  |
  |===========================================                      |  72%
  |
  |============================================                     |  74%
  |
  |=============================================                    |  76%
  |
  |==============================================                   |  78%
  |
  |===============================================                  |  80%
  |
  |================================================                 |  82%
  |
  |=================================================                |  84%
  |
  |==================================================               |  86%
  |
  |===================================================              |  88%
  |
  |====================================================             |  90%
  |
  |=====================================================            |  92%
  |
  |======================================================           |  94%
  |
  |=======================================================          |  96%
  |
  |========================================================         |  98%
  |
  |==========================================================================| 100%[1] ----------------------------
#> [1] Number of variables included: 5
#> [1] Dropped:
#> [1] x8 x6 x7
#>
  |
  |                                                                  |   0%
  |
  |=                                                                 |   2%
  |
  |===                                                               |   4%
  |
  |====                                                              |   6%
  |
  |=====                                                             |   8%
  |
  |======                                                            |  10%
  |
  |========                                                          |  12%
  |
  |=========                                                         |  14%
```

```
|
|==========                                                    |  16%
|
|============                                                  |  18%
|
|==============                                                |  20%
|
|==============                                                |  22%
|
|================                                              |  24%
|
|==================                                            |  26%
|
|====================                                          |  28%
|
|======================                                        |  30%
|
|========================                                      |  32%
|
|========================                                      |  34%
|
|==========================                                    |  36%
|
|============================                                  |  38%
|
|============================                                  |  40%
|
|==============================                                |  42%
|
|================================                              |  44%
|
|================================                              |  46%
|
|==================================                            |  48%
|
|==================================                            |  50%
|
|====================================                          |  52%
|
|======================================                        |  54%
|
|======================================                        |  56%
|
|========================================                      |  58%
|
|==========================================                    |  60%
|
|==========================================                    |  62%
|
|============================================                  |  64%
|
|============================================                  |  66%
|
```

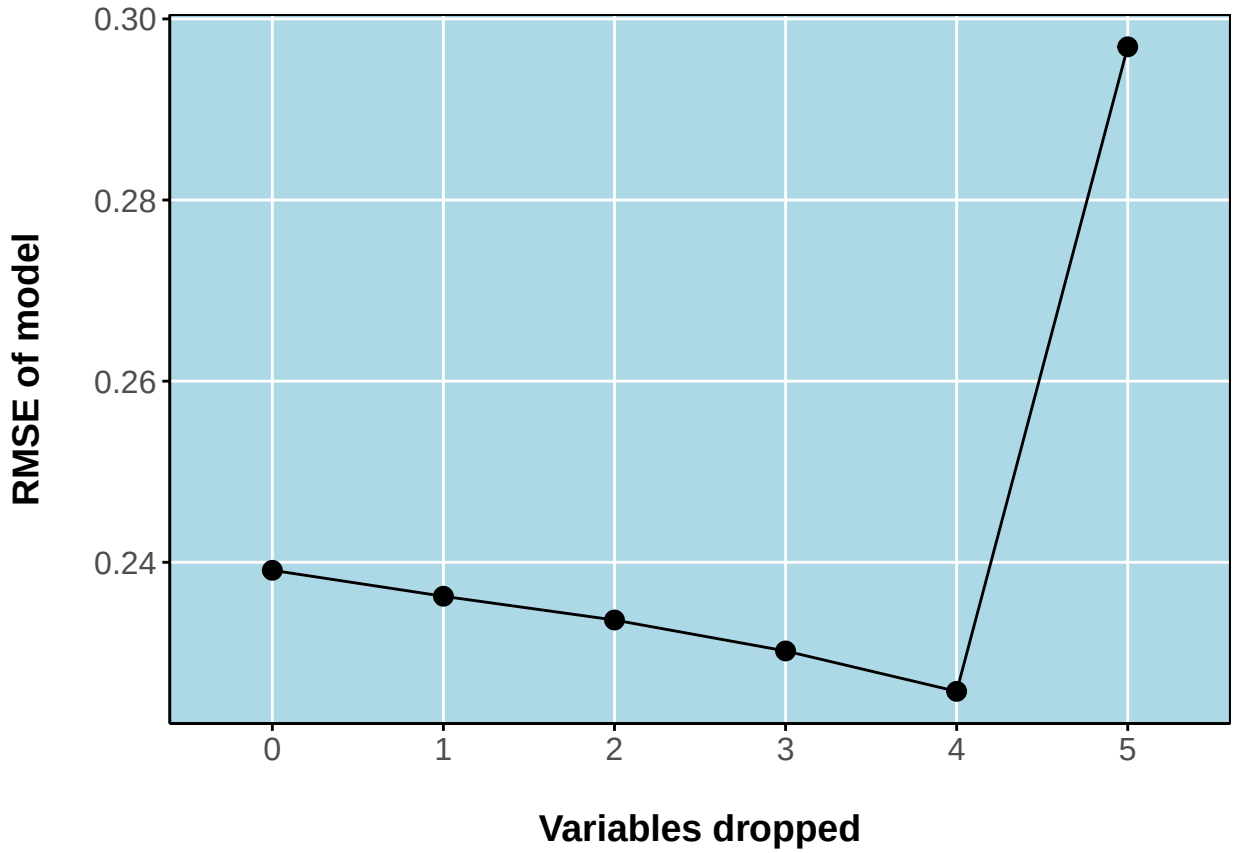
```
  |=========================================                        |  68%
  |
  |==========================================                       |  70%
  |
  |===========================================                      |  72%
  |
  |============================================                     |  74%
  |
  |=============================================                    |  76%
  |
  |==============================================                   |  78%
  |
  |===============================================                  |  80%
  |
  |================================================                 |  82%
  |
  |=================================================                |  84%
  |
  |==================================================               |  86%
  |
  |===================================================              |  88%
  |
  |====================================================             |  90%
  |
  |=====================================================            |  92%
  |
  |======================================================           |  94%
  |
  |=======================================================          |  96%
  |
  |========================================================         |  98%
  |
  |==========================================================================| 100%[1] ------------------------
#> [1] Number of variables included: 4
#> [1] Dropped:
#> [1] x8 x6 x7 x5
#>
  |
  |                                                                  |   0%
  |
  |=                                                                 |   2%
  |
  |===                                                               |   4%
  |
  |====                                                              |   6%
  |
  |=====                                                             |   8%
  |
  |======                                                            |  10%
  |
  |========                                                          |  12%
  |
  |=========                                                         |  14%
```

```
|
|==========                                                     |  16%
|
|============                                                   |  18%
|
|==============                                                 |  20%
|
|==============                                                 |  22%
|
|================                                               |  24%
|
|==================                                             |  26%
|
|====================                                           |  28%
|
|======================                                         |  30%
|
|========================                                       |  32%
|
|========================                                       |  34%
|
|==========================                                     |  36%
|
|============================                                   |  38%
|
|============================                                   |  40%
|
|==============================                                 |  42%
|
|================================                               |  44%
|
|================================                               |  46%
|
|==================================                             |  48%
|
|==================================                             |  50%
|
|====================================                           |  52%
|
|======================================                         |  54%
|
|======================================                         |  56%
|
|========================================                       |  58%
|
|========================================                       |  60%
|
|==========================================                     |  62%
|
|==========================================                     |  64%
|
|============================================                   |  66%
|
```

```
  |=========================================                        |  68%
  |
  |==========================================                       |  70%
  |
  |===========================================                      |  72%
  |
  |============================================                     |  74%
  |
  |=============================================                    |  76%
  |
  |==============================================                   |  78%
  |
  |===============================================                  |  80%
  |
  |================================================                 |  82%
  |
  |=================================================                |  84%
  |
  |==================================================               |  86%
  |
  |===================================================              |  88%
  |
  |====================================================             |  90%
  |
  |=====================================================            |  92%
  |
  |======================================================           |  94%
  |
  |=======================================================          |  96%
  |
  |========================================================         |  98%
  |
  |==========================================================| 100%[1] ------------------------
#> [1] Number of variables included: 3
#> [1] Dropped:
#> [1] x8 x6 x7 x5 x4
#>
  |
  |                                                                  |   0%
  |
  |=                                                                 |   2%
  |
  |===                                                               |   4%
  |
  |====                                                              |   6%
  |
  |=====                                                             |   8%
  |
  |======                                                            |  10%
  |
  |========                                                          |  12%
  |
  |=========                                                         |  14%
```

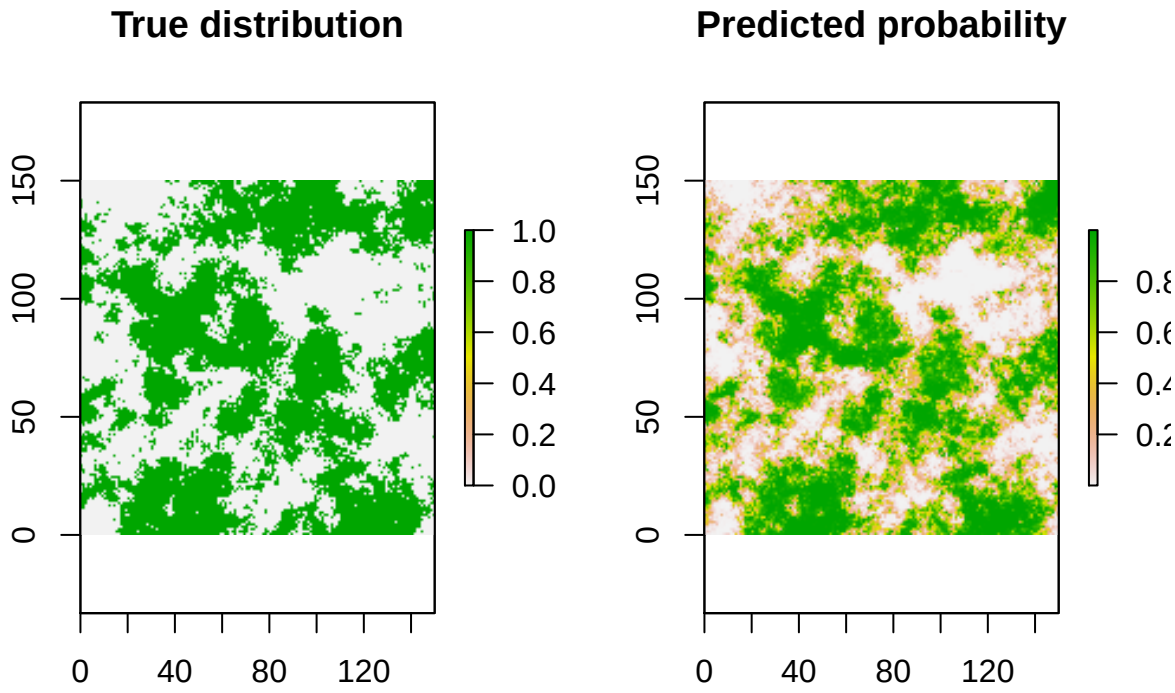
```
|
|==========                                                    |  16%
|
|============                                                  |  18%
|
|==============                                                |  20%
|
|==============                                                |  22%
|
|================                                              |  24%
|
|==================                                            |  26%
|
|====================                                          |  28%
|
|======================                                        |  30%
|
|========================                                      |  32%
|
|========================                                      |  34%
|
|==========================                                    |  36%
|
|============================                                  |  38%
|
|============================                                  |  40%
|
|==============================                                |  42%
|
|================================                              |  44%
|
|================================                              |  46%
|
|==================================                            |  48%
|
|==================================                            |  50%
|
|====================================                          |  52%
|
|====================================                          |  54%
|
|======================================                        |  56%
|
|========================================                      |  58%
|
|========================================                      |  60%
|
|==========================================                    |  62%
|
|==========================================                    |  64%
|
|============================================                  |  66%
|
```

```
|=============================================                    |  68%
|
|================================================                  |  70%
|
|================================================                  |  72%
|
|================================================                  |  74%
|
|================================================                  |  76%
|
|=================================================                 |  78%
|
|=================================================                 |  80%
|
|=================================================                 |  82%
|
|=================================================                 |  84%
|
|=================================================                 |  86%
|
|=================================================                 |  88%
|
|=================================================                 |  90%
|
|=================================================                 |  92%
|
|=================================================                 |  94%
|
|=================================================                 |  96%
|
|================================================================= |  98%
|
|=================================================================| 100%[1] ------------------------
```



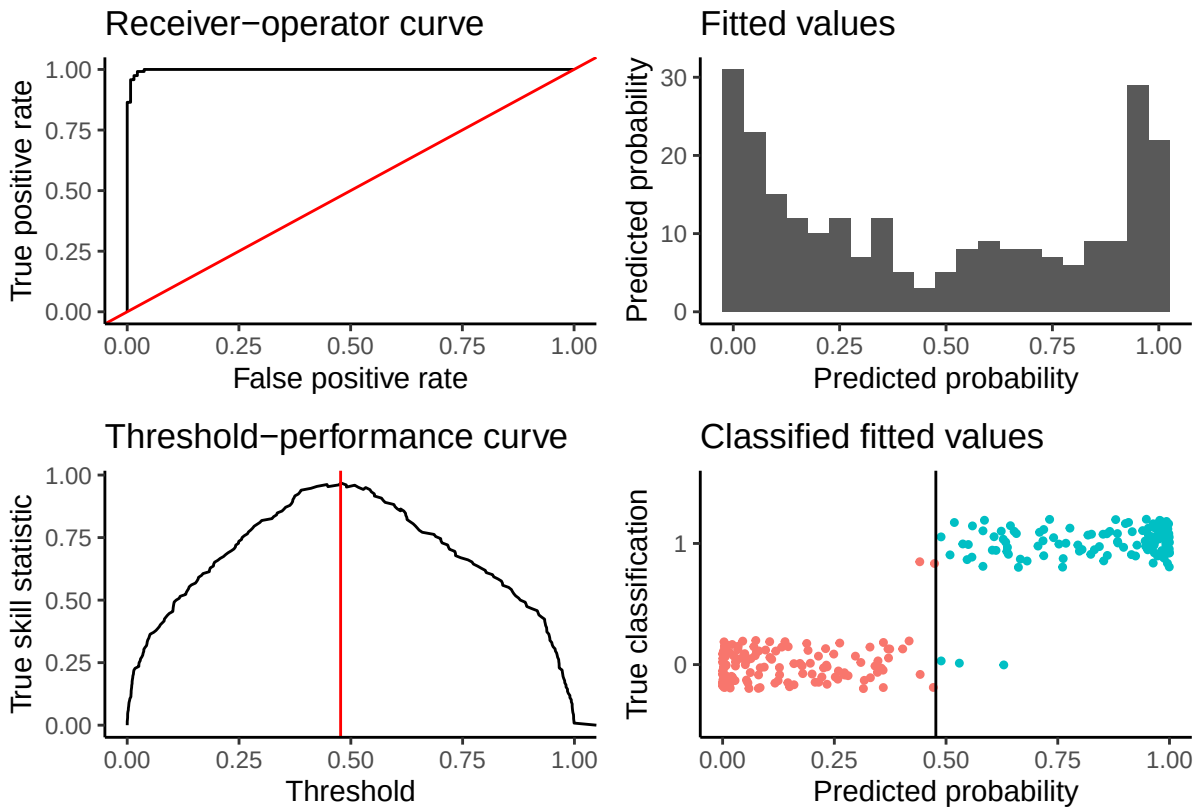

```
#> [1] ----------------------------------------
#> [1] Final recommended variable list
#> [1] x1 x2 x3 x4
step.model
#> [1] "x1" "x2" "x3" "x4"

# Retrain the model
sdm <- bart(x.train=occ.df[, step.model], y.train=occ.df[,'Observed'],
            keeptrees = TRUE)
#>
#> Running BART with binary y
#>
#> number of trees: 200
#> number of chains: 1, number of threads 1
#> Prior:
#>  k: 2.000000
#>  power and base for tree prior: 2.000000 0.950000
#>  use quantiles for rule cut points: false
#> data:
#>  number of training observations: 250
#>  number of test observations: 0
#>  number of explanatory variables: 4
#>
#> Cutoff rules c in x<=c vs x>c
#> Number of cutoffs: (var: number of possible c):
#> (1: 100) (2: 100) (3: 100) (4: 100)
```

```
#>
#> offsets:
#>  reg : 0.00 0.00 0.00 0.00 0.00
#> Running mcmc loop:
#> iteration: 100 (of 1000)
#> iteration: 200 (of 1000)
#> iteration: 300 (of 1000)
#> iteration: 400 (of 1000)
#> iteration: 500 (of 1000)
#> iteration: 600 (of 1000)
#> iteration: 700 (of 1000)
#> iteration: 800 (of 1000)
#> iteration: 900 (of 1000)
#> iteration: 1000 (of 1000)
#> total seconds in loop: 2.641285
#>
#> Tree sizes, last iteration:
#> [1] 2 3 2 3 2 4 3 2 2 2 2 2 2 3 3 2 3 2
#> 3 4 2 3 2 2 2 3 1 2 3 2 2 2 3 2 2 2 2 2
#> 3 2 2 2 4 2 2 2 2 2 2 2 2 2 3 2 2 3 3 2
#> 2 2 2 3 2 2 6 3 3 2 3 3 2 2 2 2 2 2 2 2
#> 2 3 2 1 2 3 1 2 2 4 3 2 2 4 3 2 3 4 2 3
#> 2 2 2 2 3 2 3 1 4 2 3 2 2 3 1 3 3 3 2 2
#> 3 2 4 2 2 4 3 3 3 2 2 2 4 2 2 4 4 3 3 2
#> 2 2 3 3 3 2 1 2 2 3 3 2 3 3 3 2 3 2 2 2
#> 2 4 2 2 2 3 4 2 2 2 2 3 2 3 5 2 2 3 2 2
#> 2 3 3 2 3 2 2 2 2 3 3 3 2 4 3 3 5 3 2 3
#> 2 2
#>
#> Variable Usage, last iteration (var:count):
#> (1: 64) (2: 88) (3: 78) (4: 66)
#> DONE BART

# Predict the species distribution!
map <- predict(sdm, climate)

# How's it look?
par(mfrow=c(1,2))
plot(random.sp$pa.raster, main='True distribution')
plot(map, main='Predicted probability')
```



OK. We have a good-looking model. What's inside?

```
# How good is it?
summary(sdm)
#> Call:  bart occ.df[, step.model] occ.df[, "Observed"] TRUE
#>
#> Predictor list:
#>  x1 x2 x3 x4
#>
#> Area under the receiver-operator curve
#> AUC = 0.9983308
#>
#> Recommended threshold (maximizes true skill statistic)
#> Cutoff =  0.4771741
#> TSS =  0.9687982
#> Resulting type I error rate:  0.008474576
#> Resulting type II error rate:  0.02272727
```

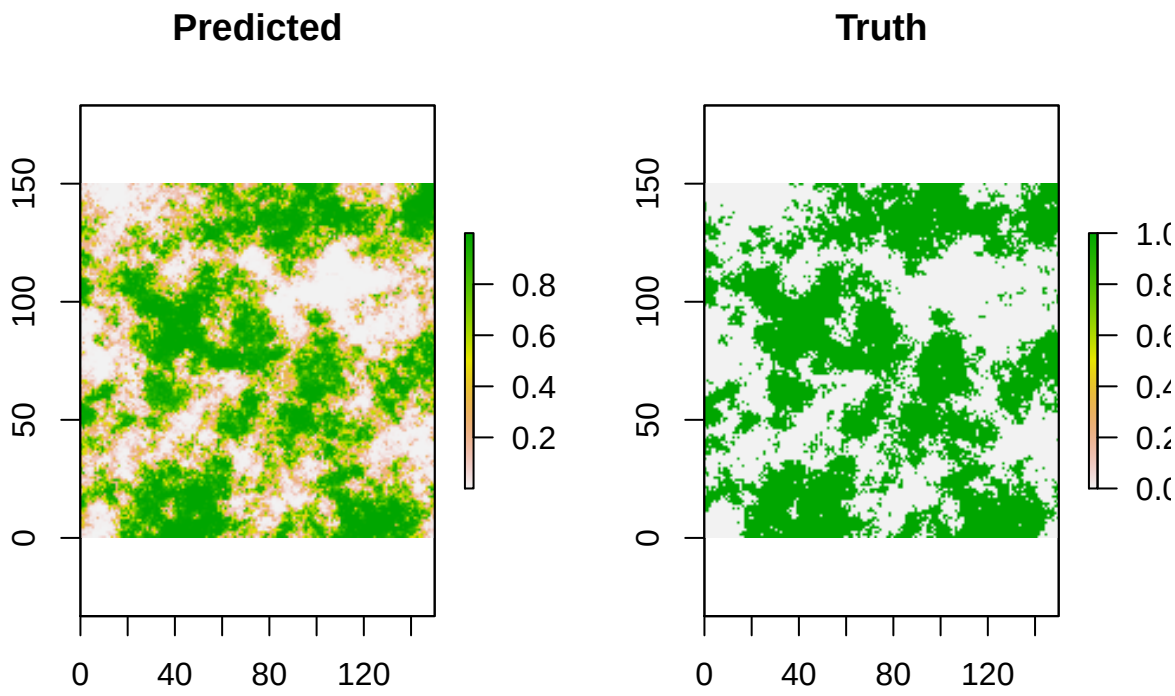



```
map <- predict(sdm, climate, splitby=50, quantiles=c(0.01))
#> Estimated time to total prediction (mins):
#> 1.139946
#>
  |
  |                                                                      |   0%
  |
  |=                                                                     |   2%
  |
  |===                                                                   |   4%
  |
  |====                                                                  |   6%
  |
  |=====                                                                 |   8%
  |
  |======                                                                |  10%
  |
  |========                                                              |  12%
  |
  |=========                                                             |  14%
  |
  |==========                                                            |  16%
  |
  |============                                                          |  18%
  |
  |=============                                                         |  20%
```

```
|
|==============                                                  |  22%
|
|================                                                |  24%
|
|==================                                              |  26%
|
|====================                                            |  28%
|
|======================                                          |  30%
|
|========================                                        |  32%
|
|==========================                                      |  34%
|
|============================                                    |  36%
|
|==============================                                  |  38%
|
|================================                                |  40%
|
|==================================                              |  42%
|
|====================================                            |  44%
|
|======================================                          |  46%
|
|========================================                        |  48%
|
|==========================================                      |  50%
|
|============================================                    |  52%
|
|==============================================                  |  54%
|
|================================================                |  56%
|
|==================================================              |  58%
|
|====================================================            |  60%
|
|======================================================          |  62%
|
|========================================================        |  64%
|
|==========================================================      |  66%
|
|============================================================    |  68%
|
|==============================================================  |  70%
|
|================================================================|  72%
|
```

```
  |=============================================                    |  74%
  |
  |==============================================                   |  76%
  |
  |===============================================                  |  78%
  |
  |================================================                 |  80%
  |
  |=================================================                |  82%
  |
  |==================================================               |  84%
  |
  |===================================================              |  86%
  |
  |====================================================             |  88%
  |
  |=====================================================            |  90%
  |
  |======================================================           |  92%
  |
  |=======================================================          |  94%
  |
  |========================================================         |  96%
  |
  |========================================================================= |  98%
  |
  |==========================================================================| 100%
```

```
# Check
par(mfrow=c(1,2))
plot(map[[1]], main='Predicted')
plot(random.sp$pa.raster, main='Truth')
```




```
#plot(values(map) ~ values(random.sp$pa.raster))
```