MENA Exceptionalism and BART

Daniel K Baissa

2023-05-10

In this example I will use BART to predict coups (not the DV) globally and use it to explain my logic for looking at MENA exceptionalism

I will start by loading the libraries and setting up bartMachine

```
library(tidyverse)
library(rJava)

options(java.parameters = "-Xmx50g")
library(bartMachine)
set_bart_machine_num_cores(30)
```

```
## bartMachine now using 30 cores.
```

This is just a subset of data for illustrative perpouses and speed on my laptop.

I will set up the data for BARTmachine

https://statisticsglobe.com/convert-factor-to-dummy-variables-in-r

```
df <- read_csv("Data/Vdem_Banks.csv") |>
  select(!c(suvival_time, coup_number, id, ...1)) |>
  mutate(Region = as.factor(Region)) |>
  filter(country_name != "Israel") # Removing Israel because its an outlier in the MENA





df <- data.frame(df[ , ! colnames(df) %in% "Region"],       # Create dummy data
                        model.matrix( ~ (Region) - 1, df) )



X <- df |>
  na.omit()

y <- as.factor(X$e_pt_coup)

X$e_pt_coup <- NULL

y <- relevel(y, "1")

X <- as.data.frame(X)
```

Here are the variables on the right hand side of this model:

```r
# Convert column names to a table for ease of reading
col_table <- matrix(colnames(X), ncol = 2, byrow = TRUE)
```

```
## Warning in matrix(colnames(X), ncol = 2, byrow = TRUE): data length [37] is not
## a sub-multiple or multiple of the number of rows [19]
```

```r
# Print the table
col_table
```

```
##       [,1]                                 [,2]
##  [1,] "country_name"                       "year"
##  [2,] "Polity.5"                           "GDP.Growth"
##  [3,] "GDP.Per.Capita"                     "Inflation"
##  [4,] "Petroleum.production.per.capita"    "Infant.Mortality"
##  [5,] "Clientelism"                        "Rule.of.Law"
##  [6,] "Party.Institutionalization"         "Legislative.Party.Coh"
##  [7,] "National.party.control"             "Political.Polarization"
##  [8,] "Defense.Exp.Per.Cap"                "Boix.Democracy"
##  [9,] "Barriers.to.parties"                "Party.Ban"
## [10,] "Opposition.parties.autonomy"        "Party.organizations"
## [11,] "Party.branches"                     "Party.linkages"
## [12,] "Distinct.party.platforms"           "Candidate.selection"
## [13,] "Party.competition.across.regions"   "Subnational.party.control"
## [14,] "v2exdfpphg"                         "Region1"
## [15,] "Region2"                            "Region3"
## [16,] "Region4"                            "Region5"
## [17,] "Region6"                            "Region7"
## [18,] "Region8"                            "Region9"
## [19,] "Region10"                           "country_name"
```

Now let's run a basic BART model. The data are extremely sparce so I will set prob_rule_class = .1 to compensate. This can be validated by looking at the confusion matrix to make sure the model has a good fit.

```r
bm <- bartMachine(y = y, X = X, prob_rule_class = .1)
```

```
## bartMachine initializing with 50 trees...
## bartMachine vars checked...
## bartMachine java init...
## bartMachine factors created...
## bartMachine before preprocess...
## bartMachine after preprocess... 134 total features...
## bartMachine training data finalized...
## Now building bartMachine for classification where "1" is considered the target level...Covariate impo
## evaluating in sample data...done
```

```r
bm
```

```
## bartMachine v1.3.2 for classification
```

```
##
## training data size: n = 1895 and p = 134
## built in 12.8 secs on 30 cores, 50 trees, 250 burn-in and 1000 post. samples
##
## confusion matrix:
##
##              predicted 1 predicted 0 model errors
## actual 1          30.000      14.000        0.318
## actual 0          67.000    1784.000        0.036
## use errors         0.691       0.008        0.043
```

Here we see that the model accurately classifies 96.4% of non-coups and 68.2% of coups with an overall accuracy rate of 95.7%

Variable Selection

Note that there are 134 variables in the final model. Let's select a smaller number of variables to allow for an easier comparison to see how Region might differ.

```r
vs <-  var_selection_by_permute(bm, plot  = FALSE)
```

```
## avg..........null...........................................................................
```

Now we will use local names to select the following Party.competition.across.regions, Boix.Democracy, Political.Polarization, Infant.Mortality as variables

```r
X2 <- df |>
  select(c(Region3, e_pt_coup, vs$important_vars_local_names)) |>
  na.omit()

y2 <- X2$e_pt_coup

X2$e_pt_coup <- NULL


bm2 <- bartMachine(y = y2, X = X2, prob_rule_class = .1)
```

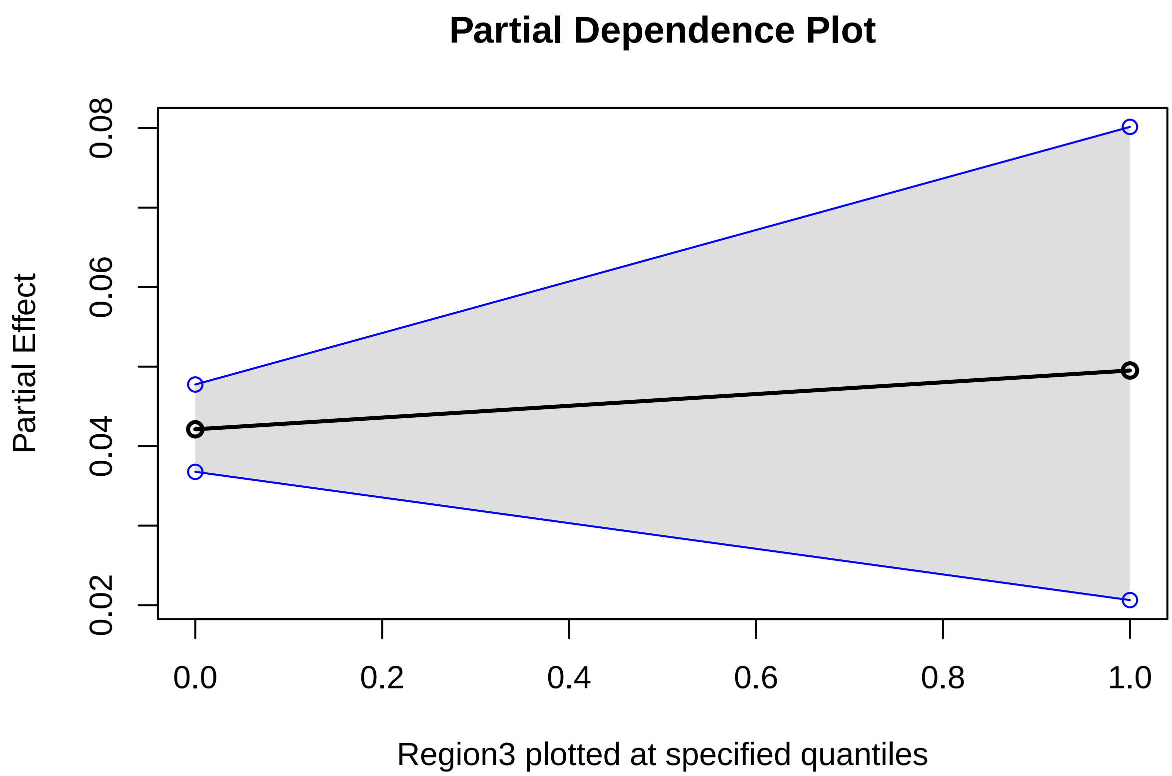
```
## bartMachine initializing with 50 trees...
## bartMachine vars checked...
## bartMachine java init...
## bartMachine factors created...
## bartMachine before preprocess...
## bartMachine after preprocess... 5 total features...
## bartMachine sigsq estimated...
## bartMachine training data finalized...
## Now building bartMachine for regression...
## evaluating in sample data...done
```

PD Plot

Region 3 is for the MENA

```
pd_plot(bm2,  "Region3")
```

```
## ..
```

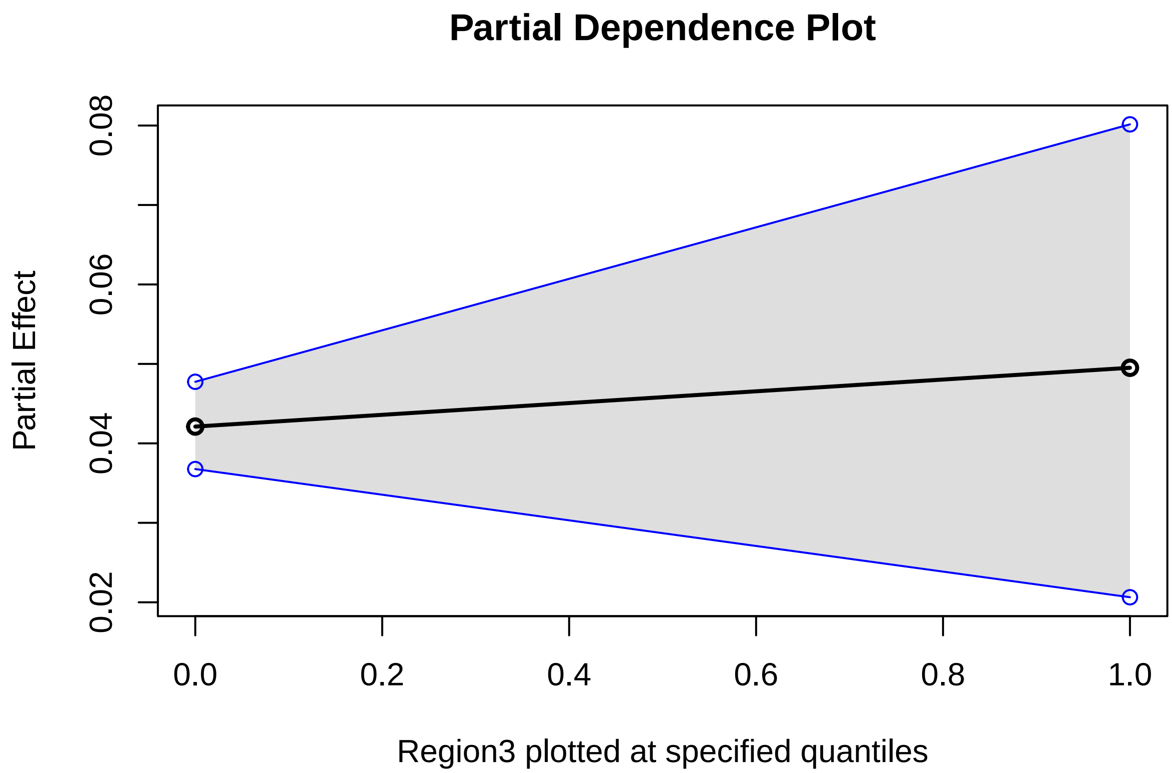


Here we find that the MENA region is about the same with respect to coup chance as any other region.

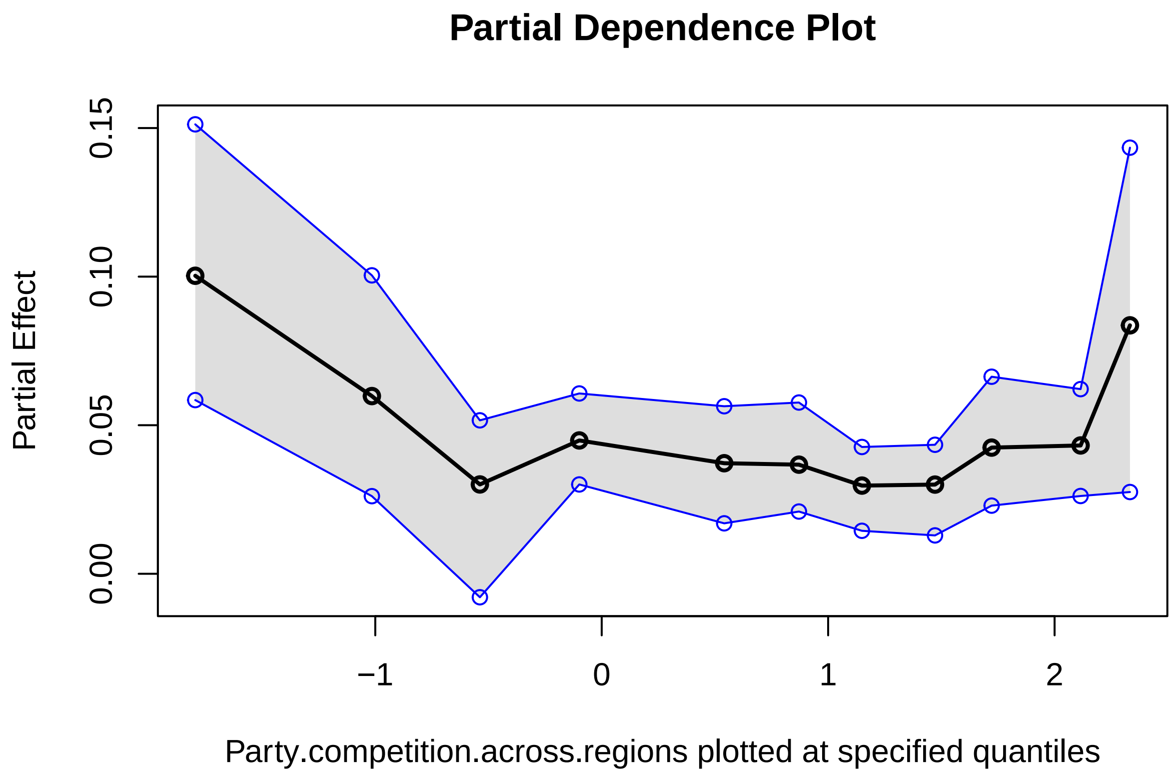
```
for (i in 1:length(bm2$training_data_features)) {
  pd_plot(bm2,  bm2$training_data_features[i])

}
```

```
## ..
```

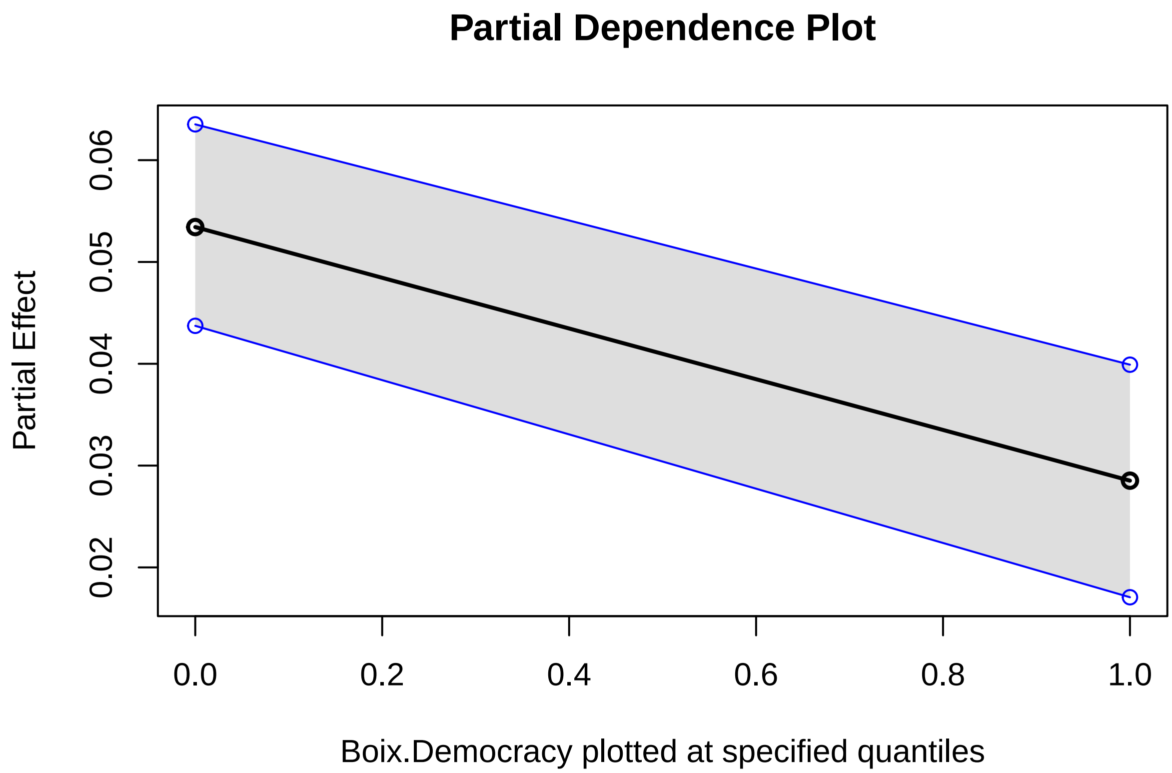

Partial Dependence Plot
Partial Effect
0.02
0.04
0.06
0.08
0.0
0.2
0.4
0.6
0.8
1.0
Region3 plotted at specified quantiles

```
## ...........
```



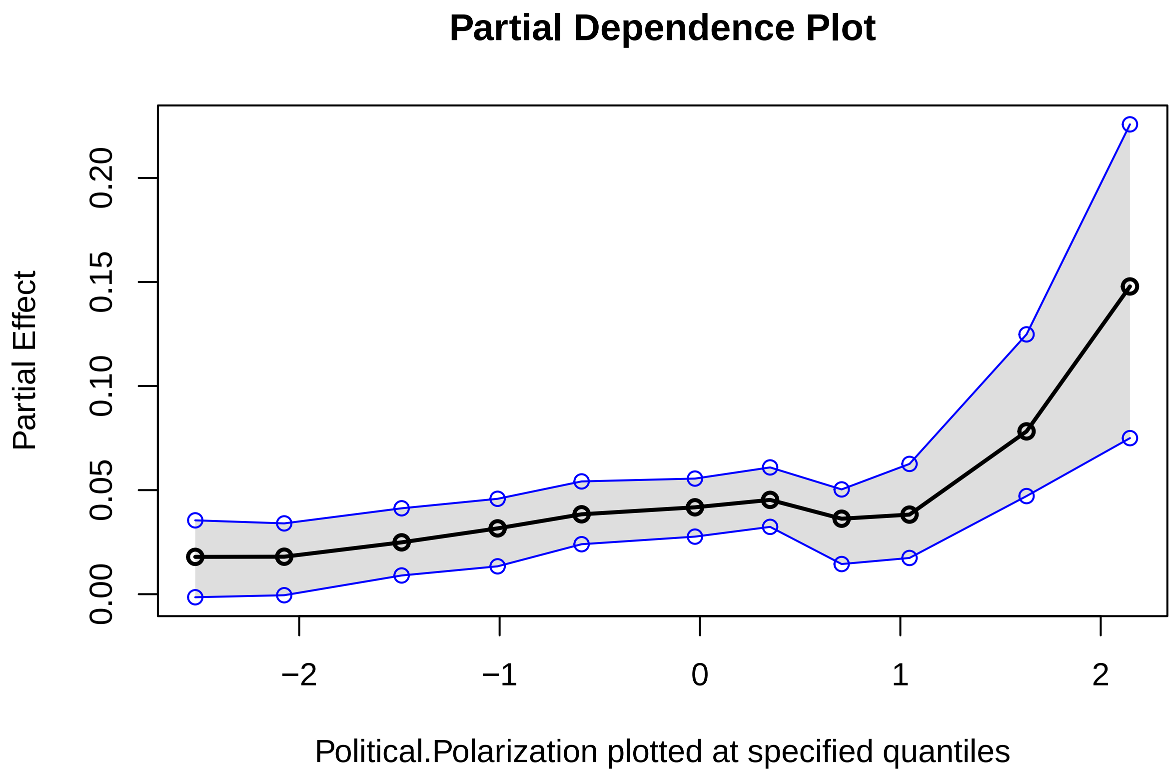
Partial Dependence Plot
Partial Effect
0.00
0.05
0.10
0.15
−1
0
1
2
Party.competition.across.regions plotted at specified quantiles


```
## ..
```

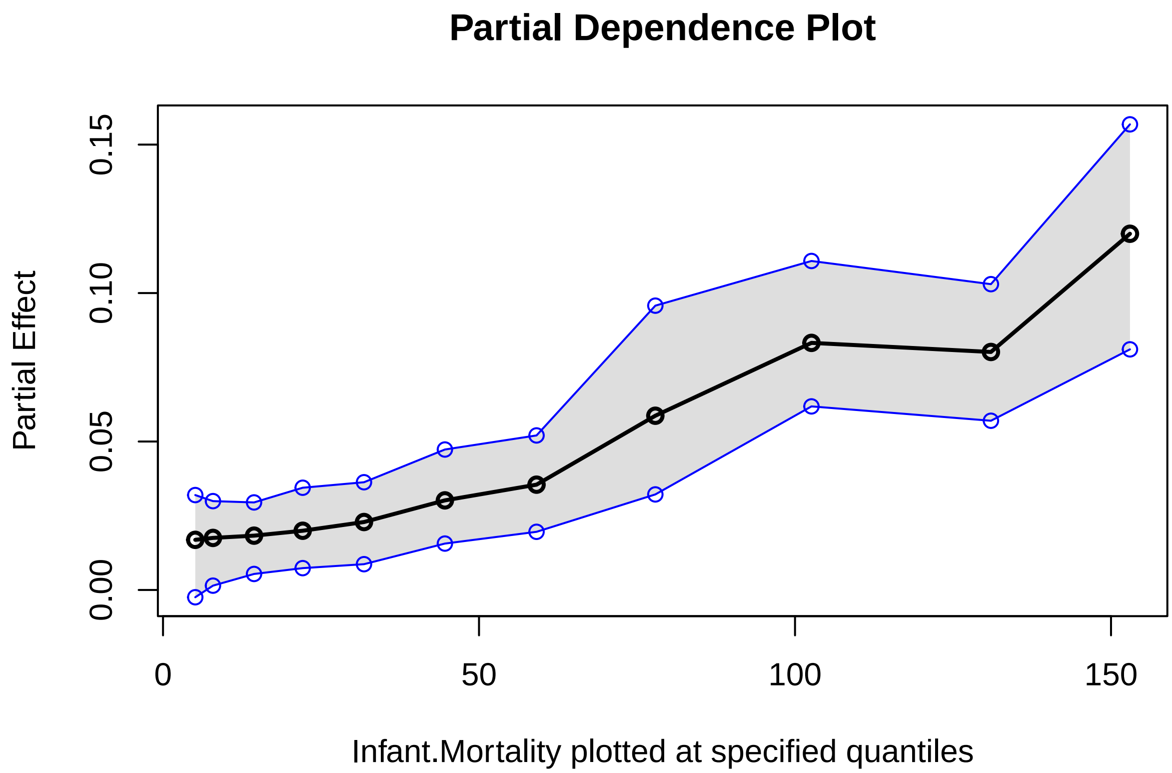


Partial Dependence Plot
Partial Effect
0.02
0.03
0.04
0.05
0.06
0.0
0.2
0.4
0.6
0.8
1.0
Boix.Democracy plotted at specified quantiles


```
## ...........
```


Partial Dependence Plot
Partial Effect
0.00
0.05
0.10
0.15
0.20
−2
−1
0
1
2
Political.Polarization plotted at specified quantiles

```
## ...........
```



Here we can see that Party Competition has a U shaped Function. Democracy is negativly associated with coups, Infant mortality is poitivly assoicated with coups, and political polarization is moderatly associated with coups at the extremel levels.

Within the MENA

Now let's look at the results within the MENA... This unfortunately would be like what we did before, because we want to know about variation within the region. Hariri used a 2 stage model to say that MENA is unique. I am not aware of such a BART model today. But I am not sure what he did made sense anyways.

Countries in the MENA dataset

```
df |>
  filter(Region3 == 1) |>
  select(country_name) |>
  unique()
```

```
##               country_name
## 135                Algeria
## 479                Bahrain
## 2250                 Egypt
## 3551                  Iran
## 3619                  Iraq
```

```
## 4006               Jordan
## 4158               Kuwait
## 4333              Lebanon
## 4521                Libya
## 5279              Morocco
## 5979                 Oman
## 6555                Qatar
## 6785         Saudi Arabia
## 7551                Syria
## 7994              Tunisia
## 8055               Turkey
## 8229 United Arab Emirates
## 8647                Yemen
```

Stage 1

Stage 1 would be something like starting by predicting the MENA region. Technically that would be finding everything that is related to being in the MENA region.

```
X3 <- df |>
  select(-c("Region1",
            "Region2",
            "Region4",
            "Region5",
            "Region6",
            "Region7",
            "Region8",
            "Region9",
            "Region10",
            "country_name")) |>
  na.omit()

y3 <- as.factor(X3$Region3)
y3 <- relevel(y3, "1")

X3$Region3 <- NULL

bm3 <- bartMachine(y = y3, X = X3, prob_rule_class = .5)
```

```
## bartMachine initializing with 50 trees...
## bartMachine vars checked...
## bartMachine java init...
## bartMachine factors created...
## bartMachine before preprocess...
## bartMachine after preprocess... 27 total features...
## bartMachine training data finalized...
## Now building bartMachine for classification where "1" is considered the target level...
## evaluating in sample data...done
```

```
bm3
```

```
## bartMachine v1.3.2 for classification
```

```
##
## training data size: n = 1895 and p = 27
## built in 9.4 secs on 30 cores, 50 trees, 250 burn-in and 1000 post. samples
##
## confusion matrix:
##
##            predicted 1 predicted 0 model errors
## actual 1       290.000       1.000        0.003
## actual 0         1.000    1603.000        0.001
## use errors       0.003       0.001        0.001
```

This model was extremely effetive at predicting MENA countries.

Predictors of the MENA

Now what variables predict being in the MENA?

```
vs <-  var_selection_by_permute(bm3, plot  = FALSE)
```

```
## avg..........null.......................................................................................
```

The following variables predict being a MENA countryv2exdfpphg, Barriers.to.parties, Party.competition.across.regions, GDP.Per.Capita, Legislative.Party.Coh

Effect of Predictors of MENA

Now let's see how those predictors determine impact the MENA region

```
X4 <- df |>
  select(c(Region3, vs$important_vars_local_names)) |>
  na.omit()

y4 <- as.factor(X4$Region3)
y4 <- relevel(y4, "1")

X4$Region3 <- NULL

bm4 <- bartMachine(y = y4, X = X4, prob_rule_class = .5)
```

```
## bartMachine initializing with 50 trees...
## bartMachine vars checked...
## bartMachine java init...
## bartMachine factors created...
## bartMachine before preprocess...
## bartMachine after preprocess... 5 total features...
## bartMachine training data finalized...
## Now building bartMachine for classification where "1" is considered the target level...
## evaluating in sample data...done
```

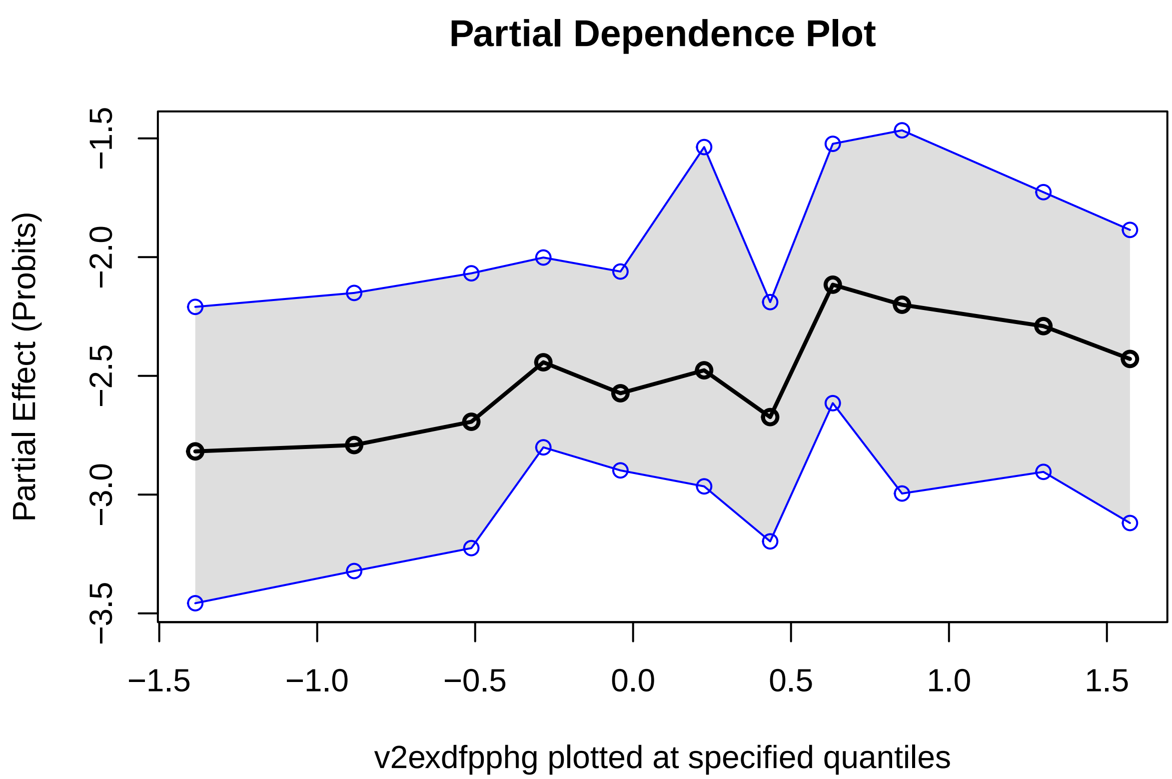
```
bm4
```

```
## bartMachine v1.3.2 for classification
##
## training data size: n = 4837 and p = 5
## built in 18.1 secs on 30 cores, 50 trees, 250 burn-in and 1000 post. samples
##
## confusion matrix:
##
##              predicted 1 predicted 0 model errors
## actual 1         750.000      44.000        0.055
## actual 0           8.000    4035.000        0.002
## use errors         0.011       0.011        0.011
```

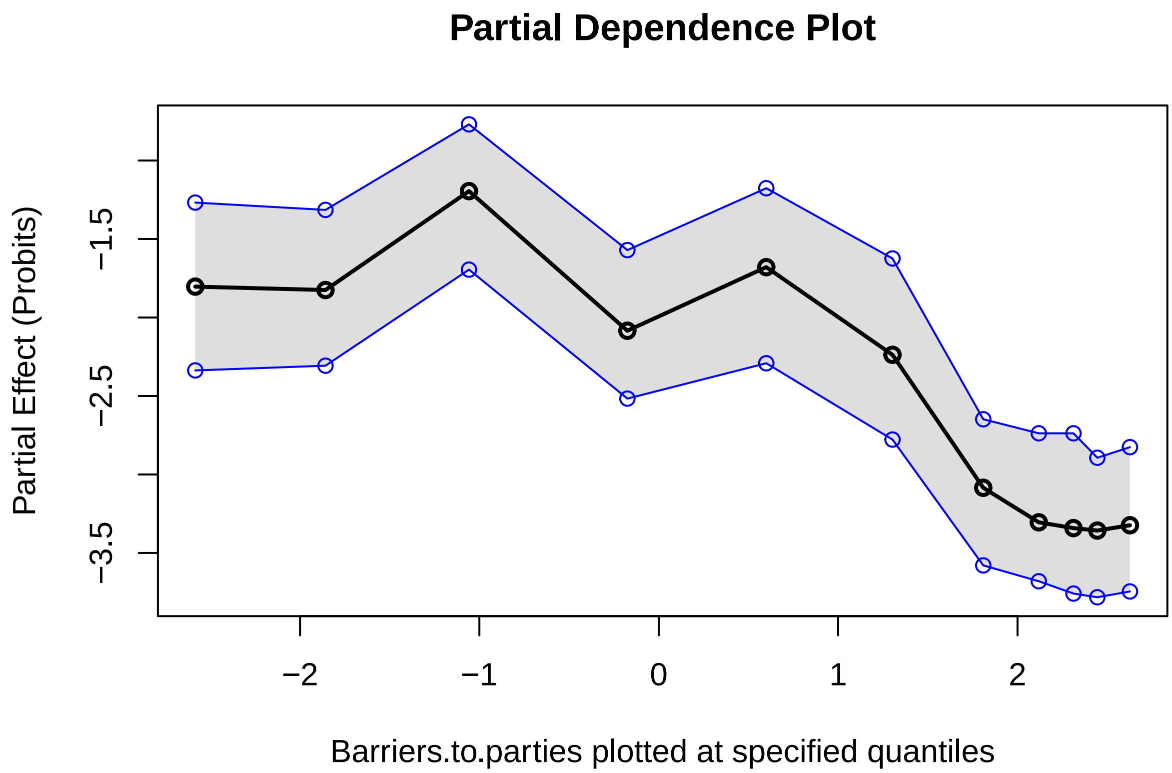
```
for (i in 1:length(bm4$training_data_features)) {
  pd_plot(bm4,  bm4$training_data_features[i])

}
```

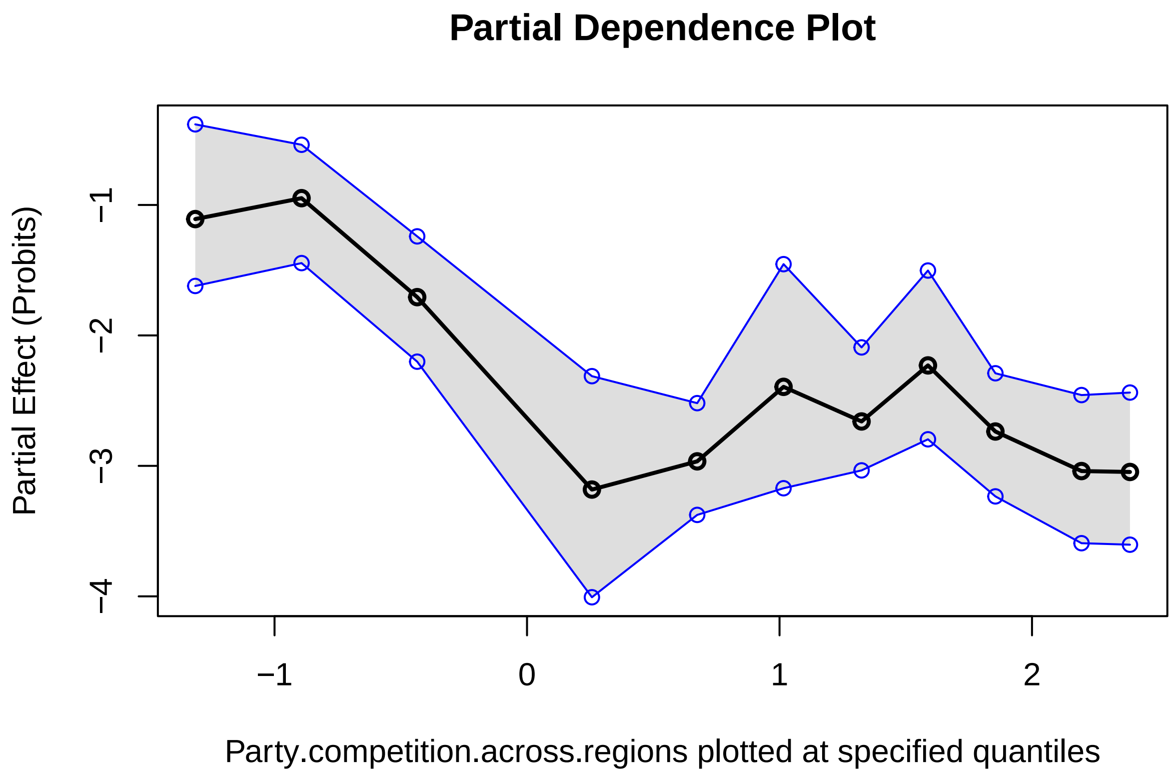
```
## ...........
```



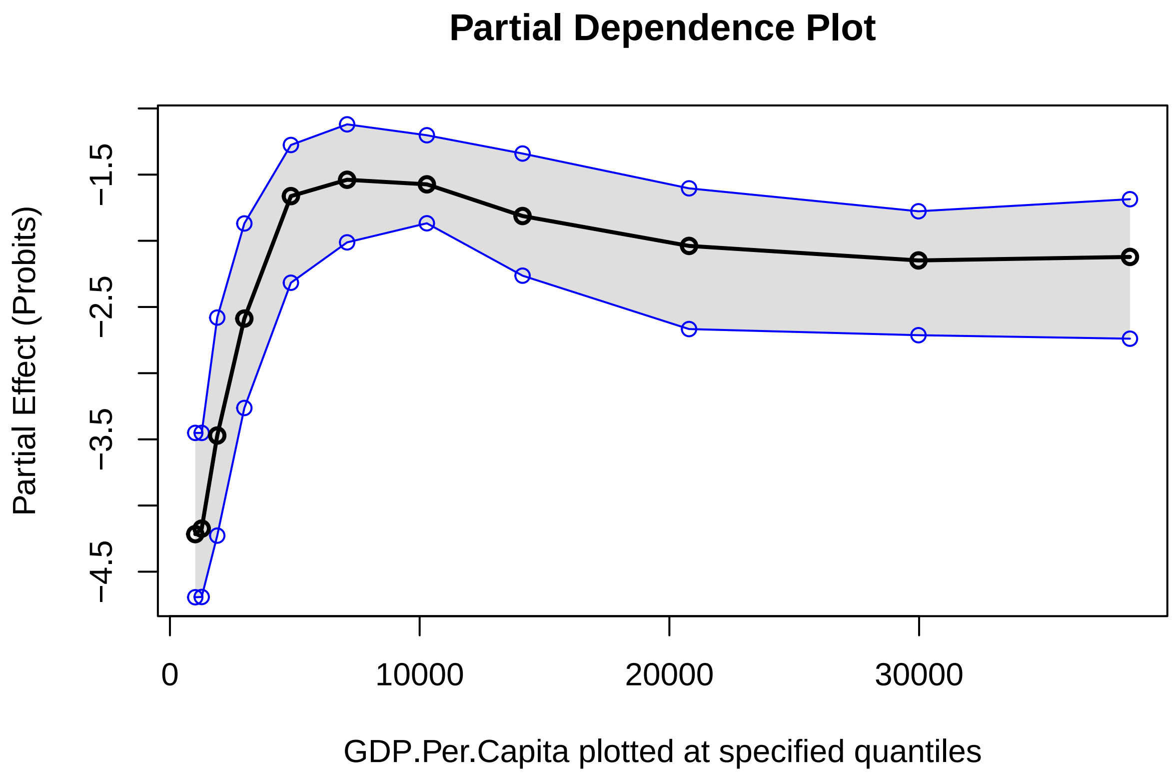

```
## ...........
```


Partial Dependence Plot
Partial Effect (Probits)
−1.5
−2.5
−3.5
−2
−1
0
1
2
Barriers.to.parties plotted at specified quantiles

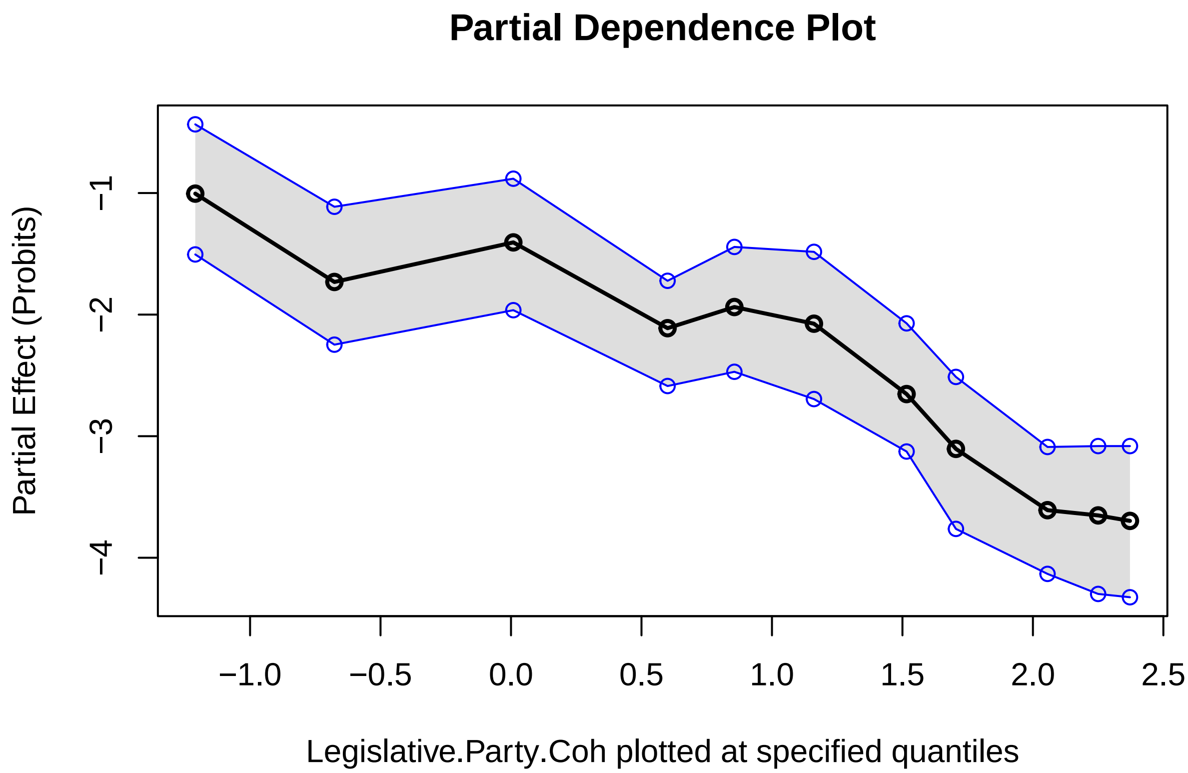
```
## ...........
```


Partial Dependence Plot
Partial Effect (Probits)
−1
−2
−3
−4
−1
0
1
2
Party.competition.across.regions plotted at specified quantiles

```
## ...........
```


Partial Dependence Plot
Partial Effect (Probits)
−1.5
−2.5
−3.5
−4.5
0
10000
20000
30000
GDP.Per.Capita plotted at specified quantiles

```
## ...........
```



This findings indicate that the MENA countries have lower party competition across their regions in their countries, Higher GDP Per Capitas, though not the highest. Produce more oil per capita, and are more likely to have their HOG propose legislation in practice (v2exdfpphg) though that may not be signficant. Oddly, it looks like V-dem also codes the MENA countries as having fewer barriers to political parties?