# XAI FINAL

Tufan BOSTAN

25127215588

```
#libraries
library(dplyr)
library(xgboost)
library(DALEX)
library(lightgbm)
library(party)
library(e1071)
library(gbm)
library(randomForest)
```

## forestfires

```
forestfires <- read.csv(paste0(getwd(),"/forestfires.csv"))
head(forestfires)
```

```
##   X Y month day FFMC  DMC    DC  ISI temp RH wind rain area
## 1 7 5   mar fri 86.2 26.2  94.3  5.1  8.2 51  6.7  0.0    0
## 2 7 4   oct tue 90.6 35.4 669.1  6.7 18.0 33  0.9  0.0    0
## 3 7 4   oct sat 90.6 43.7 686.9  6.7 14.6 33  1.3  0.0    0
## 4 8 6   mar fri 91.7 33.3  77.5  9.0  8.3 97  4.0  0.2    0
## 5 8 6   mar sun 89.3 51.3 102.2  9.6 11.4 99  1.8  0.0    0
## 6 8 6   aug sun 92.3 85.3 488.0 14.7 22.2 29  5.4  0.0    0
```

```
summary(forestfires)
```

```
##        X               Y          month               day
##  Min.   :1.000   Min.   :2.0   Length:517         Length:517
##  1st Qu.:3.000   1st Qu.:4.0   Class :character   Class :character
##  Median :4.000   Median :4.0   Mode  :character   Mode  :character
##  Mean   :4.669   Mean   :4.3
##  3rd Qu.:7.000   3rd Qu.:5.0
##  Max.   :9.000   Max.   :9.0
##       FFMC            DMC              DC             ISI
##  Min.   :18.70   Min.   :  1.1   Min.   :  7.9   Min.   : 0.000
##  1st Qu.:90.20   1st Qu.: 68.6   1st Qu.:437.7   1st Qu.: 6.500
##  Median :91.60   Median :108.3   Median :664.2   Median : 8.400
##  Mean   :90.64   Mean   :110.9   Mean   :547.9   Mean   : 9.022
##  3rd Qu.:92.90   3rd Qu.:142.4   3rd Qu.:713.9   3rd Qu.:10.800
```

```
##  Max.   :96.20   Max.   :291.3   Max.   :860.6   Max.   :56.100  
##       temp             RH              wind            rain        
##  Min.   : 2.20   Min.   : 15.00   Min.   :0.400   Min.   :0.00000  
##  1st Qu.:15.50   1st Qu.: 33.00   1st Qu.:2.700   1st Qu.:0.00000  
##  Median :19.30   Median : 42.00   Median :4.000   Median :0.00000  
##  Mean   :18.89   Mean   : 44.29   Mean   :4.018   Mean   :0.02166  
##  3rd Qu.:22.80   3rd Qu.: 53.00   3rd Qu.:4.900   3rd Qu.:0.00000  
##  Max.   :33.30   Max.   :100.00   Max.   :9.400   Max.   :6.40000  
##       area        
##  Min.   :   0.00  
##  1st Qu.:   0.00  
##  Median :   0.52  
##  Mean   :  12.85  
##  3rd Qu.:   6.57  
##  Max.   :1090.84
```

Bu veri kümesi, Portekiz'in kuzeydoğu bölgesindeki orman yangınlarının etkilediği alanın tahmin edilmesi amacıyla oluşturulmuştur. Veri seti, yangınların koordinatları, yangın tarihi ve çeşitli meteorolojik koşullar gibi bir dizi faktörü içermektedir. Bu faktörler, orman yangınlarının şiddetini ve yayılmasını etkileyen önemli bileşenlerdir. Veri seti, öncelikle "area" adlı çıktı değişkeni üzerine odaklanır. Bu değişken, orman yangınının neden olduğu alanın büyüklüğünü ifade eder.

1. X - x-ekseni koordinatı (1 ile 9 arasında değerler alır).
2. Y - y-ekseni koordinatı (2 ile 9 arasında değerler alır).
3. Month - Yangın ayı (Ocak'tan Aralık'a kadar 12 farklı ayı temsil eden üç harfli kısaltmalarla, örneğin "jan" veya "aug").
4. Day - Yangın günü (Pazartesiden Pazar'a kadar 7 farklı günü temsil eden 1 ile 7 arasında tam sayılar).
5. FFMC - Fine Fuel Moisture Code, nem kodu (18.7 ile 96.20 arasında değerler alır).
6. DMC - Duff Moisture Code, duff nem kodu (1.1 ile 291.3 arasında değerler alır).
7. DC - Drought Code, kuruluk kodu (7.9 ile 860.6 arasında değerler alır).
8. ISI - Initial Spread Index, başlangıç yayılma indeksi (0.0 ile 56.10 arasında değerler alır).
9. Temp - Sıcaklık (2.2 ile 33.30 arasında değerler alır).
10. RH - Relatif Nem (15 ile 100 arasında değerler alır).
11. Wind - Rüzgar hızı (0.40 ile 9.40 arasında değerler alır).
12. Rain - Yağış (0.0 ile 6.4 arasında değerler alır).
13. Area - Orman yangınının etkilediği alanın büyüklüğü (sıfır veya pozitif gerçek sayılar, çoğunlukla sıfıra yakın).

```r
# Ay ve gün eşleme tablolarının oluşturulması
month_map <- c(
  'jan' = 1,
  'feb' = 2,
  'mar' = 3,
  'apr' = 4,
  'may' = 5,
  'jun' = 6,
  'jul' = 7,
  'aug' = 8,
  'sep' = 9,
  'oct' = 10,
  'nov' = 11,
  'dec' = 12
)
```

```r
day_map <- c(
  'sun' = 7,
  'mon' = 1,
  'tue' = 2,
  'wed' = 3,
  'thu' = 4,
  'fri' = 5,
  'sat' = 6
)


# forestfires veri setindeki 'month' sütununu eşleme tablosuyla değiştirme
forestfires <- forestfires %>%
  mutate(month = month_map[month])

# forestfires veri setindeki 'day' sütununu eşleme tablosuyla değiştirme
forestfires <- forestfires %>%
  mutate(day = day_map[day])

# forestfires veri setinin yapısını inceleme
str(forestfires)
```

```
## 'data.frame':    517 obs. of  13 variables:
##  $ X    : int  7 7 7 8 8 8 8 8 8 7 ...
##  $ Y    : int  5 4 4 6 6 6 6 6 6 5 ...
##  $ month: Named num  3 10 10 3 3 8 8 8 9 9 ...
##   ..- attr(*, "names")= chr [1:517] "mar" "oct" "oct" "mar" ...
##  $ day  : Named num  5 2 6 5 7 7 1 1 2 6 ...
##   ..- attr(*, "names")= chr [1:517] "fri" "tue" "sat" "fri" ...
##  $ FFMC : num  86.2 90.6 90.6 91.7 89.3 92.3 92.3 91.5 91 92.5 ...
##  $ DMC  : num  26.2 35.4 43.7 33.3 51.3 ...
##  $ DC   : num  94.3 669.1 686.9 77.5 102.2 ...
##  $ ISI  : num  5.1 6.7 6.7 9 9.6 14.7 8.5 10.7 7 7.1 ...
##  $ temp : num  8.2 18 14.6 8.3 11.4 22.2 24.1 8 13.1 22.8 ...
##  $ RH   : int  51 33 33 97 99 29 27 86 63 40 ...
##  $ wind : num  6.7 0.9 1.3 4 1.8 5.4 3.1 2.2 5.4 4 ...
##  $ rain : num  0 0 0 0.2 0 0 0 0 0 0 ...
##  $ area : num  0 0 0 0 0 0 0 0 0 0 ...
```

```r
# Veri Setinin Bölünmesi
set.seed(123)  # Tekrarlanabilirlik için
sample <-
  sample.int(
    n = nrow(forestfires),
    size = floor(.8 * nrow(forestfires)),
    replace = FALSE
  )

# Eğitim veri setini oluşturma
train_data <- forestfires[sample,]
train_data_ff  <- train_data
glimpse(train_data, width = 44)
```

```
## Rows: 413
## Columns: 13
## $ X     <int> 5, 1, 2, 2, 3, 8, 4, 7, 6, 5~
## $ Y     <int> 4, 4, 5, 2, 4, 6, 6, 4, 5, 4~
## $ month <dbl> 8, 9, 9, 8, 3, 6, 9, 8, 9, 8~
## $ day   <dbl> 7, 7, 3, 2, 6, 3, 7, 7, 1, 4~
## $ FFMC  <dbl> 93.6, 91.0, 90.1, 94.8, 91.7~
## $ DMC   <dbl> 235.1, 276.3, 82.9, 108.3, 3~
## $ DC    <dbl> 723.1, 825.1, 735.7, 647.1, ~
## $ ISI   <dbl> 10.1, 7.1, 6.2, 17.0, 7.8, 1~
## $ temp  <dbl> 24.1, 14.5, 18.3, 24.6, 15.2~
## $ RH    <int> 50, 76, 45, 22, 27, 43, 26, ~
## $ wind  <dbl> 4.0, 7.6, 2.2, 4.5, 4.9, 4.9~
## $ rain  <dbl> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0~
## $ area  <dbl> 0.00, 3.71, 4.88, 10.01, 0.0~
```

```r
# Test veri setini oluşturma
test_data <- forestfires[-sample,]
glimpse(test_data, width = 44)
```

```
## Rows: 104
## Columns: 13
## $ X     <int> 7, 8, 6, 5, 8, 6, 6, 7, 4, 5~
## $ Y     <int> 5, 6, 5, 5, 5, 4, 3, 4, 4, 6~
## $ month <dbl> 3, 9, 9, 3, 10, 3, 10, 10, 9~
## $ day   <dbl> 5, 2, 3, 6, 1, 3, 2, 5, 6, 3~
## $ FFMC  <dbl> 86.2, 91.0, 92.9, 91.7, 84.9~
## $ DMC   <dbl> 26.2, 129.5, 133.3, 35.8, 32~
## $ DC    <dbl> 94.3, 692.6, 699.6, 80.8, 66~
## $ ISI   <dbl> 5.1, 7.0, 9.2, 7.8, 3.0, 6.3~
## $ temp  <dbl> 8.2, 13.1, 26.4, 15.1, 16.7,~
## $ RH    <int> 51, 63, 21, 27, 47, 35, 24, ~
## $ wind  <dbl> 6.7, 5.4, 4.5, 5.4, 4.9, 4.0~
## $ rain  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ area  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

# XGBoost

```r
# XGBoost Modelinin Eğitilmesi
# xgb.DMatrix ile eğitim verisi matrisini ve hedef değişkeni oluşturma
xgb_data <-
  xgb.DMatrix(data = as.matrix(train_data[, -13]), label = train_data$area)

# xgb.train ile XGBoost modelini eğitme
xgb_model <-
  xgb.train(data = xgb_data,
            nrounds = 100,
            objective = "reg:squarederror")

# DALEX::explain ile XGBoost modelini açıklama nesnesi oluşturma
```

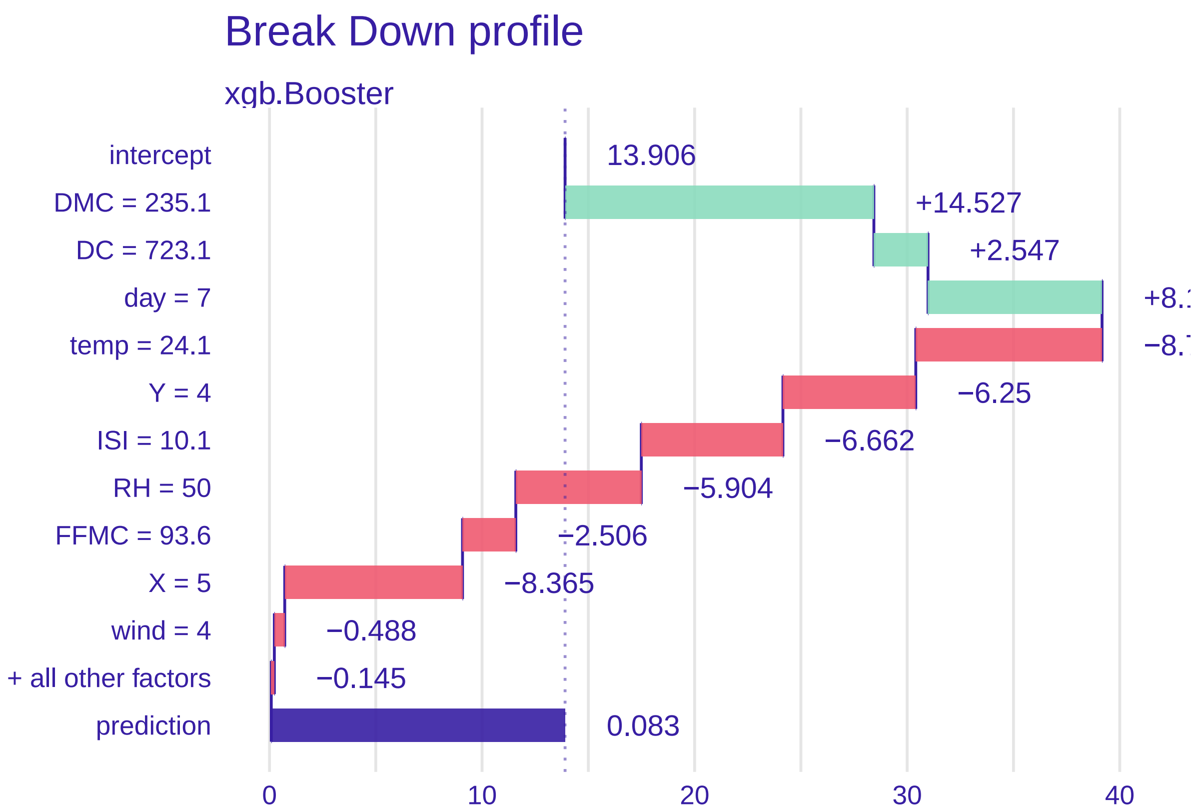
```r
xgb_explainer <-
  DALEX::explain(xgb_model, data = as.matrix(train_data[, -13]), y = train_data$area)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  xgb.Booster  (  default  )
##   -> data              :  413  rows  12  cols
##   -> target variable   :  413  values
##   -> predict function  :  yhat.default will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package Model of class: xgb.Booster package unrecognized , ver. Unknown , t
##   -> predicted values  :  numerical, min =  -0.4886776 , mean =  13.90567 , max =  1090.009
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -6.749003 , mean =  0.0008390863 , max =  5.510997
##   A new explainer has been created!
```

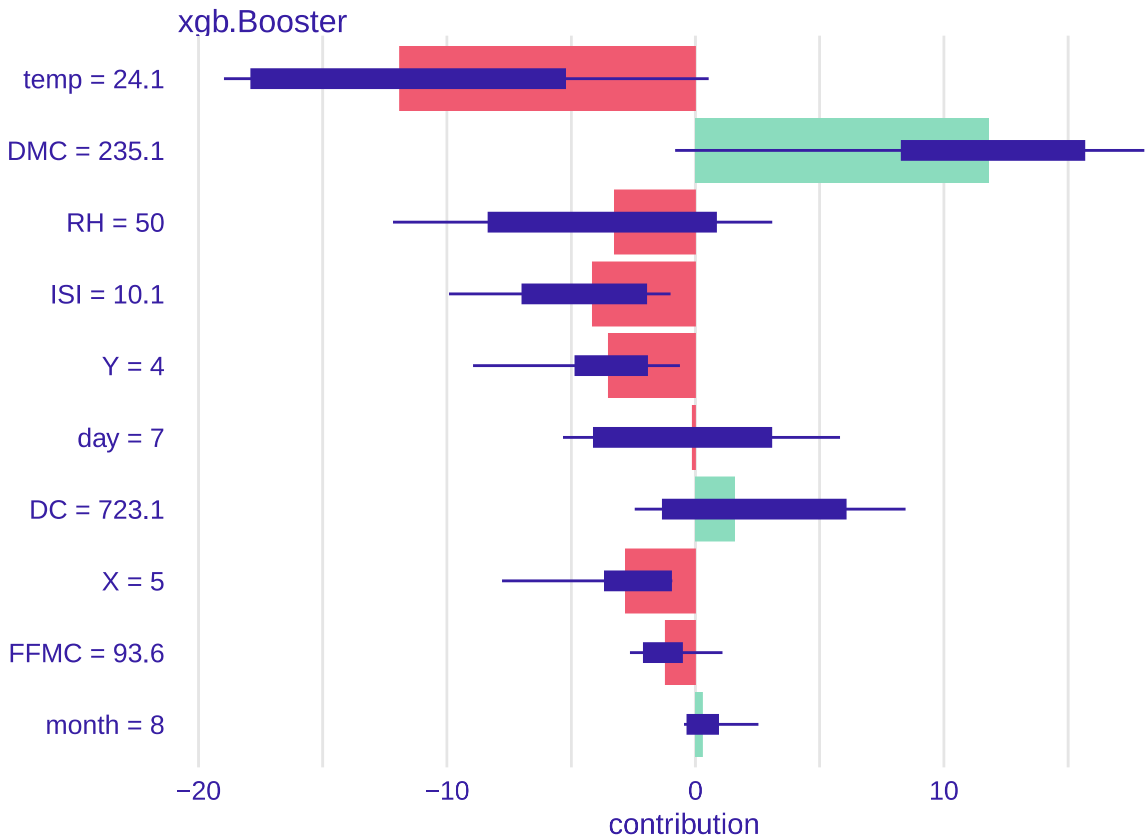
```r
# İlk gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[1, ]
```

```
##     X Y month day FFMC   DMC    DC  ISI temp RH wind rain area
## 415 5 4     8   7 93.6 235.1 723.1 10.1 24.1 50    4    0    0
```

```r
bd_xgb_1 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[1, -13]),
                type = "break_down")
shap_xgb_1 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[1, -13]),
                type = "shap")
plot(bd_xgb_1)
```



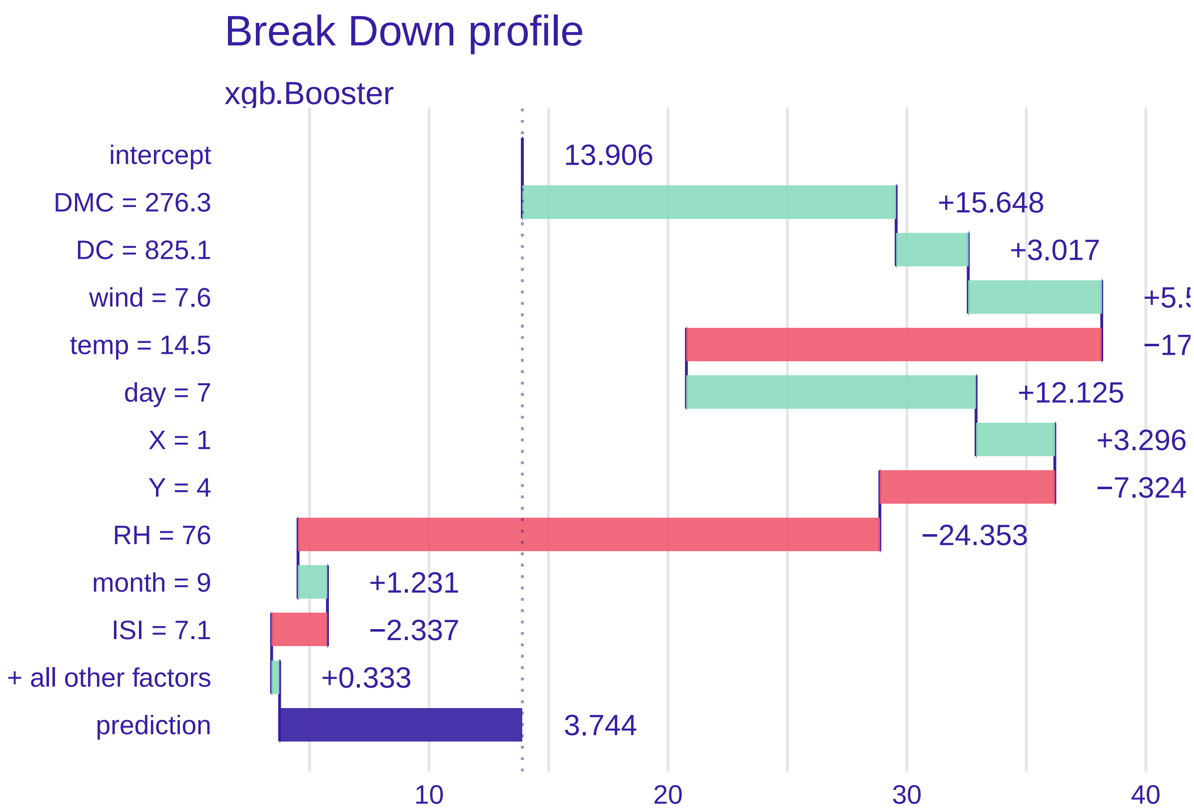

```
plot(shap_xgb_1)
```



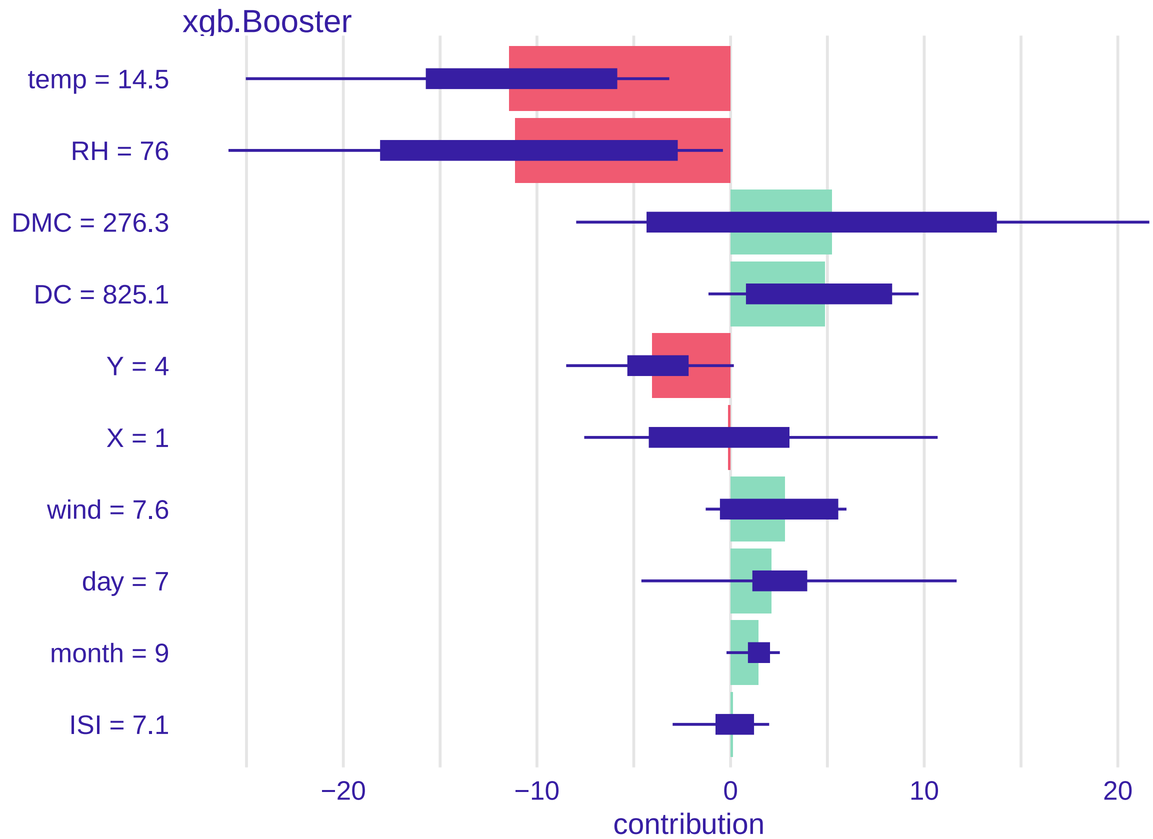

```
# İkinci gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[2, ]
```

```
##       X Y month day FFMC   DMC    DC ISI temp RH wind rain area
## 463 1 4     9   7   91 276.3 825.1 7.1 14.5 76  7.6    0 3.71
```

```
bd_xgb_2 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[2, -13]),
                type = "break_down")
shap_xgb_2 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[2, -13]),
                type = "shap")
plot(bd_xgb_2)
```



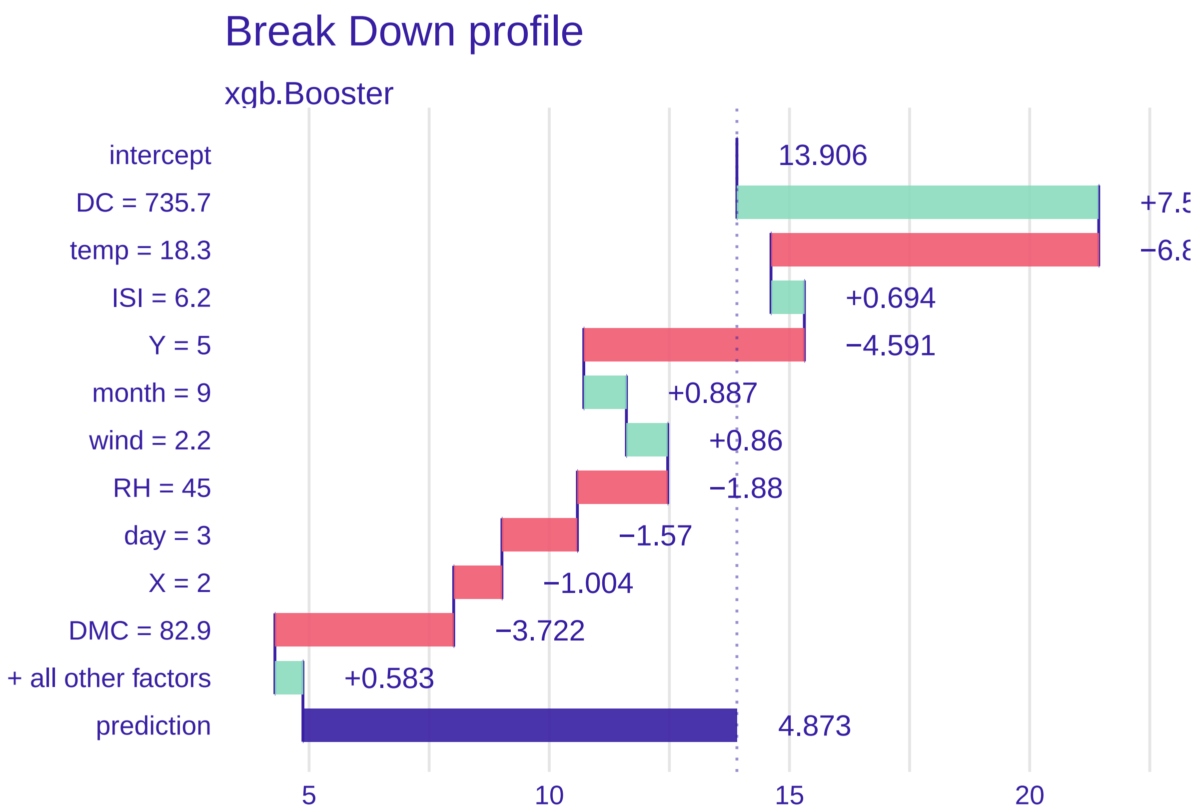

```
plot(shap_xgb_2)
```



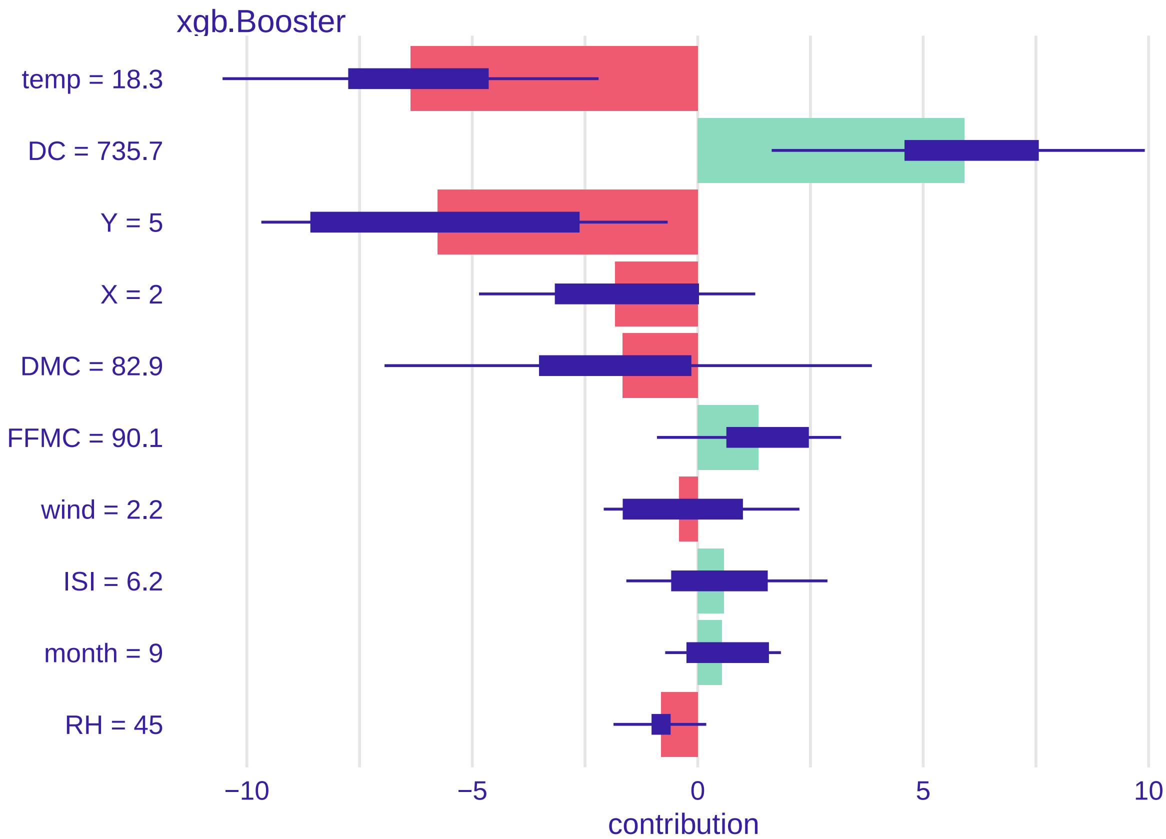

```
# Üçüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[3, ]
```

```
##     X Y month day FFMC  DMC    DC ISI temp RH wind rain area
## 179 2 5     9   3 90.1 82.9 735.7 6.2 18.3 45  2.2    0 4.88
```

```
bd_xgb_3 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[3, -13]),
                type = "break_down")
shap_xgb_3 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[3, -13]),
                type = "shap")
plot(bd_xgb_3)
```



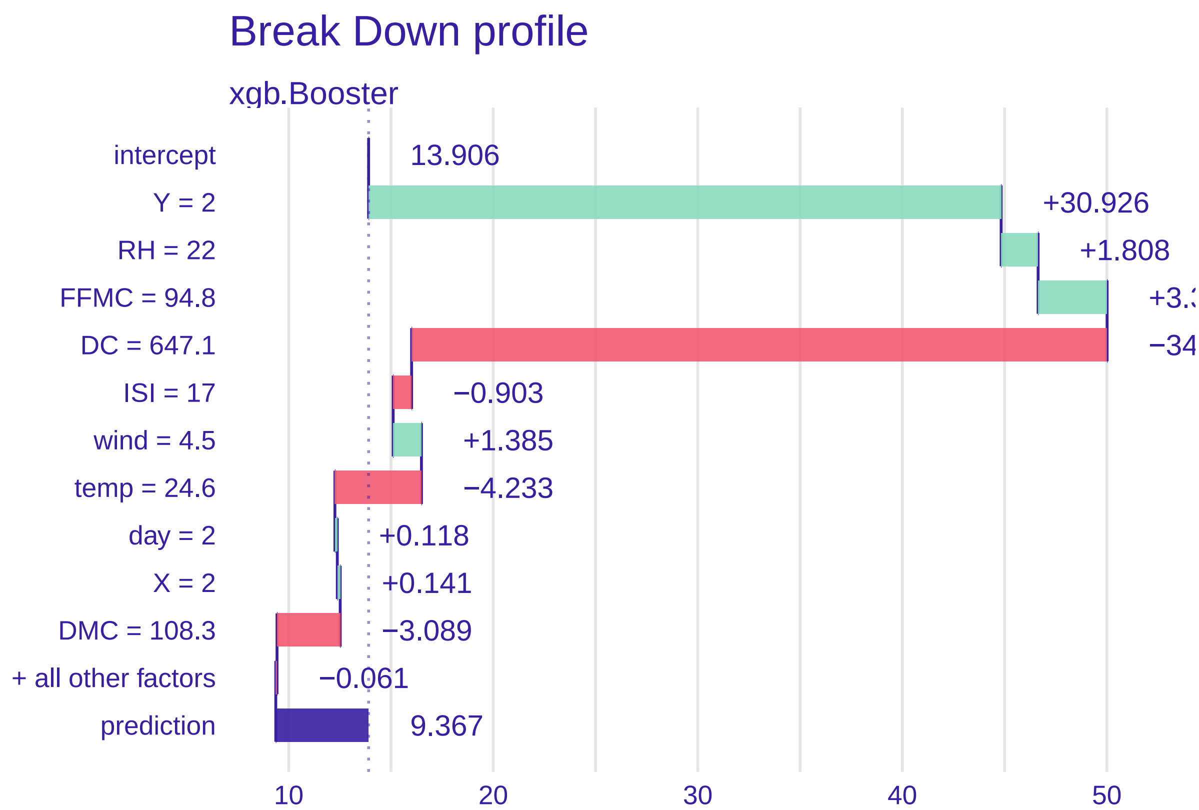

```
plot(shap_xgb_3)
```



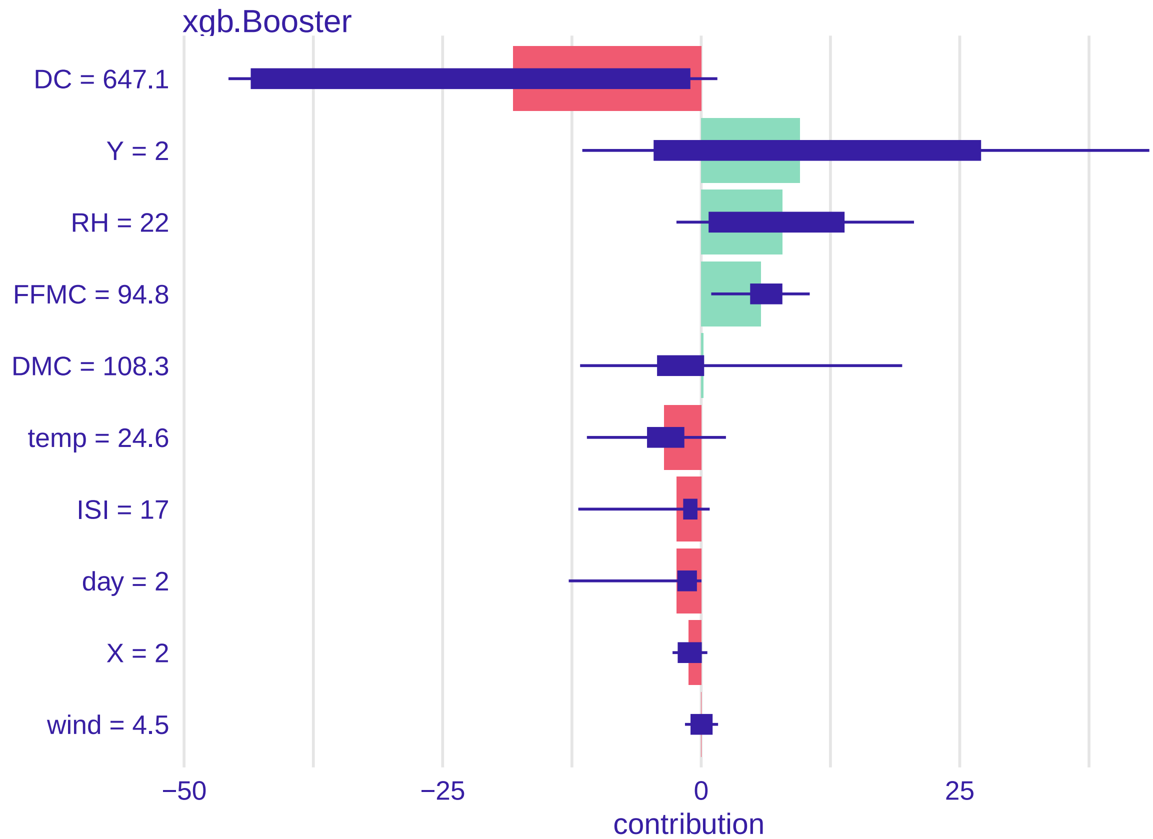

```
# Dördüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[4, ]
```

```
##     X Y month day FFMC   DMC    DC ISI temp RH wind rain  area
## 195 2 2     8   2 94.8 108.3 647.1  17 24.6 22  4.5    0 10.01
```

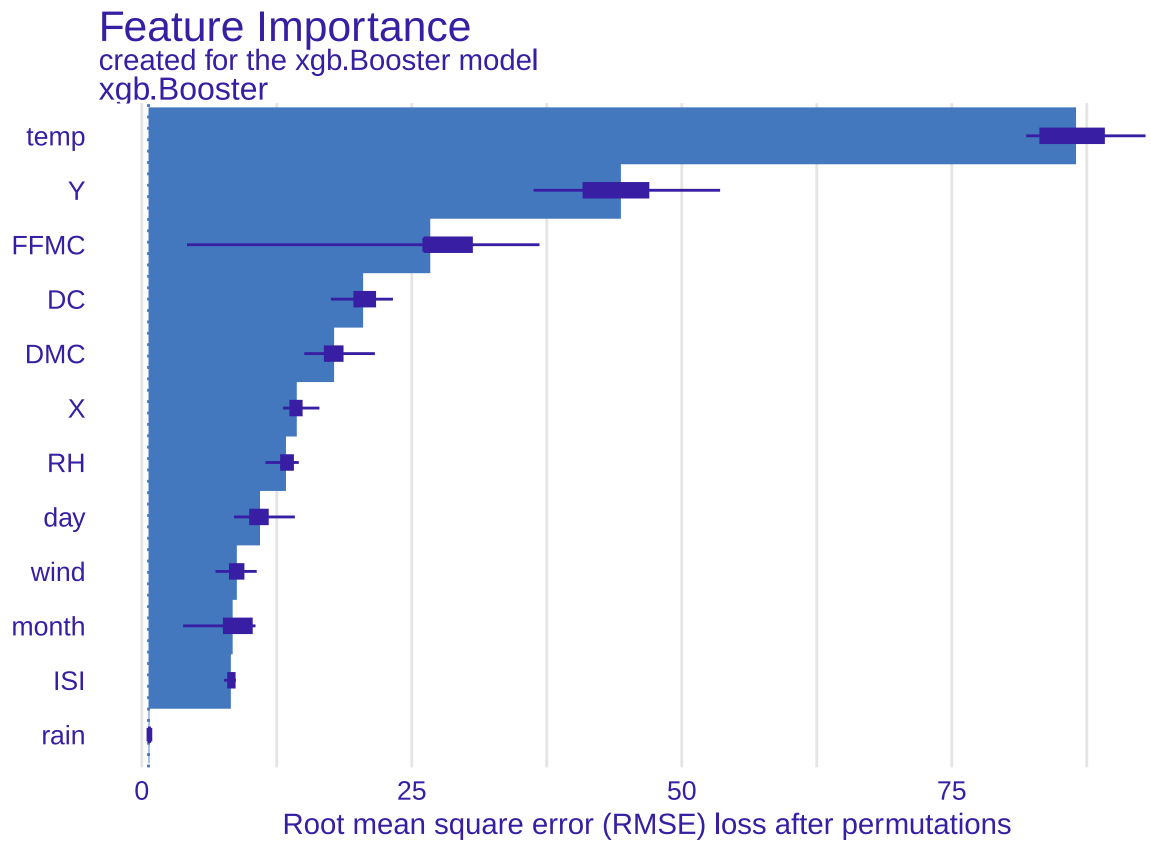
```
bd_xgb_4 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[4, -13]),
                type = "break_down")
shap_xgb_4 <-
  predict_parts(xgb_explainer,
                new_observation = as.matrix(train_data[4, -13]),
                type = "shap")
plot(bd_xgb_4)
```




```
plot(shap_xgb_4)
```




```
# Değişken önem sıralamasını hesaplama ve grafikleme
xgb_var_imp <- DALEX::variable_importance(xgb_explainer)
plot(xgb_var_imp)
```




### lightgbm

```
# LightGBM veri setlerini tanımlama
train_x <- data.matrix(train_data[,-13])
train_y <- data.matrix(train_data[, 13])
test_x <- data.matrix(test_data[,-13])
test_y <- data.matrix(test_data[, 13])
lgb_train <- lgb.Dataset(train_x, label = train_y)
lgb_test <-
  lgb.Dataset.create.valid(lgb_train, test_x, label = test_y)

# LightGBM modelini early stopping ile eğitme
lgb_model <- lgb.train(
  params = list('verbose' = -1),
  # Çalışma sırasında çıktıyı gizler
  data = lgb_train,
  nrounds = 100,
  # Maksimum iterasyon sayısı
  valids = list(test = lgb_test) # Doğrulama (validation) seti
)

# LightGBM modelini açıklama nesnesiyle açıklama
lgb_explainer <-
```

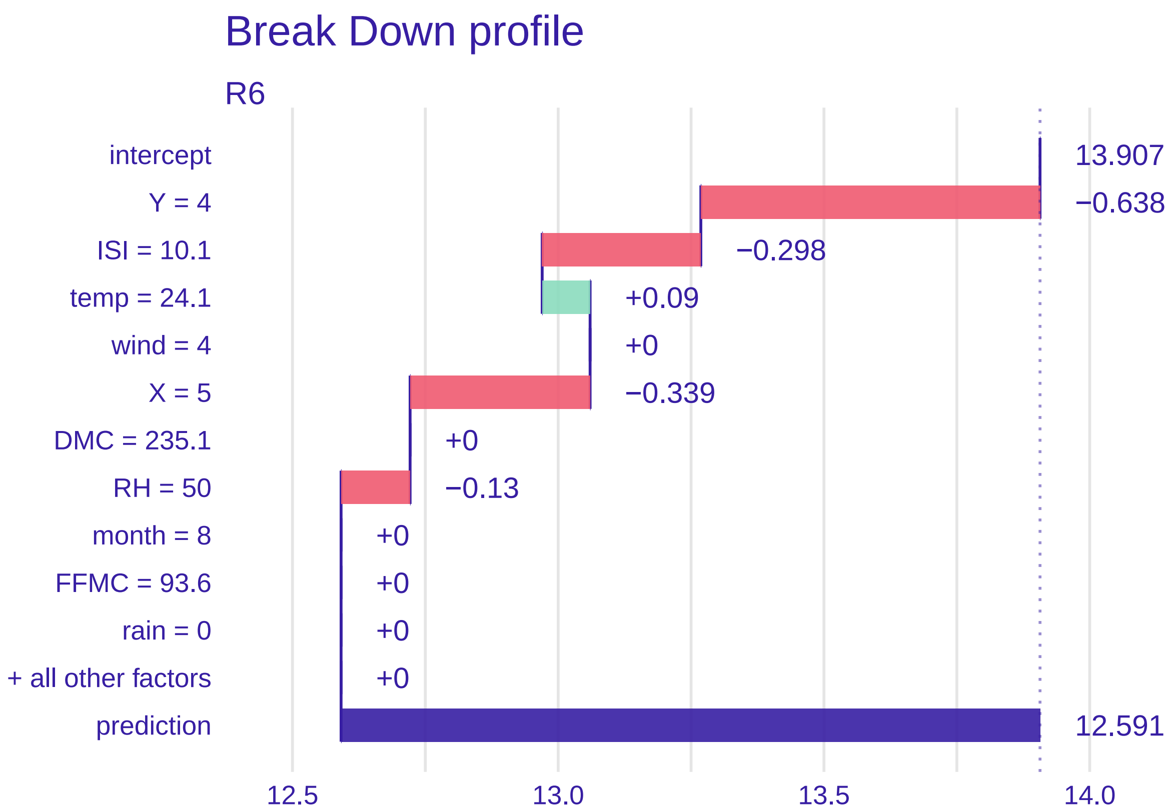
```
  DALEX::explain(lgb_model, data = as.matrix(train_data[, -13]),
                 y = train_data$area)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  R6  (  default  )
##   -> data              :  413  rows  12  cols
##   -> target variable   :  413  values
##   -> predict function  :  yhat.default will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package Model of class: lgb.Booster package unrecognized , ver. Unknown , t
##   -> predicted values  :  numerical, min =  12.5915 , mean =  13.90651 , max =  22.42649
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -22.42649 , mean =  -3.257473e-08 , max =  1068.414
##   A new explainer has been created!
```

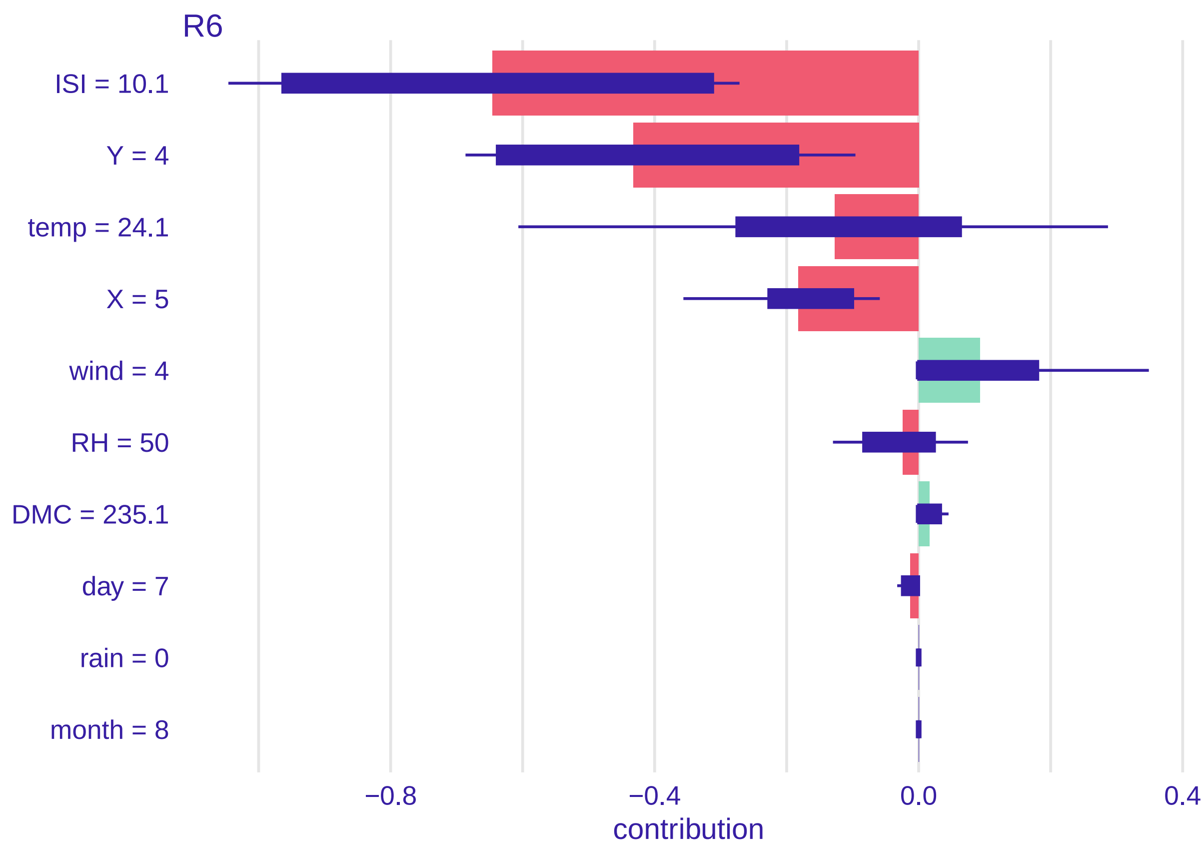
```
# İlk gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[1, ]
```

```
##     X Y month day FFMC   DMC    DC  ISI temp RH wind rain area
## 415 5 4     8   7 93.6 235.1 723.1 10.1 24.1 50    4    0    0
```

```
bd_lgb_1 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[1, -13]),
                type = "break_down")
shap_lgb_1 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[1, -13]),
                type = "shap")
plot(bd_lgb_1)
```



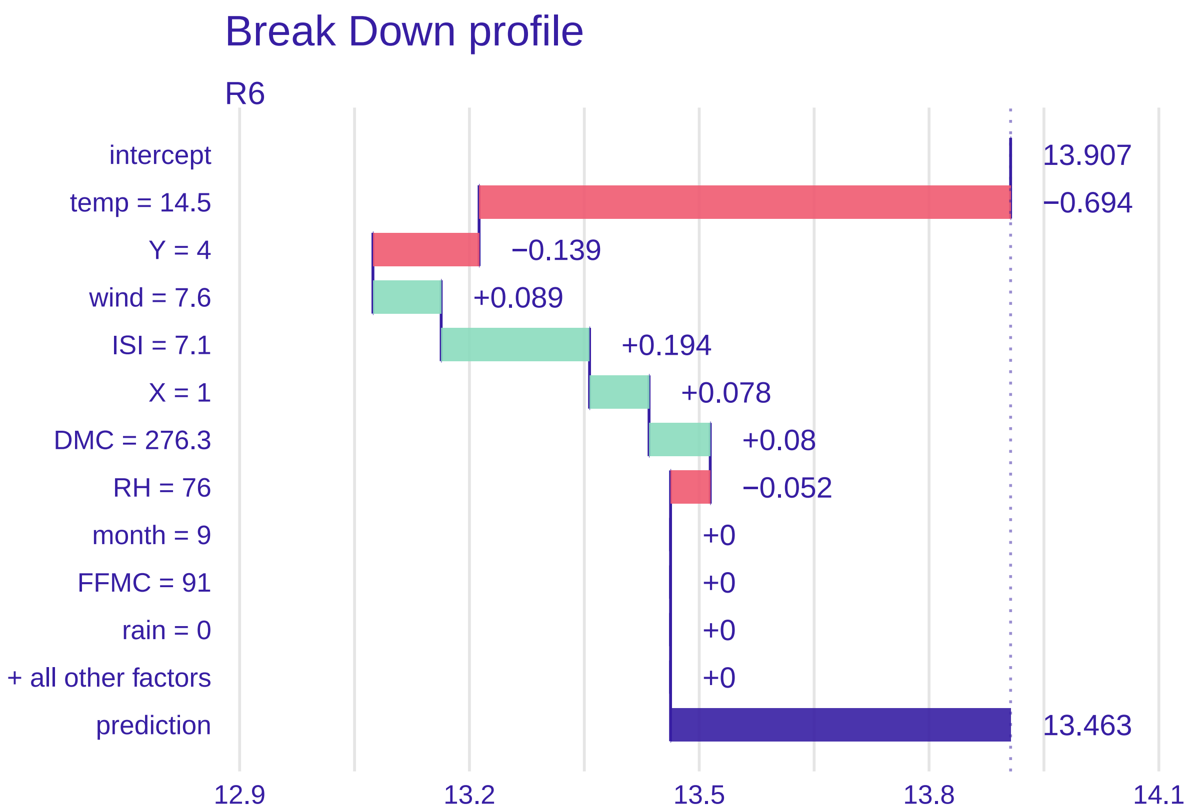

```
plot(shap_lgb_1)
```



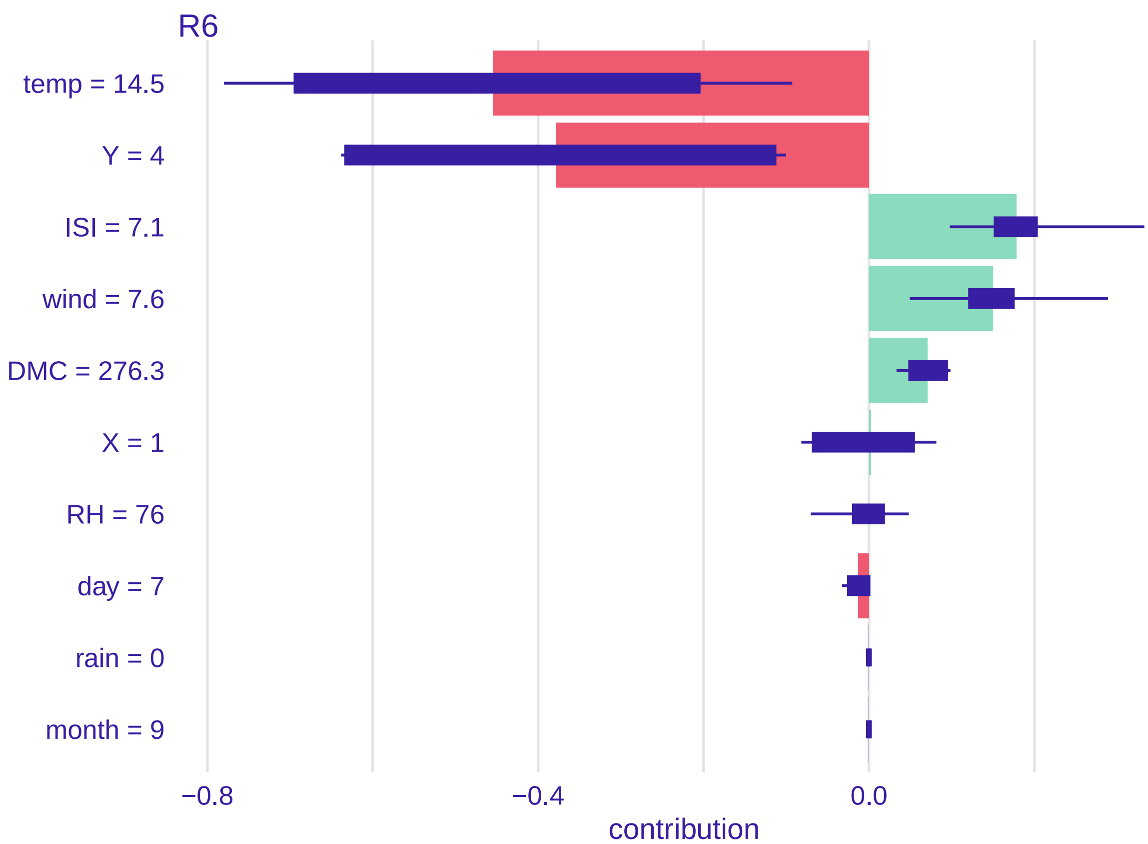

```
# İkinci gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[2, ]
```

```
##       X Y month day FFMC   DMC    DC ISI temp RH wind rain area
## 463 1 4     9   7   91 276.3 825.1 7.1 14.5 76  7.6    0 3.71
```

```
bd_lgb_2 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[2, -13]),
                type = "break_down")
shap_lgb_2 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[2, -13]),
                type = "shap")
plot(bd_lgb_2)
```



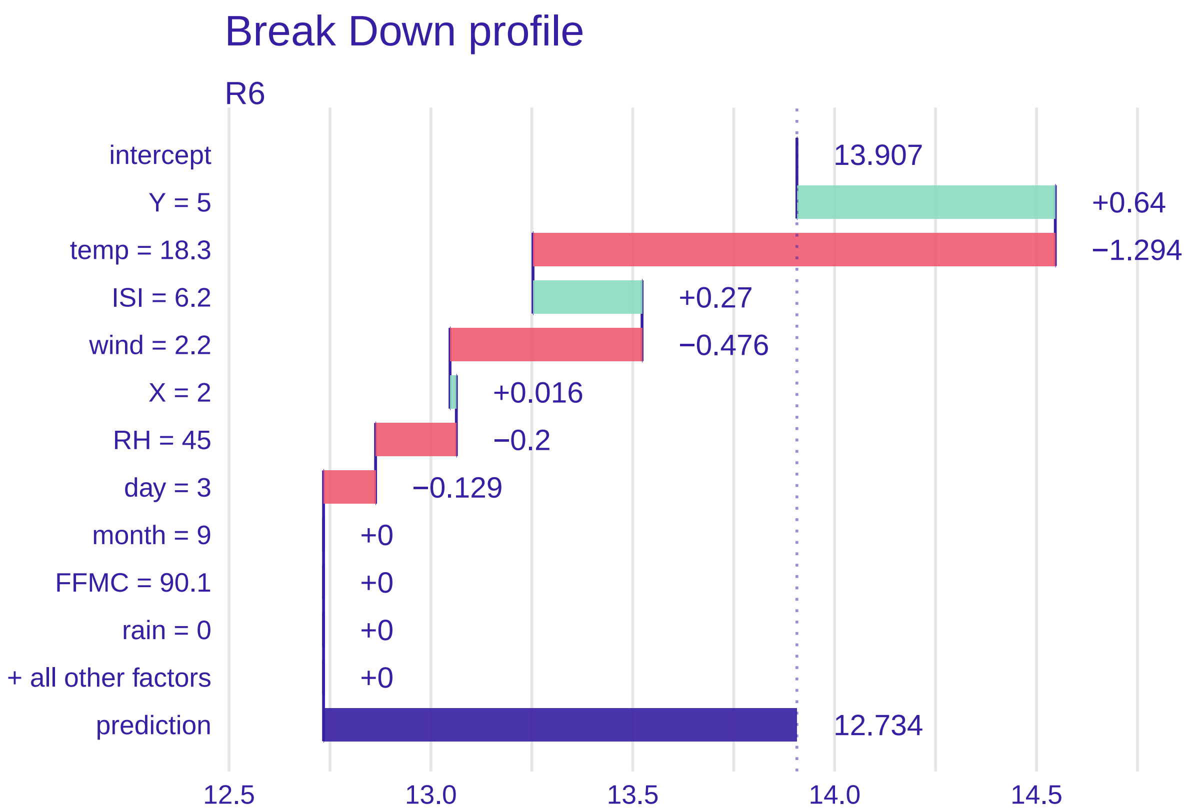

```
plot(shap_lgb_2)
```



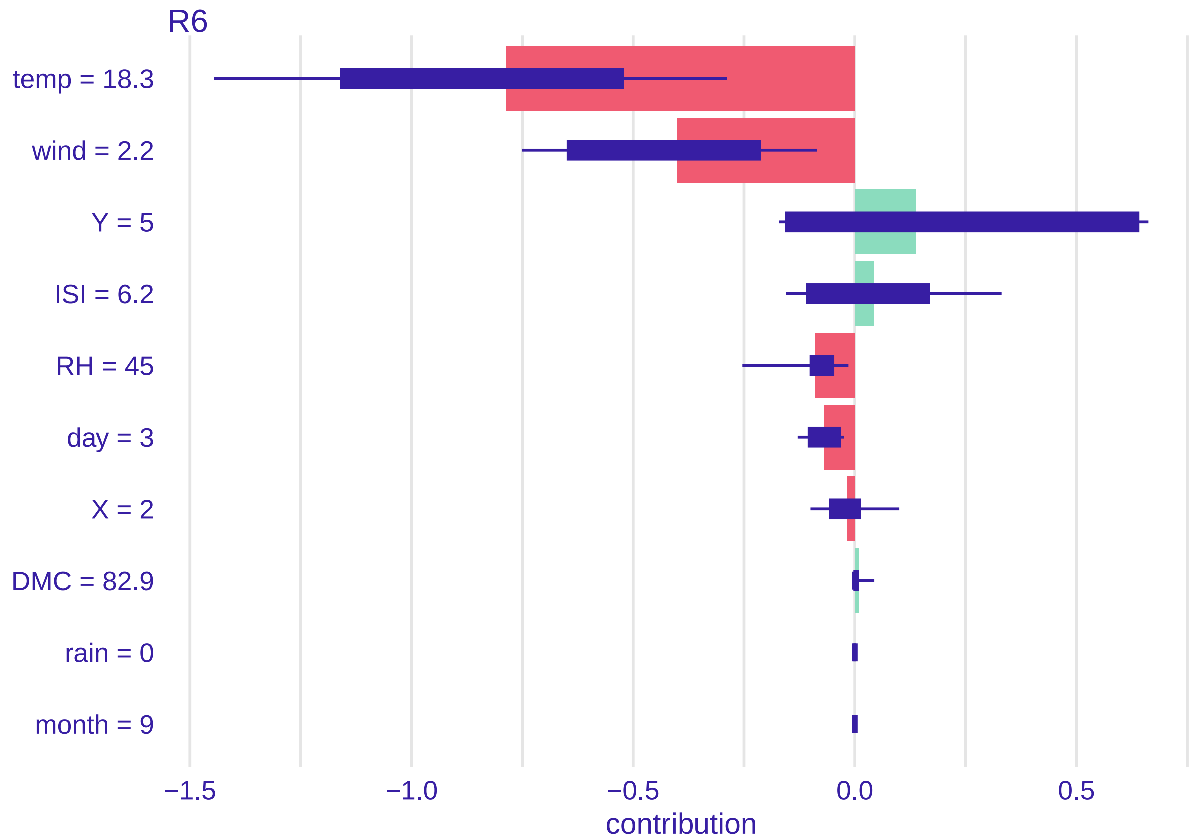

```
# Üçüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[3, ]
```

```
##     X Y month day FFMC  DMC    DC ISI temp RH wind rain area
## 179 2 5     9   3 90.1 82.9 735.7 6.2 18.3 45  2.2    0 4.88
```

```
bd_lgb_3 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[3, -13]),
                type = "break_down")
shap_lgb_3 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[3, -13]),
                type = "shap")
plot(bd_lgb_3)
```



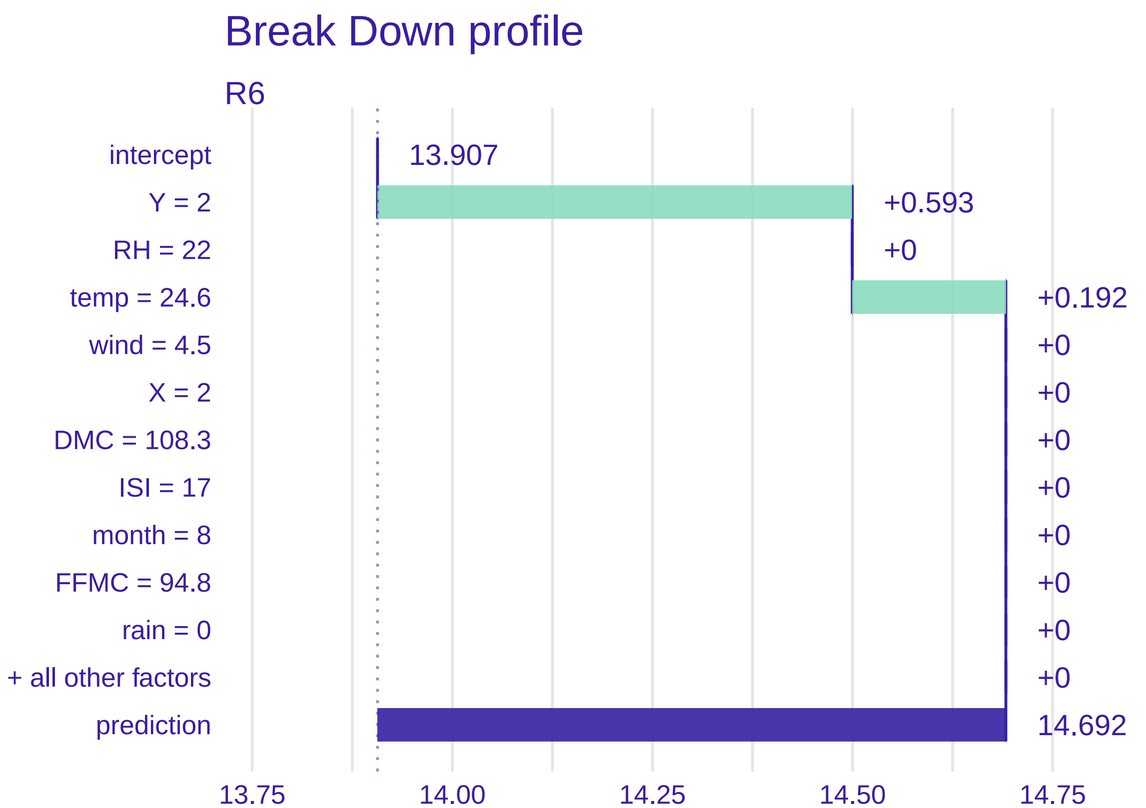

```
plot(shap_lgb_3)
```



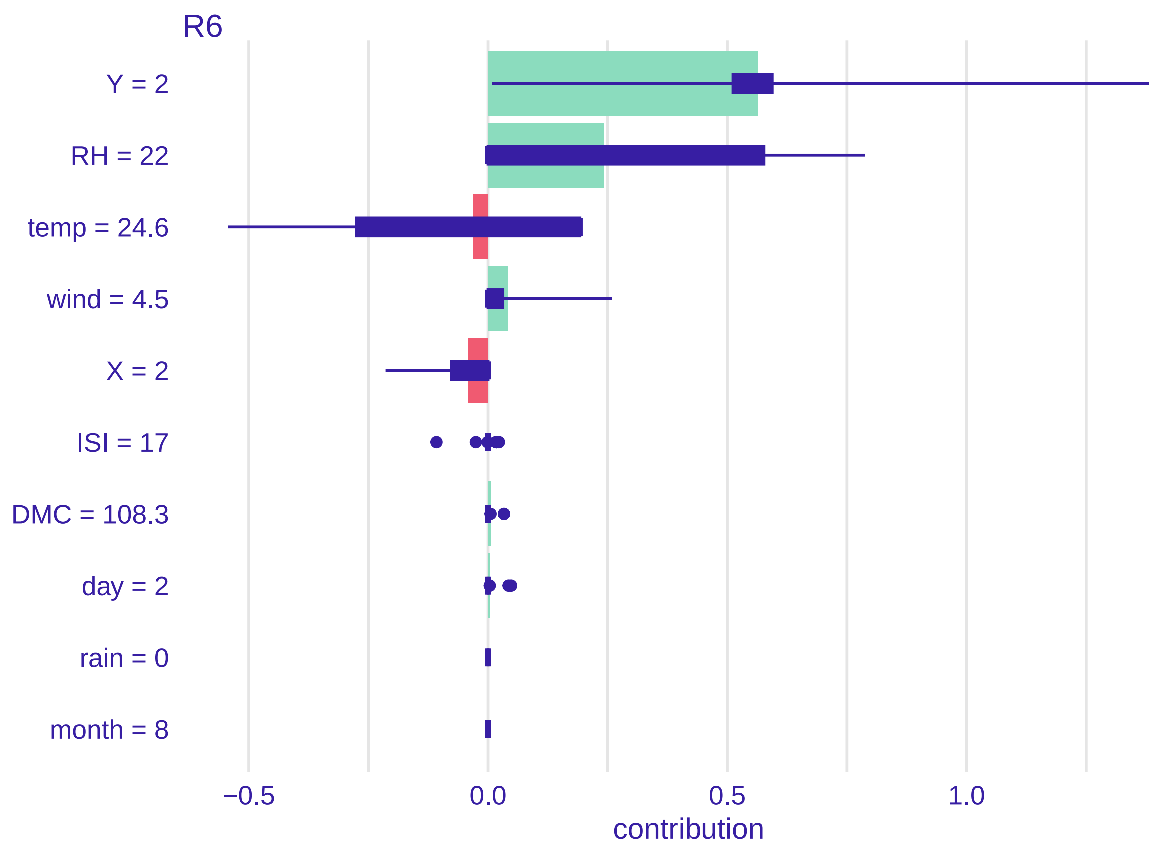

```
# Dördüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[4, ]
```

```
##       X Y month day FFMC   DMC    DC ISI temp RH wind rain  area
## 195 2 2     8   2 94.8 108.3 647.1  17 24.6 22  4.5    0 10.01
```

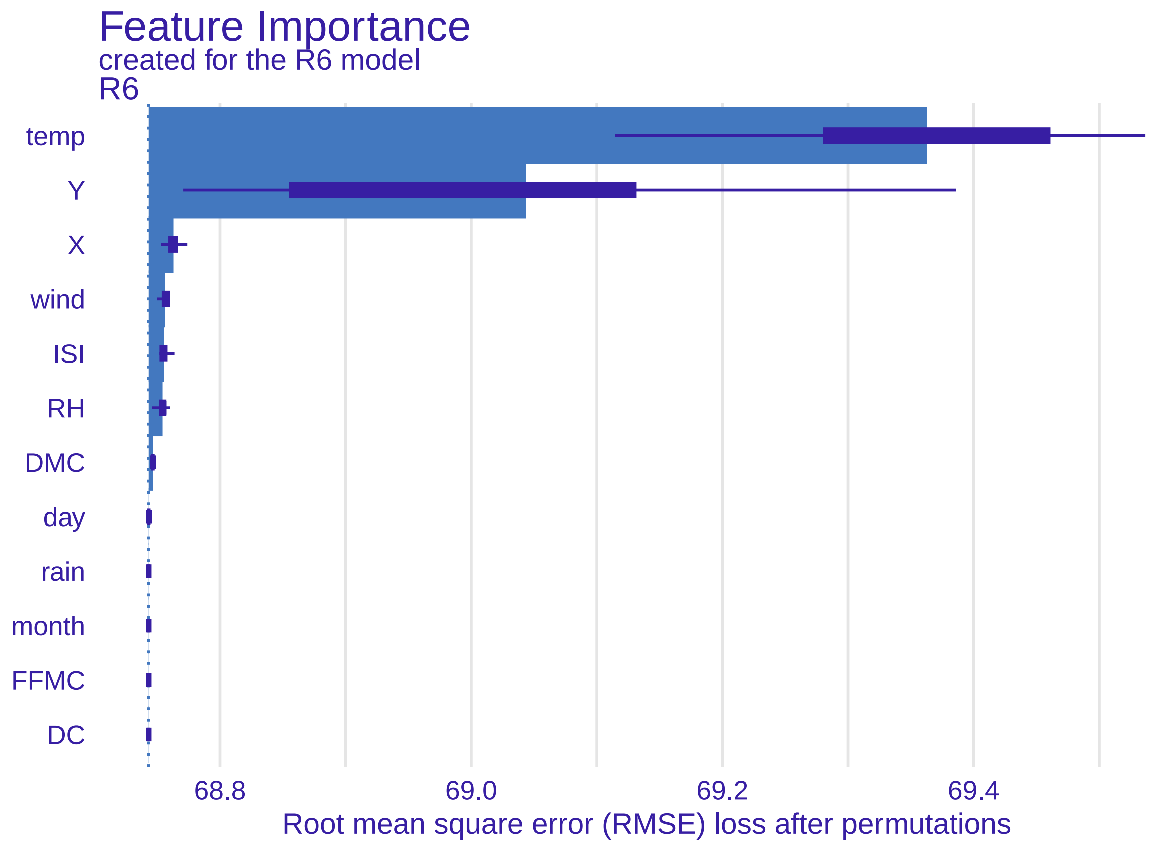
```
bd_lgb_4 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[4, -13]),
                type = "break_down")
shap_lgb_4 <-
  predict_parts(lgb_explainer,
                new_observation = as.matrix(train_data[4, -13]),
                type = "shap")
plot(bd_lgb_4)
```




```
plot(shap_lgb_4)
```




```
# Değişken önem sıralamasını hesaplama ve grafikleme
lgb_var_imp <- DALEX::variable_importance(lgb_explainer)
plot(lgb_var_imp)
```




## gbm

```r
# GBM modelini oluşturma
gbm_model <-
  gbm(area ~ .,
      data = train_data,
      n.trees = 100,
      distribution = "gaussian")

# GBM modeli ile tahmin yapma
gbm_predictions <- predict(gbm_model, test_data[, -13])

# GBM modelinin ortalama karesel hatasını hesaplama
gbm_mse <- mean((test_data$area - gbm_predictions) ^ 2)

# GBM modelini açıklama nesnesiyle açıklama
gbm_explainer <-
  DALEX::explain(gbm_model, data = train_data[,-13], y = train_data$area)
```

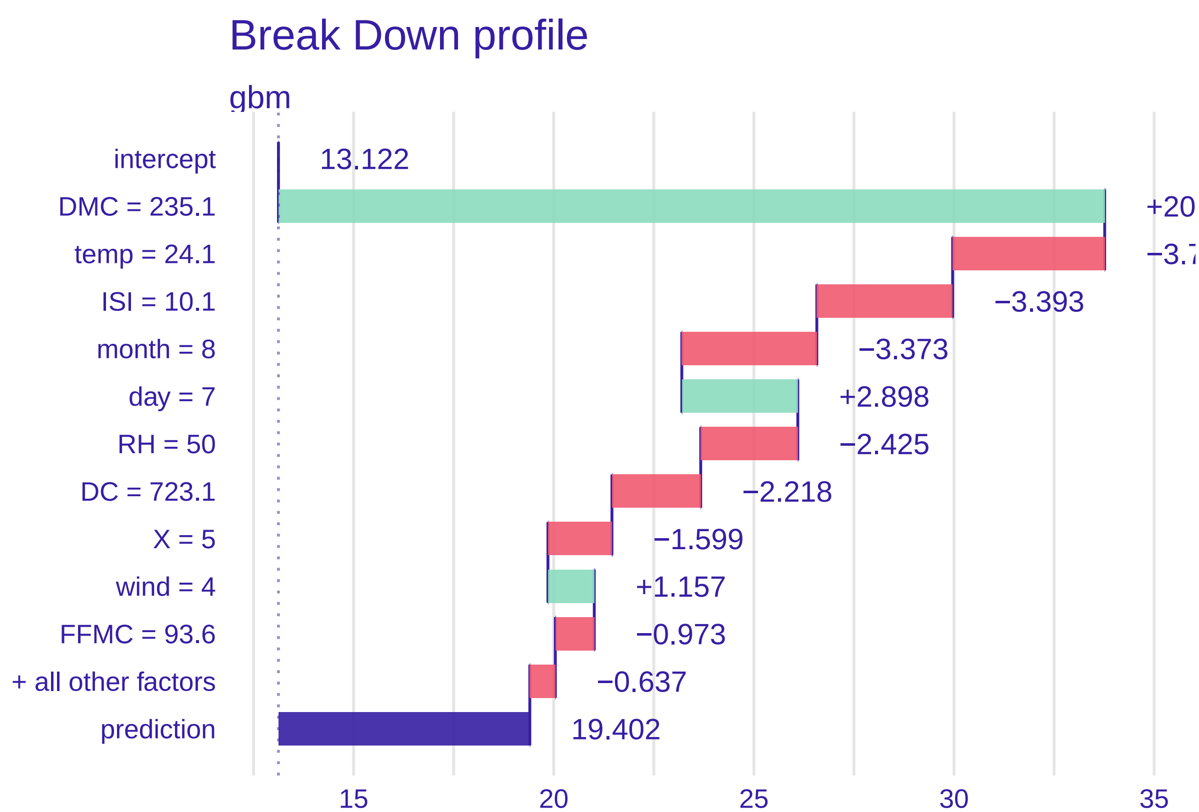
```
## Preparation of a new explainer is initiated
##   -> model label       :  gbm  (  default  )
##   -> data              :  413  rows  12  cols
##   -> target variable   :  413  values
```

```
##    -> predict function   :  yhat.gbm  will be used (  default  )
##    -> predicted values   :  No value for predict function target column. (  default  )
##    -> model_info         :  package gbm , ver. 2.1.8.1 , task regression (  default  )
##    -> predicted values   :  numerical, min =  -6.030842 , mean =  13.12233 , max =  91.90969
##    -> residual function :  difference between y and yhat (  default  )
##    -> residuals          :  numerical, min =  -59.41337 , mean =  0.7841794 , max =  998.9303
##    A new explainer has been created!
```

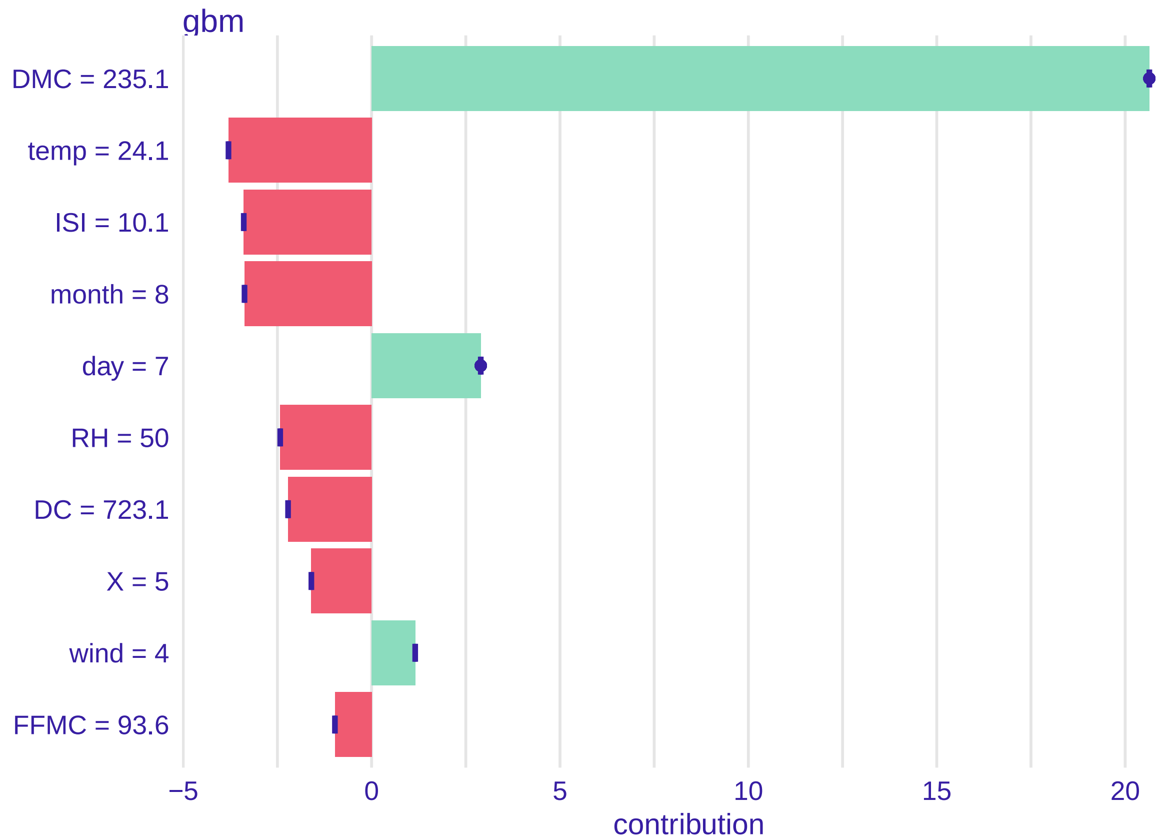
```
# İlk gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[1,]
```

```
##      X Y month day FFMC   DMC    DC  ISI temp RH wind rain area
## 415 5 4      8   7 93.6 235.1 723.1 10.1 24.1 50    4    0    0
```

```
bd_gbm_1 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[1,-13],
                type = "break_down")
shap_gb_1 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[1,-13],
                type = "shap")
plot(bd_gbm_1)
```



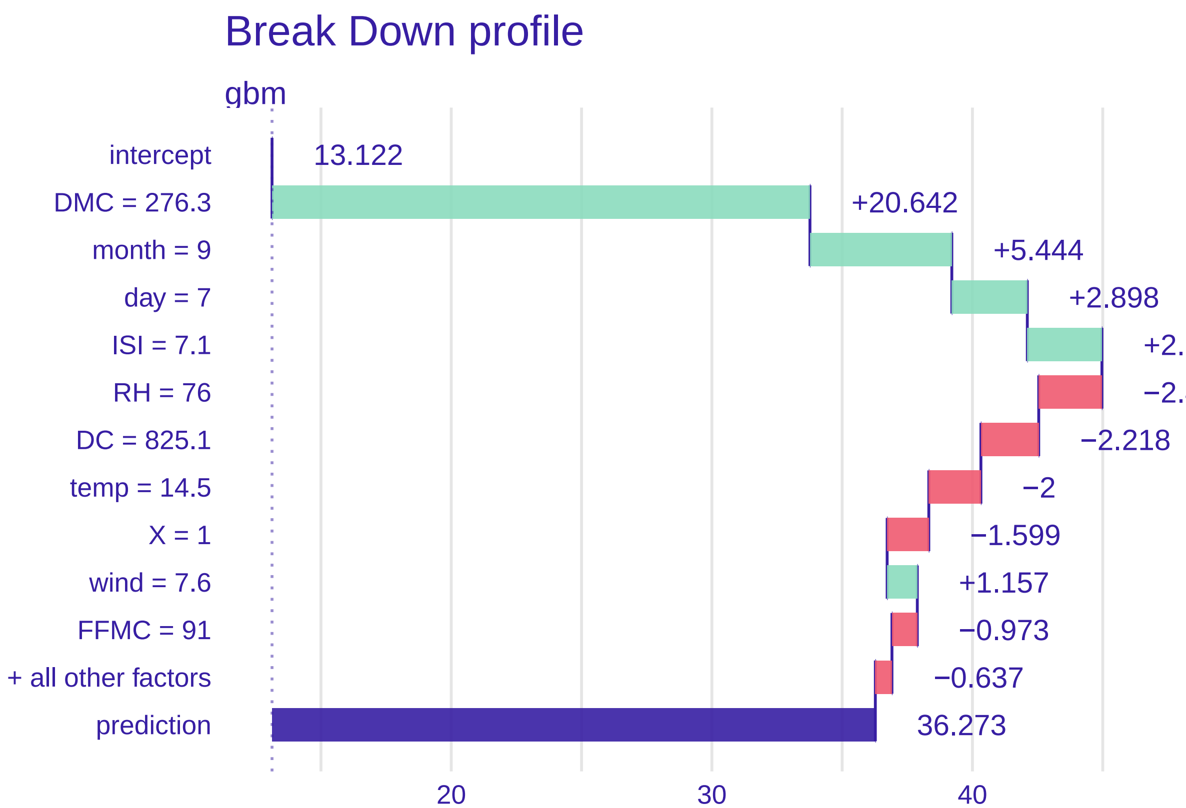
```
plot(shap_gb_1)
```



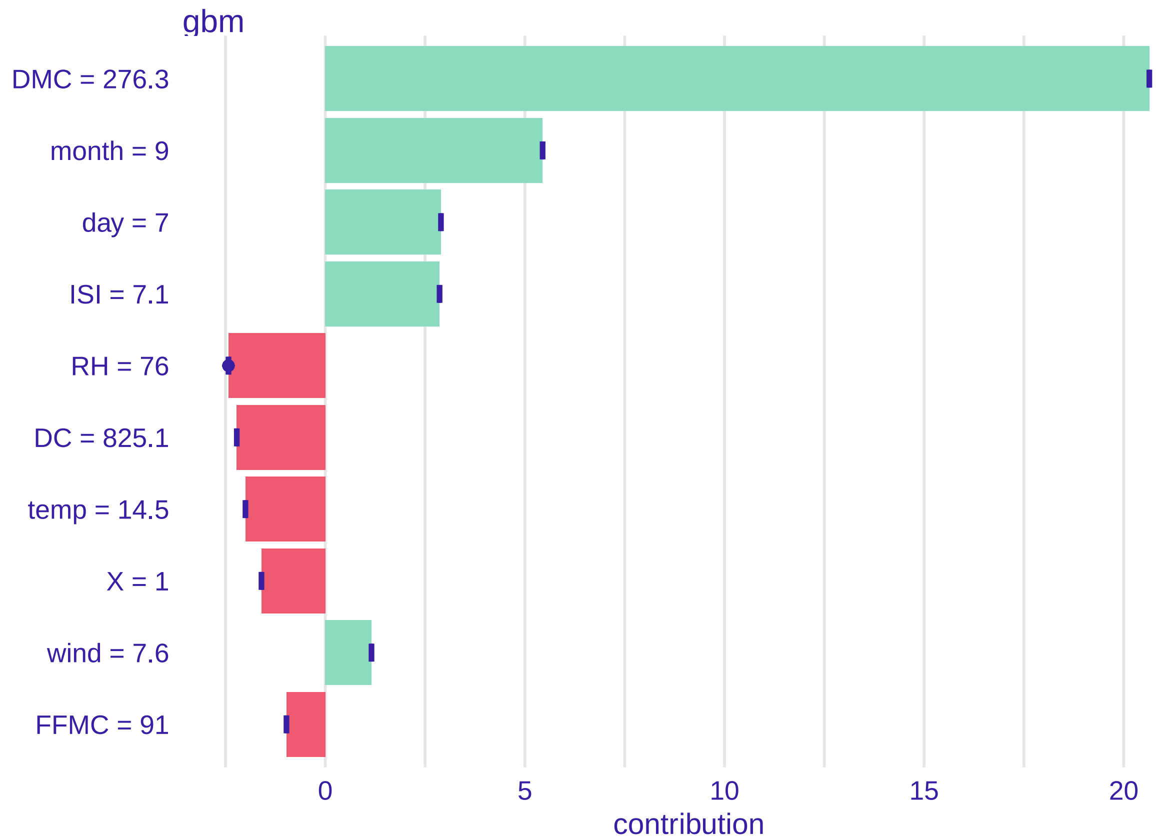

```
# İkinci gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[2,]
```

```
##      X Y month day FFMC   DMC    DC ISI temp RH wind rain area
## 463 1 4     9   7   91 276.3 825.1 7.1 14.5 76  7.6    0 3.71
```

```
bd_gbm_2 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[2,-13],
                type = "break_down")
shap_gb_2 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[2,-13],
                type = "shap")
plot(bd_gbm_2)
```



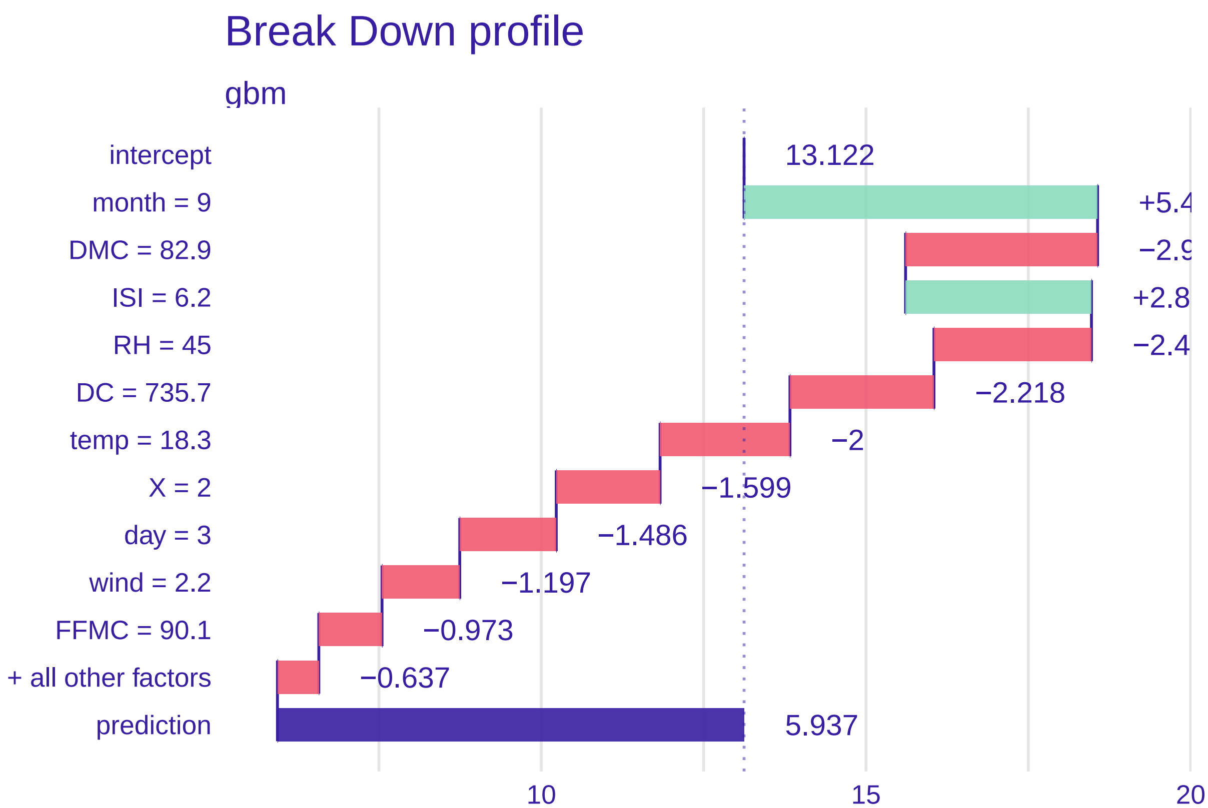

```
plot(shap_gb_2)
```



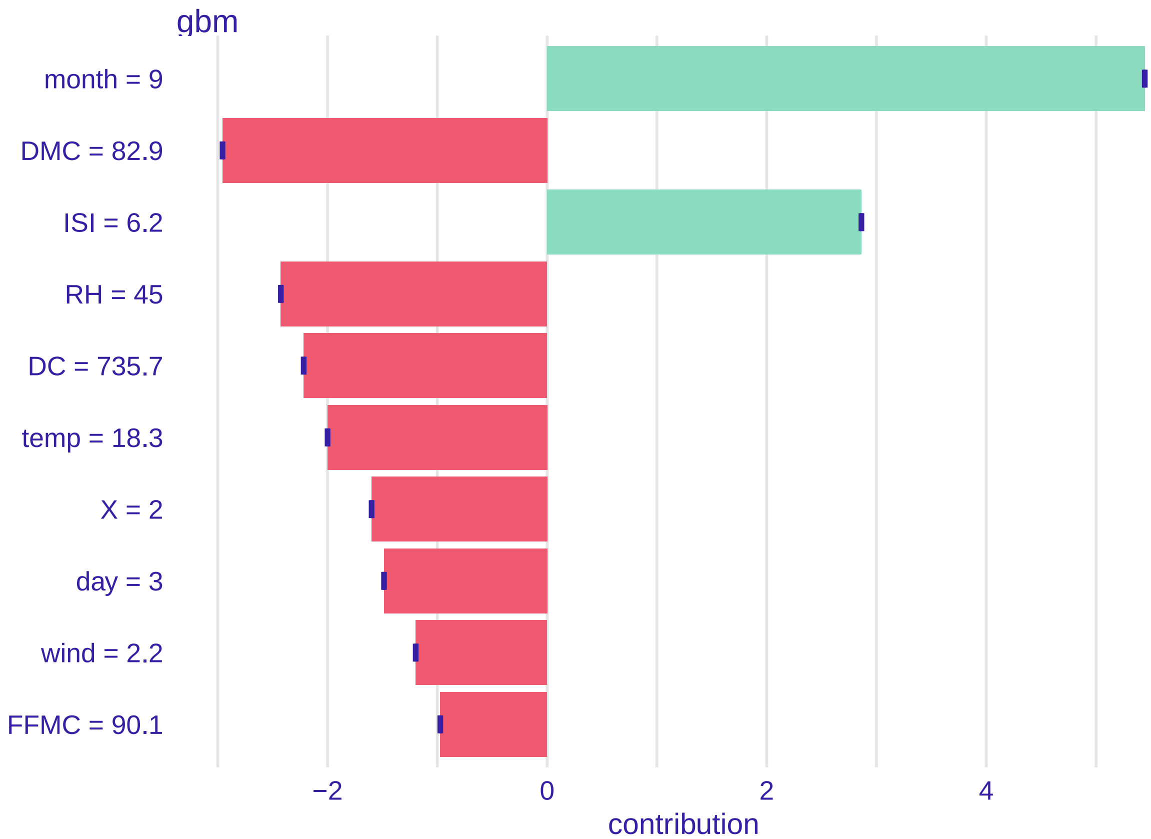

```
# Üçüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[3,]
```

```
##     X Y month day FFMC  DMC    DC ISI temp RH wind rain area
## 179 2 5     9   3 90.1 82.9 735.7 6.2 18.3 45  2.2    0 4.88
```

```
bd_gbm_3 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[3,-13],
                type = "break_down")
shap_gb_3 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[3,-13],
                type = "shap")
plot(bd_gbm_3)
```



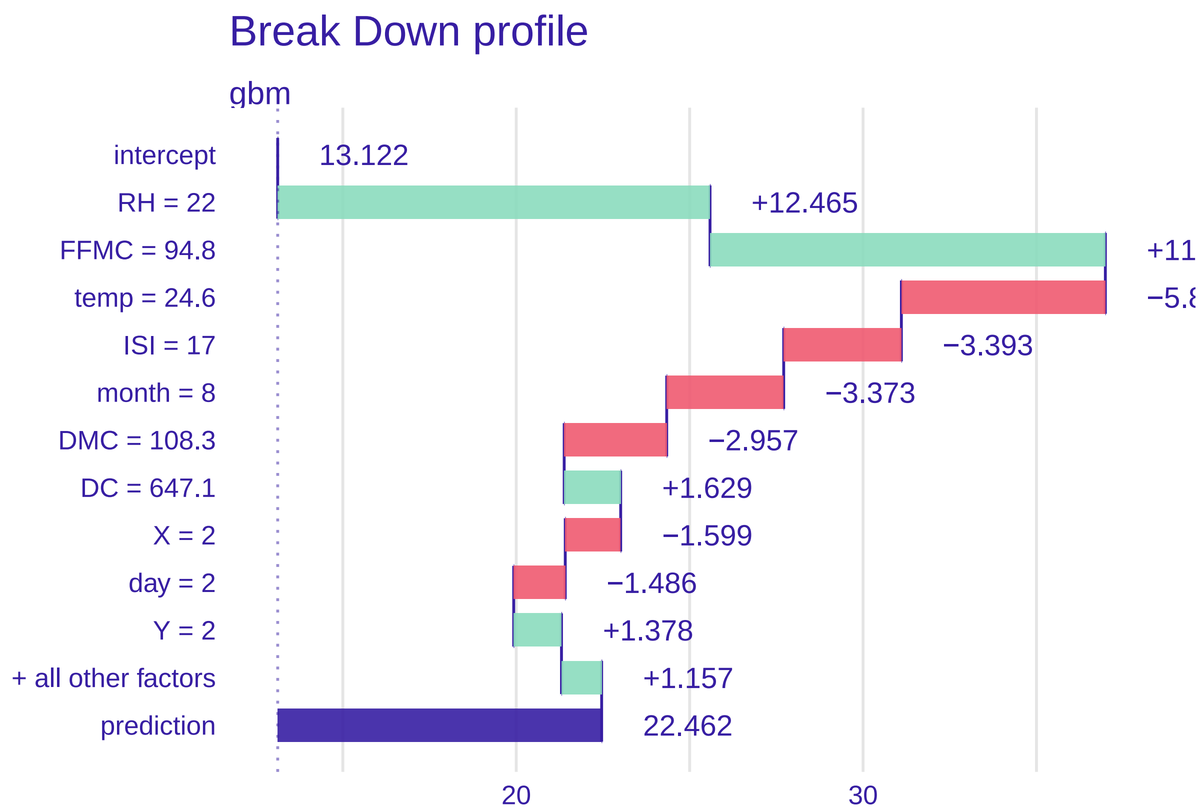

```
plot(shap_gb_3)
```



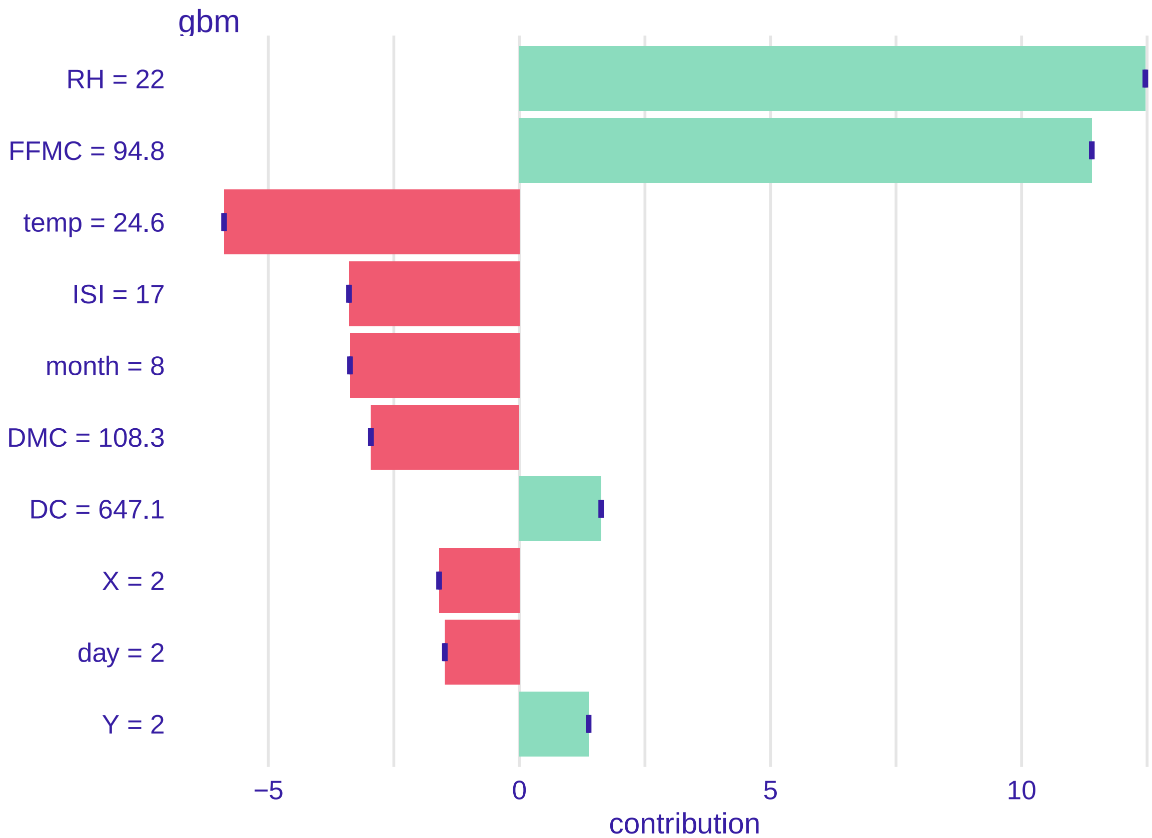

```
# Dördüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[4,]
```

```
##     X Y month day FFMC   DMC    DC ISI temp RH wind rain  area
## 195 2 2     8   2 94.8 108.3 647.1  17 24.6 22  4.5    0 10.01
```

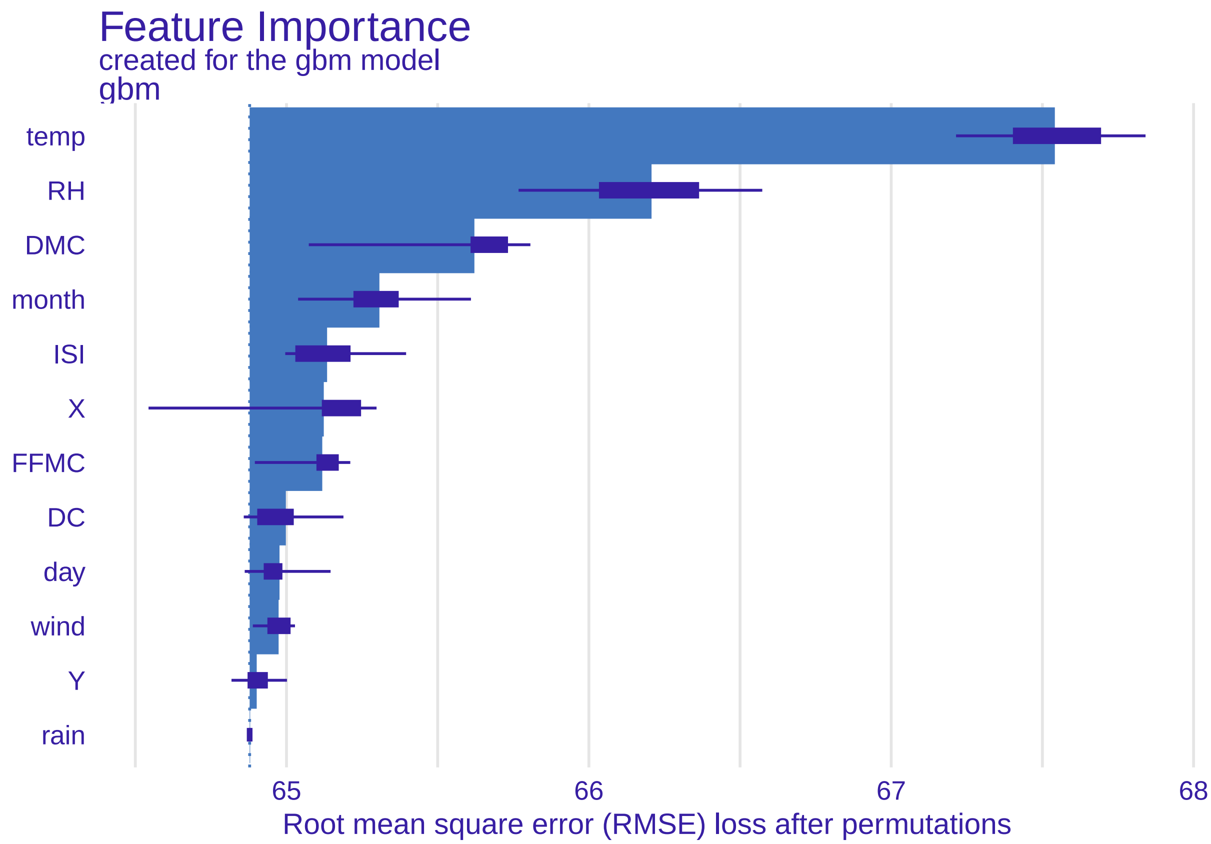
```
bd_gbm_4 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[4,-13],
                type = "break_down")
shap_gb_4 <-
  predict_parts(gbm_explainer,
                new_observation = train_data[4,-13],
                type = "shap")
plot(bd_gbm_4)
```




```
plot(shap_gb_4)
```




```
# GBM modelinin değişken önem sıralamasını hesaplama ve grafikleme
gbm_var_imp <- DALEX::variable_importance(gbm_explainer)
plot(gbm_var_imp)
```




## ionosphere

```
# Ionosphere veri setini CSV dosyasından okuma
ionosphere <-
  read.csv(paste0(getwd(),"/ionosphere.data.csv"),
           header = FALSE)

# Hedef değişkeni binary hale getirme: "g" ise 1, "b" ise 0
ionosphere$V35 <- ifelse(ionosphere$V35 == "g", 1, 0)

# Veri setinin genel yapısını inceleme
glimpse(ionosphere)
```

```
## Rows: 351
## Columns: 35
## $ V1  <int> 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, ~
## $ V2  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ V3  <dbl> 0.99539, 1.00000, 1.00000, 1.00000, 1.00000, 0.02337, 0.97588, 0.0~
## $ V4  <dbl> -0.05889, -0.18829, -0.03365, -0.45161, -0.02401, -0.00592, -0.106~
## $ V5  <dbl> 0.85243, 0.93035, 1.00000, 1.00000, 0.94140, -0.09924, 0.94601, 0.~
## $ V6  <dbl> 0.02306, -0.36156, 0.00485, 1.00000, 0.06531, -0.11949, -0.20800, ~
## $ V7  <dbl> 0.83398, -0.10868, 1.00000, 0.71216, 0.92106, -0.00763, 0.92806, 1~
## $ V8  <dbl> -0.37708, -0.93597, -0.12062, -1.00000, -0.23255, -0.11824, -0.283~
```

```
## $ V9  <dbl> 1.00000, 1.00000, 0.88965, 0.00000, 0.77152, 0.14706, 0.85996, 0.0~
## $ V10 <dbl> 0.03760, -0.04549, 0.01198, 0.00000, -0.16399, 0.06637, -0.27342, ~
## $ V11 <dbl> 0.85243, 0.50874, 0.73082, 0.00000, 0.52798, 0.03786, 0.79766, -1.~
## $ V12 <dbl> -0.17755, -0.67743, 0.05346, 0.00000, -0.20275, -0.06302, -0.47929~
## $ V13 <dbl> 0.59755, 0.34432, 0.85443, 0.00000, 0.56409, 0.00000, 0.78225, 0.0~
## $ V14 <dbl> -0.44945, -0.69707, 0.00827, 0.00000, -0.00712, 0.00000, -0.50764,~
## $ V15 <dbl> 0.60536, -0.51685, 0.54591, -1.00000, 0.34395, -0.04572, 0.74628, ~
## $ V16 <dbl> -0.38223, -0.97515, 0.00299, 0.14516, -0.27457, -0.15540, -0.61436~
## $ V17 <dbl> 0.84356, 0.05499, 0.83775, 0.54094, 0.52940, -0.00343, 0.57945, 1.~
## $ V18 <dbl> -0.38542, -0.62237, -0.13644, -0.39330, -0.21780, -0.10196, -0.680~
## $ V19 <dbl> 0.58212, 0.33109, 0.75535, -1.00000, 0.45107, -0.11575, 0.37852, -~
## $ V20 <dbl> -0.32192, -1.00000, -0.08540, -0.54467, -0.17813, -0.05414, -0.736~
## $ V21 <dbl> 0.56971, -0.13151, 0.70887, -0.69975, 0.05982, 0.01838, 0.36324, 0~
## $ V22 <dbl> -0.29674, -0.45300, -0.27502, 1.00000, -0.35575, 0.03669, -0.76562~
## $ V23 <dbl> 0.36946, -0.18056, 0.43385, 0.00000, 0.02309, 0.01519, 0.31898, 0.~
## $ V24 <dbl> -0.47357, -0.35734, -0.12062, 0.00000, -0.52879, 0.00888, -0.79753~
## $ V25 <dbl> 0.56811, -0.20332, 0.57528, 1.00000, 0.03286, 0.03513, 0.22792, 1.~
## $ V26 <dbl> -0.51171, -0.26569, -0.40220, 0.90695, -0.65158, -0.01535, -0.8163~
## $ V27 <dbl> 0.41078, -0.20468, 0.58984, 0.51613, 0.13290, -0.03240, 0.13659, 1~
## $ V28 <dbl> -0.46168, -0.18401, -0.22145, 1.00000, -0.53206, 0.09223, -0.82510~
## $ V29 <dbl> 0.21266, -0.19040, 0.43100, 1.00000, 0.02431, -0.07859, 0.04606, 0~
## $ V30 <dbl> -0.34090, -0.11593, -0.17365, -0.20099, -0.62197, 0.00732, -0.8239~
## $ V31 <dbl> 0.42267, -0.16626, 0.60436, 0.25682, -0.05707, 0.00000, -0.04262, ~
## $ V32 <dbl> -0.54487, -0.06288, -0.24180, 1.00000, -0.59573, 0.00000, -0.81318~
## $ V33 <dbl> 0.18641, -0.13738, 0.56045, -0.32382, -0.04608, -0.00039, -0.13832~
## $ V34 <dbl> -0.45300, -0.02447, -0.38238, 1.00000, -0.65697, 0.12011, -0.80975~
## $ V35 <dbl> 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, ~
```

```r
# Veri setinin özet istatistiklerini inceleme
summary(ionosphere)
```

```
##        V1               V2          V3                 V4          
##  Min.   :0.0000   Min.   :0   Min.   :-1.0000   Min.   :-1.00000  
##  1st Qu.:1.0000   1st Qu.:0   1st Qu.: 0.4721   1st Qu.:-0.06474  
##  Median :1.0000   Median :0   Median : 0.8711   Median : 0.01631  
##  Mean   :0.8917   Mean   :0   Mean   : 0.6413   Mean   : 0.04437  
##  3rd Qu.:1.0000   3rd Qu.:0   3rd Qu.: 1.0000   3rd Qu.: 0.19418  
##  Max.   :1.0000   Max.   :0   Max.   : 1.0000   Max.   : 1.00000  
##        V5                V6                V7                V8          
##  Min.   :-1.0000   Min.   :-1.0000   Min.   :-1.0000   Min.   :-1.00000  
##  1st Qu.: 0.4127   1st Qu.:-0.0248   1st Qu.: 0.2113   1st Qu.:-0.05484  
##  Median : 0.8092   Median : 0.0228   Median : 0.7287   Median : 0.01471  
##  Mean   : 0.6011   Mean   : 0.1159   Mean   : 0.5501   Mean   : 0.11936  
##  3rd Qu.: 1.0000   3rd Qu.: 0.3347   3rd Qu.: 0.9692   3rd Qu.: 0.44567  
##  Max.   : 1.0000   Max.   : 1.0000   Max.   : 1.0000   Max.   : 1.00000  
##        V9                V10                V11                V12          
##  Min.   :-1.00000   Min.   :-1.00000   Min.   :-1.00000   Min.   :-1.00000  
##  1st Qu.: 0.08711   1st Qu.:-0.04807   1st Qu.: 0.02112   1st Qu.:-0.06527  
##  Median : 0.68421   Median : 0.01829   Median : 0.66798   Median : 0.02825  
##  Mean   : 0.51185   Mean   : 0.18135   Mean   : 0.47618   Mean   : 0.15504  
##  3rd Qu.: 0.95324   3rd Qu.: 0.53419   3rd Qu.: 0.95790   3rd Qu.: 0.48237  
##  Max.   : 1.00000   Max.   : 1.00000   Max.   : 1.00000   Max.   : 1.00000  
##       V13                V14                V15                V16          
##  Min.   :-1.0000   Min.   :-1.00000   Min.   :-1.0000   Min.   :-1.00000
```

```
##  1st Qu.: 0.0000   1st Qu.:-0.07372   1st Qu.: 0.0000   1st Qu.:-0.08170
##  Median : 0.6441   Median : 0.03027   Median : 0.6019   Median : 0.00000
##  Mean   : 0.4008   Mean   : 0.09341   Mean   : 0.3442   Mean   : 0.07113
##  3rd Qu.: 0.9555   3rd Qu.: 0.37486   3rd Qu.: 0.9193   3rd Qu.: 0.30897
##  Max.   : 1.0000   Max.   : 1.00000   Max.   : 1.0000   Max.   : 1.00000
##       V17                V18                  V19                V20
##  Min.   :-1.0000   Min.   :-1.000000   Min.   :-1.0000   Min.   :-1.00000
##  1st Qu.: 0.0000   1st Qu.:-0.225690   1st Qu.: 0.0000   1st Qu.:-0.23467
##  Median : 0.5909   Median : 0.000000   Median : 0.5762   Median : 0.00000
##  Mean   : 0.3819   Mean   :-0.003617   Mean   : 0.3594   Mean   :-0.02402
##  3rd Qu.: 0.9357   3rd Qu.: 0.195285   3rd Qu.: 0.8993   3rd Qu.: 0.13437
##  Max.   : 1.0000   Max.   : 1.000000   Max.   : 1.0000   Max.   : 1.00000
##       V21                V22                  V23                V24
##  Min.   :-1.0000   Min.   :-1.000000   Min.   :-1.0000   Min.   :-1.00000
##  1st Qu.: 0.0000   1st Qu.:-0.243870   1st Qu.: 0.0000   1st Qu.:-0.36689
##  Median : 0.4991   Median : 0.000000   Median : 0.5318   Median : 0.00000
##  Mean   : 0.3367   Mean   : 0.008296   Mean   : 0.3625   Mean   :-0.05741
##  3rd Qu.: 0.8949   3rd Qu.: 0.188760   3rd Qu.: 0.9112   3rd Qu.: 0.16463
##  Max.   : 1.0000   Max.   : 1.000000   Max.   : 1.0000   Max.   : 1.00000
##       V25                V26                V27                V28
##  Min.   :-1.0000   Min.   :-1.00000   Min.   :-1.0000   Min.   :-1.00000
##  1st Qu.: 0.0000   1st Qu.:-0.33239   1st Qu.: 0.2864   1st Qu.:-0.44316
##  Median : 0.5539   Median :-0.01505   Median : 0.7082   Median :-0.01769
##  Mean   : 0.3961   Mean   :-0.07119   Mean   : 0.5416   Mean   :-0.06954
##  3rd Qu.: 0.9052   3rd Qu.: 0.15676   3rd Qu.: 0.9999   3rd Qu.: 0.15354
##  Max.   : 1.0000   Max.   : 1.00000   Max.   : 1.0000   Max.   : 1.00000
##       V29                V30                V31                V32
##  Min.   :-1.0000   Min.   :-1.00000   Min.   :-1.0000   Min.   :-1.000000
##  1st Qu.: 0.0000   1st Qu.:-0.23689   1st Qu.: 0.0000   1st Qu.:-0.242595
##  Median : 0.4966   Median : 0.00000   Median : 0.4428   Median : 0.000000
##  Mean   : 0.3784   Mean   :-0.02791   Mean   : 0.3525   Mean   :-0.003794
##  3rd Qu.: 0.8835   3rd Qu.: 0.15407   3rd Qu.: 0.8576   3rd Qu.: 0.200120
##  Max.   : 1.0000   Max.   : 1.00000   Max.   : 1.0000   Max.   : 1.000000
##       V33                V34                V35
##  Min.   :-1.0000   Min.   :-1.00000   Min.   :0.000
##  1st Qu.: 0.0000   1st Qu.:-0.16535   1st Qu.:0.000
##  Median : 0.4096   Median : 0.00000   Median :1.000
##  Mean   : 0.3494   Mean   : 0.01448   Mean   :0.641
##  3rd Qu.: 0.8138   3rd Qu.: 0.17166   3rd Qu.:1.000
##  Max.   : 1.0000   Max.   : 1.00000   Max.   :1.000
```

```r
# Veri setini eğitim ve test setlerine ayırma
sample <-
  sample.int(
    n = nrow(ionosphere),
    size = floor(.8 * nrow(ionosphere)),
    replace = FALSE
  )
train_data <- ionosphere[sample,]
glimpse(train_data, width = 44)
```

```
## Rows: 280
## Columns: 35
## $ V1  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
```

```
## $ V2  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ V3  <dbl> 1.00000, 0.68729, 0.50000, 0.5~
## $ V4  <dbl> 0.11765, 1.00000, 0.00000, -0.~
## $ V5  <dbl> 1.00000, 0.91973, 0.38696, 0.9~
## $ V6  <dbl> 0.23529, -0.76087, 0.10435, 0.~
## $ V7  <dbl> 1.00000, 0.81773, 0.49130, 0.5~
## $ V8  <dbl> 0.41176, 0.04348, 0.06522, -0.~
## $ V9  <dbl> 1.00000, 0.76087, 0.46957, 0.3~
## $ V10 <dbl> 0.05882, 0.10702, -0.03913, -0~
## $ V11 <dbl> 1.00000, 0.86789, 0.35652, 0.5~
## $ V12 <dbl> 0.23529, 0.73746, -0.12609, -0~
## $ V13 <dbl> 1.00000, 0.70067, 0.45652, 0.6~
## $ V14 <dbl> 0.11765, 0.18227, 0.04783, 0.2~
## $ V15 <dbl> 1.00000, 0.75920, 0.50435, 0.8~
## $ V16 <dbl> 0.47059, 0.13712, 0.02609, -0.~
## $ V17 <dbl> 1.00000, 0.93478, 0.35652, 0.5~
## $ V18 <dbl> -0.05882, -0.25084, 0.19565, 0~
## $ V19 <dbl> 1.00000, 0.70736, 0.42174, 0.5~
## $ V20 <dbl> -0.11765, 0.18729, 0.14783, 0.~
## $ V21 <dbl> 1.00000, 0.64883, 0.42174, 0.5~
## $ V22 <dbl> 0.35294, 0.24582, -0.02609, 0.~
## $ V23 <dbl> 1.00000, 0.60201, 0.32174, 0.8~
## $ V24 <dbl> 0.41176, 0.77425, -0.11304, 0.~
## $ V25 <dbl> 1.00000, 1.00000, 0.47391, 0.4~
## $ V26 <dbl> -0.11765, -0.53846, -0.00870, ~
## $ V27 <dbl> 1.00000, 0.89262, 0.41789, 0.6~
## $ V28 <dbl> 0.20225, 0.22216, 0.06908, -0.~
## $ V29 <dbl> 1.00000, 0.71070, 0.38696, 0.6~
## $ V30 <dbl> 0.05882, 0.53846, 0.03913, -0.~
## $ V31 <dbl> 1.00000, 1.00000, 0.35217, 0.8~
## $ V32 <dbl> 0.35294, -0.06522, 0.14783, -0~
## $ V33 <dbl> 1.00000, 0.56522, 0.44783, 0.7~
## $ V34 <dbl> 0.23529, 0.23913, 0.17391, -0.~
## $ V35 <dbl> 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, ~
```

```r
test_data <- ionosphere[-sample,]
glimpse(test_data, width = 44)
```

```
## Rows: 71
## Columns: 35
## $ V1  <int> 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, ~
## $ V2  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ V3  <dbl> 1.00000, 0.97588, 1.00000, -1.~
## $ V4  <dbl> -0.18829, -0.10602, -0.16316, ~
## $ V5  <dbl> 0.93035, 0.94601, 1.00000, 0.0~
## $ V6  <dbl> -0.36156, -0.20800, -0.10169, ~
## $ V7  <dbl> -0.10868, 0.92806, 0.99999, 0.~
## $ V8  <dbl> -0.93597, -0.28350, -0.15197, ~
## $ V9  <dbl> 1.00000, 0.85996, 1.00000, -1.~
## $ V10 <dbl> -0.04549, -0.27342, -0.19277, ~
## $ V11 <dbl> 0.50874, 0.79766, 0.94055, 1.0~
## $ V12 <dbl> -0.67743, -0.47929, -0.35151, ~
## $ V13 <dbl> 0.34432, 0.78225, 0.95735, 0.0~
## $ V14 <dbl> -0.69707, -0.50764, -0.29785, ~
```

```
## $ V15 <dbl> -0.51685, 0.74628, 0.93719, 0.~
## $ V16 <dbl> -0.97515, -0.61436, -0.34412, ~
## $ V17 <dbl> 0.05499, 0.57945, 0.94486, 1.0~
## $ V18 <dbl> -0.62237, -0.68086, -0.28106, ~
## $ V19 <dbl> 0.33109, 0.37852, 0.90137, -1.~
## $ V20 <dbl> -1.00000, -0.73641, -0.43383, ~
## $ V21 <dbl> -0.13151, 0.36324, 0.86043, 1.~
## $ V22 <dbl> -0.45300, -0.76562, -0.47308, ~
## $ V23 <dbl> -0.18056, 0.31898, 0.82987, 0.~
## $ V24 <dbl> -0.35734, -0.79753, -0.51220, ~
## $ V25 <dbl> -0.20332, 0.22792, 0.84080, -1~
## $ V26 <dbl> -0.26569, -0.81634, -0.47137, ~
## $ V27 <dbl> -0.20468, 0.13659, 0.76224, 1.~
## $ V28 <dbl> -0.18401, -0.82510, -0.58370, ~
## $ V29 <dbl> -0.19040, 0.04606, 0.65723, 1.~
## $ V30 <dbl> -0.11593, -0.82395, -0.68794, ~
## $ V31 <dbl> -0.16626, -0.04262, 0.68714, -~
## $ V32 <dbl> -0.06288, -0.81318, -0.64537, ~
## $ V33 <dbl> -0.13738, -0.13832, 0.64727, 0~
## $ V34 <dbl> -0.02447, -0.80975, -0.67226, ~
## $ V35 <dbl> 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, ~
```

Ionosphere veri kümesi, radyo frekansı yansımasını incelemek ve atmosferdeki iyonosfer tabakasına yönelik radyo dalgalarının yansımasını tahmin etmek amacıyla toplanmıştır. Bu veri kümesi, özellikle sınıflandırma problemleri için tasarlanmıştır. Veri setinde iki sınıf bulunur: "g" (iyi) ve "b" (kötü). Bu sınıflar, iyonosfer tabakasına yansıyan radyo sinyallerinin özelliklerine dayanarak belirlenir.

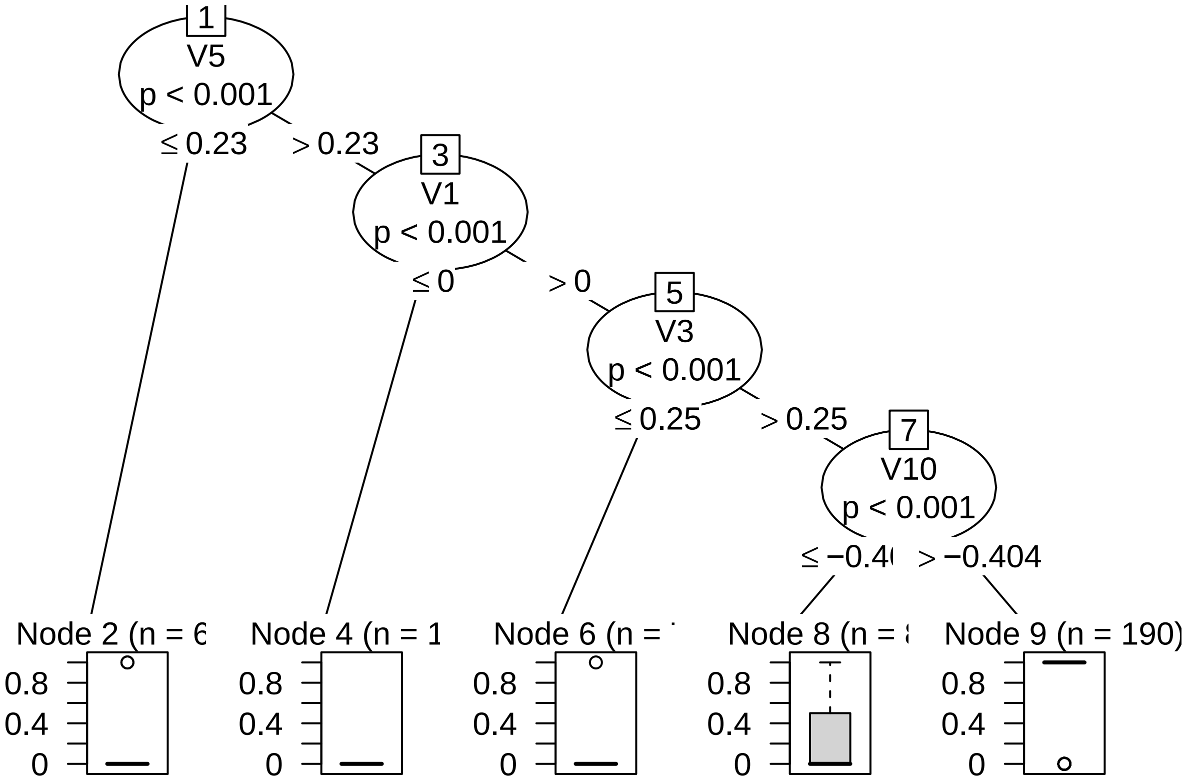
Veri kümesi, 34 sayısal özellik içerir. Bu özellikler, radyo dalgalarının farklı açılarda yansımasını temsil eder. Bu özelliklerin her biri, bir radyo sinyalinin iyonosfer tabakasına ne kadar etkili bir şekilde yansıdığını veya yansımadığını açıklar. Bu özelliklerin değerleri, ölçülen fiziksel özelliklerin sonuçlarıdır ve veri kümesinin amacı, bu özelliklerin değerlerine dayanarak bir radyo sinyalinin "iyi" veya "kötü" olduğunu sınıflandırmaktır.

Bu veri kümesi, sınıflandırma algoritmalarının eğitilmesi ve test edilmesi için kullanılabilir. Özellikle, yeni radyo sinyallerinin iyonosfer tabakasına yansıma potansiyelini değerlendirmek için kullanılabilecek bir modelin geliştirilmesine yardımcı olabilir. Ayrıca, bu veri kümesi, makine öğrenimi uygulamalarında sınıflandırma, özellik seçimi ve model değerlendirmesi için birçok farklı deneyin yapıldığı bir test sahası olarak da kullanılabilir.

# decision treees

```r
# Decision Tree (Karar Ağacı) modelini oluşturma
dt_model <- ctree(V35 ~ ., data = train_data)

# Karar Ağacı modelini görselleştirme
plot(dt_model)
```



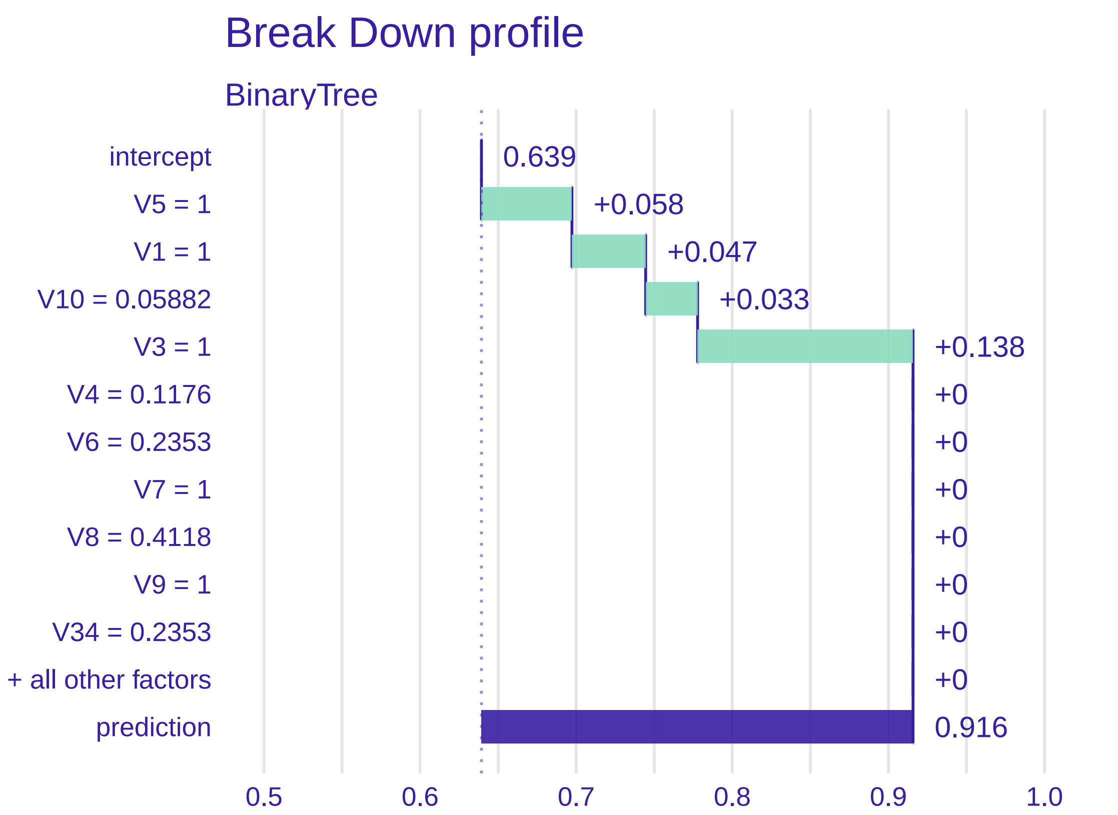

```r
# Decision Tree modelini açıklama nesnesi ile açıklama
dt_explainer <-
  DALEX::explain(dt_model, data = train_data[, -35], y = train_data$V35)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  BinaryTree  (  default  )
##   -> data              :  280  rows  34  cols
##   -> target variable   :  280  values
##   -> predict function  :  yhat.default will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package Model of class: BinaryTree package unrecognized , ver. Unknown , ta
##   -> predicted values  :  numerical, min =  0 , mean =  0.6392857 , max =  0.9157895
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -0.9157895 , mean =  2.750035e-17 , max =  0.9666667
##   A new explainer has been created!
```

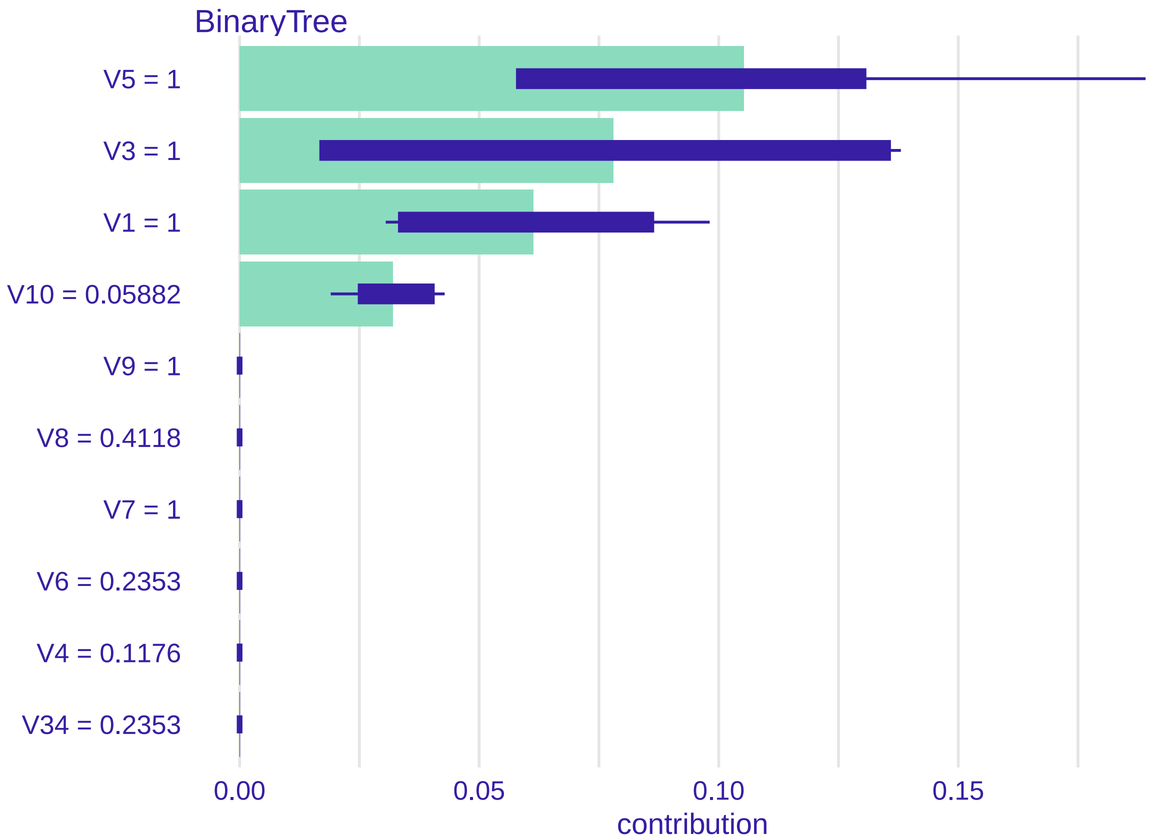
```r
# İlk gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[1, ]
```

```
##     V1 V2 V3      V4 V5      V6 V7      V8 V9     V10 V11     V12 V13     V14
## 176  1  0  1 0.11765  1 0.23529  1 0.41176  1 0.05882   1 0.23529   1 0.11765
##     V15     V16 V17      V18 V19      V20 V21     V22 V23     V24 V25      V26
## 176   1 0.47059   1 -0.05882   1 -0.11765   1 0.35294   1 0.41176   1 -0.11765
##     V27     V28 V29     V30 V31     V32 V33     V34 V35
## 176   1 0.20225   1 0.05882   1 0.35294   1 0.23529   1
```

```
bd_dt_1 <-
  predict_parts(dt_explainer,
                new_observation = train_data[1, -35], type = "break_down")
shap_dt_1 <-
  predict_parts(dt_explainer,
                new_observation = train_data[1, -35], type = "shap")
plot(bd_dt_1)
```



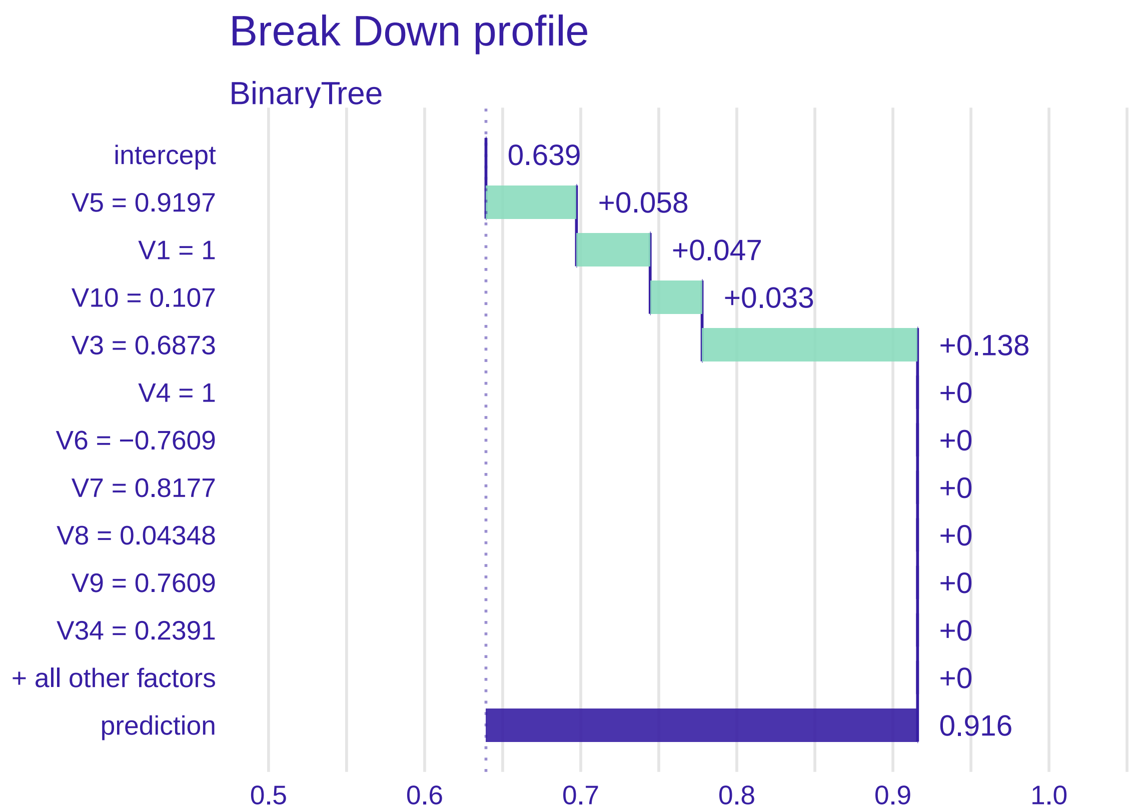

```
plot(shap_dt_1)
```



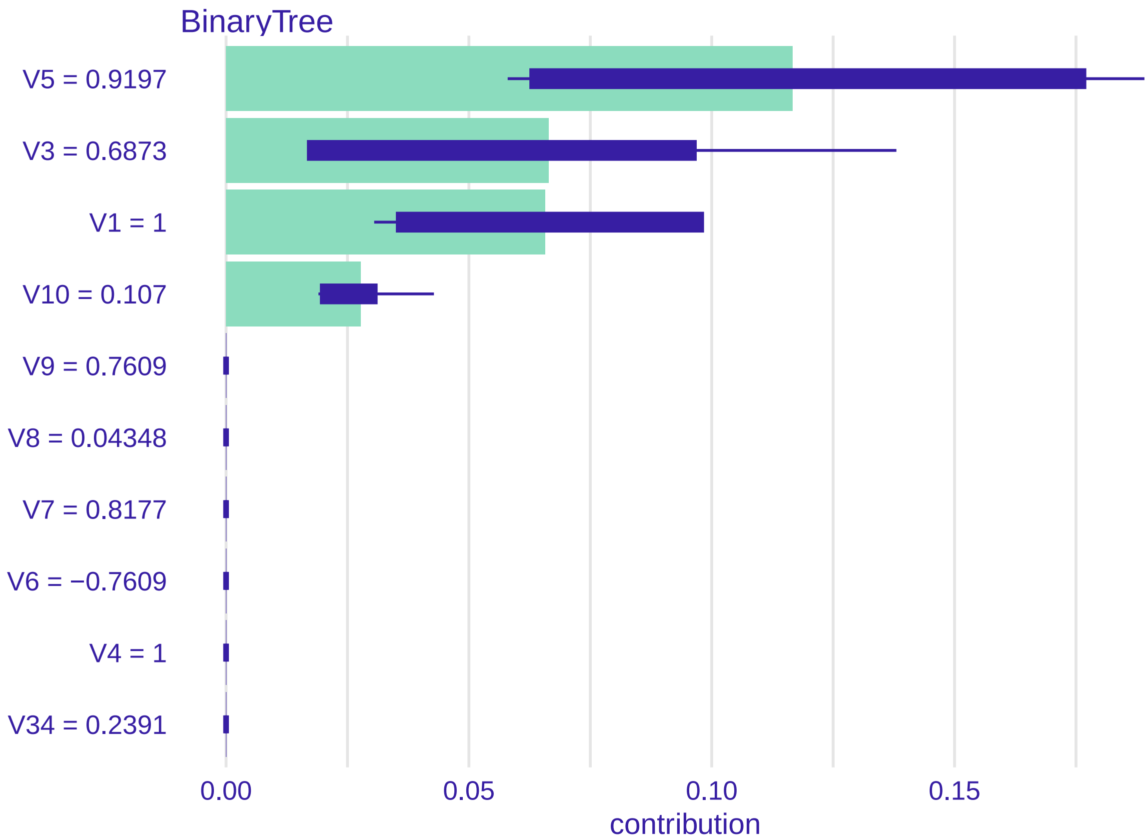

```
# İkinci gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[2, ]
```

```
##      V1 V2      V3 V4      V5       V6      V7      V8      V9     V10     V11
## 235  1  0 0.68729  1 0.91973 -0.76087 0.81773 0.04348 0.76087 0.10702 0.86789
##         V12     V13     V14    V15     V16     V17      V18     V19     V20
## 235 0.73746 0.70067 0.18227 0.7592 0.13712 0.93478 -0.25084 0.70736 0.18729
##         V21     V22     V23     V24 V25      V26     V27     V28    V29     V30
## 235 0.64883 0.24582 0.60201 0.77425   1 -0.53846 0.89262 0.22216 0.7107 0.53846
##     V31      V32     V33     V34 V35
## 235   1 -0.06522 0.56522 0.23913   0
```

```
bd_dt_2 <-
  predict_parts(dt_explainer,
              new_observation = train_data[2, -35], type = "break_down")
shap_dt_2 <-
  predict_parts(dt_explainer,
              new_observation = train_data[2, -35], type = "shap")
plot(bd_dt_2)
```



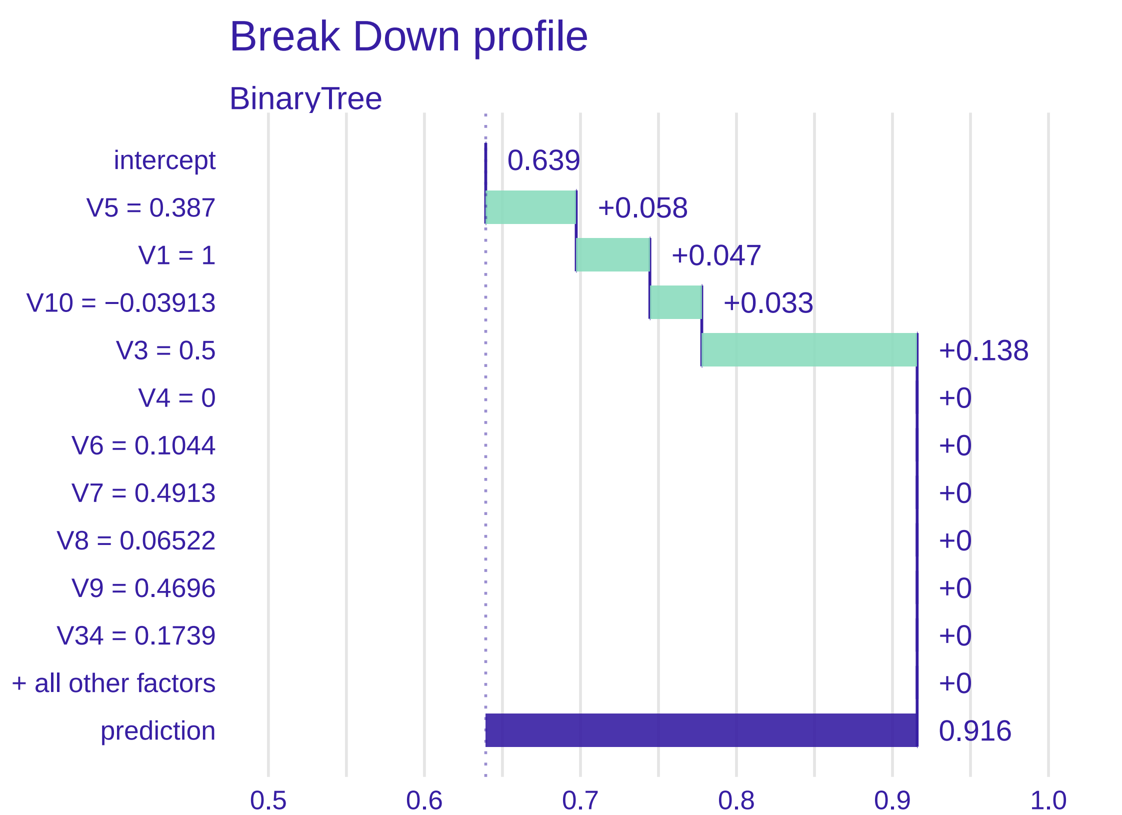

```
plot(shap_dt_2)
```



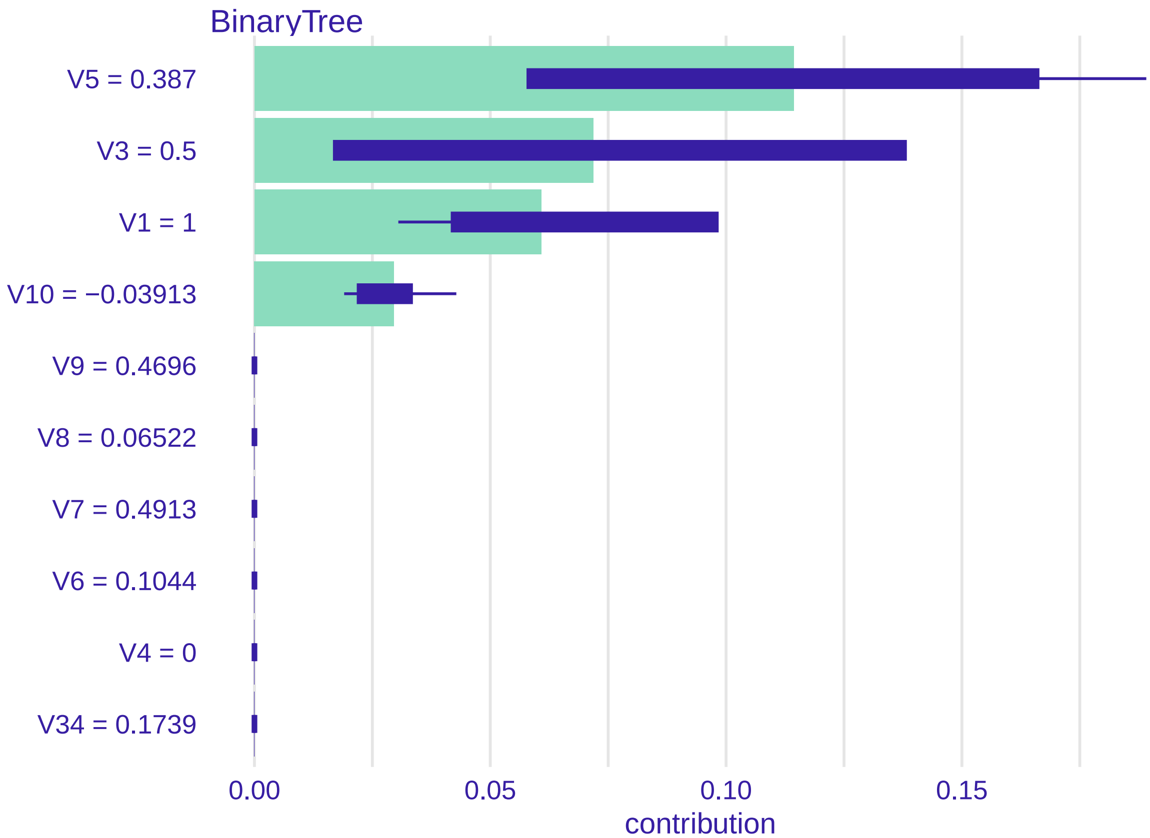

```
# Üçüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[3, ]
```

```
##      V1  V2  V3 V4       V5       V6      V7       V8       V9      V10      V11
## 261   1   0 0.5  0 0.38696 0.10435 0.4913 0.06522 0.46957 -0.03913 0.35652
##           V12      V13      V14      V15      V16      V17      V18      V19      V20
## 261 -0.12609 0.45652 0.04783 0.50435 0.02609 0.35652 0.19565 0.42174 0.14783
##          V21       V22      V23       V24      V25     V26      V27      V28      V29
## 261 0.42174 -0.02609 0.32174 -0.11304 0.47391 -0.0087 0.41789 0.06908 0.38696
##          V30      V31      V32      V33      V34 V35
## 261 0.03913 0.35217 0.14783 0.44783 0.17391   1
```

```
bd_dt_3 <-
  predict_parts(dt_explainer,
                new_observation = train_data[3, -35], type = "break_down")
shap_dt_3 <-
  predict_parts(dt_explainer,
                new_observation = train_data[3, -35], type = "shap")
plot(bd_dt_3)
```



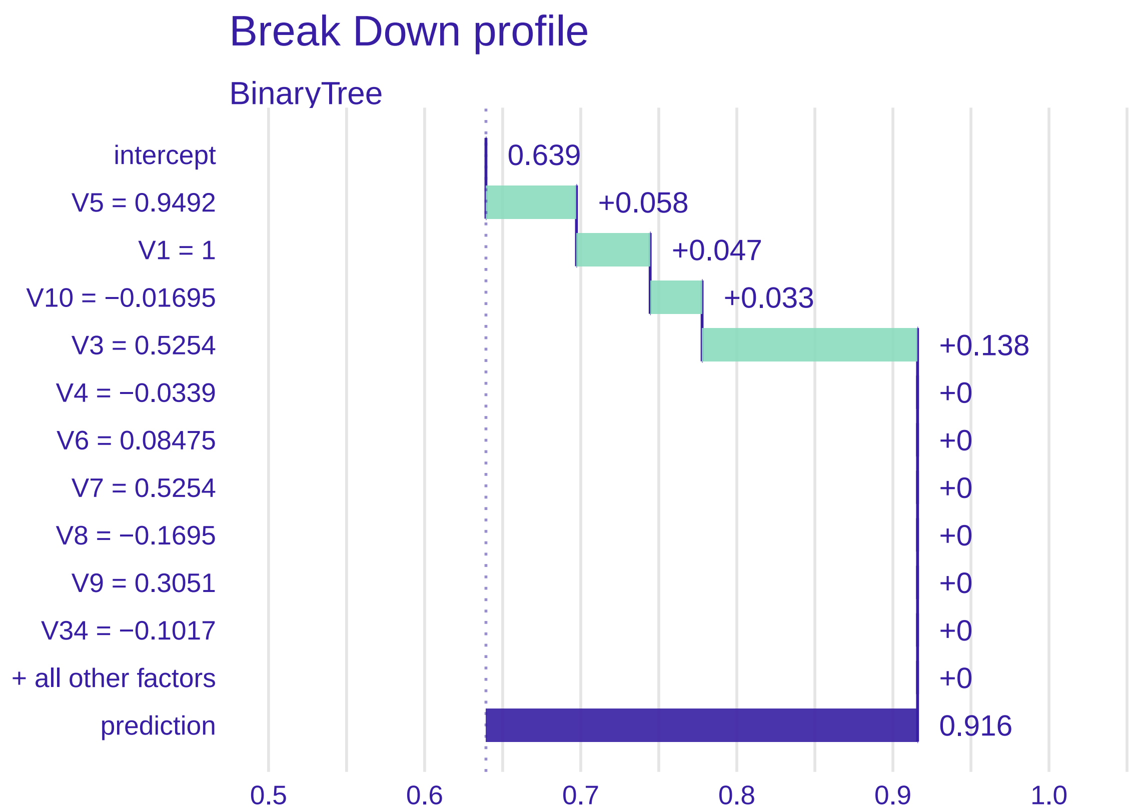

```
plot(shap_dt_3)
```



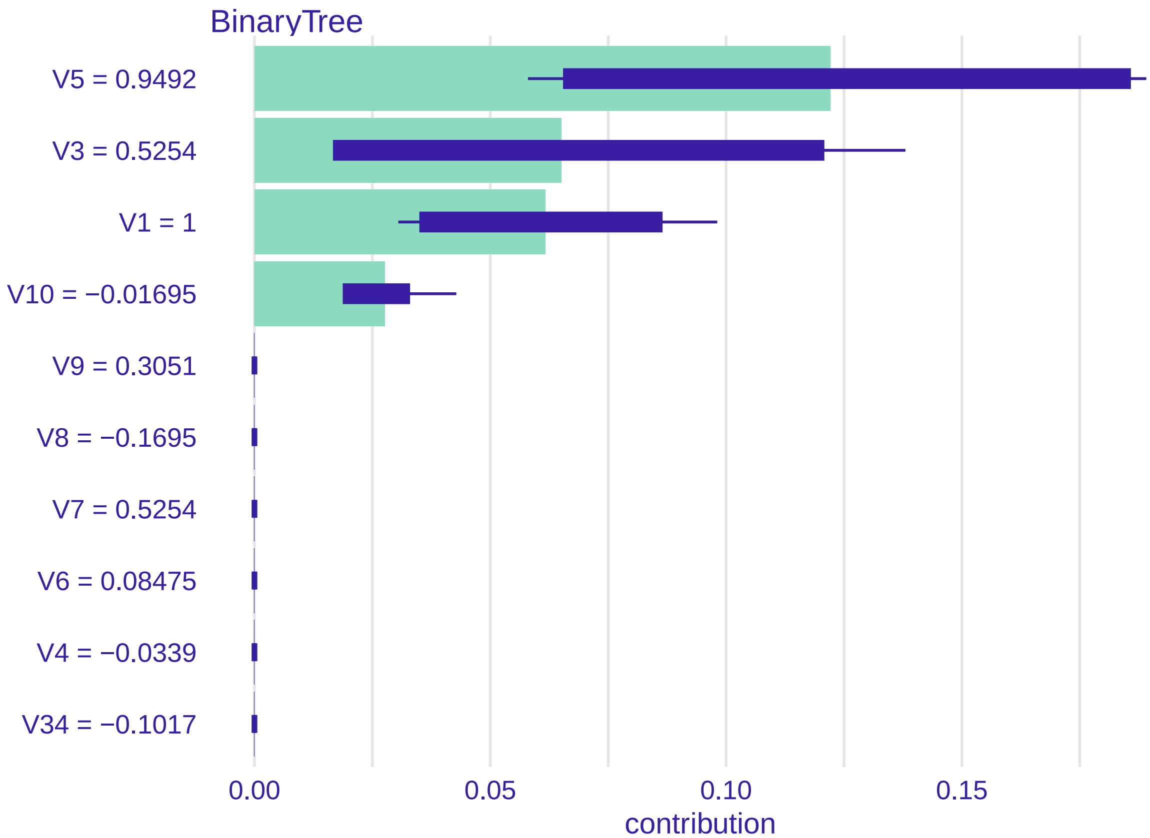

```
# Dördüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[4, ]
```

```
##      V1 V2       V3       V4      V5      V6      V7       V8      V9      V10
## 306   1  0 0.52542 -0.0339 0.94915 0.08475 0.52542 -0.16949 0.30508 -0.01695
##         V11      V12     V13     V14     V15      V16     V17     V18     V19
## 306 0.50847 -0.13559 0.64407 0.28814 0.83051 -0.35593 0.54237 0.01695 0.55932
##        V20     V21     V22     V23     V24     V25     V26     V27      V28
## 306 0.0339 0.59322 0.30508 0.86441 0.05085 0.40678 0.15254 0.67287 -0.00266
##         V29     V30     V31      V32     V33      V34 V35
## 306 0.66102 -0.0339 0.83051 -0.15254 0.76271 -0.10169   1
```

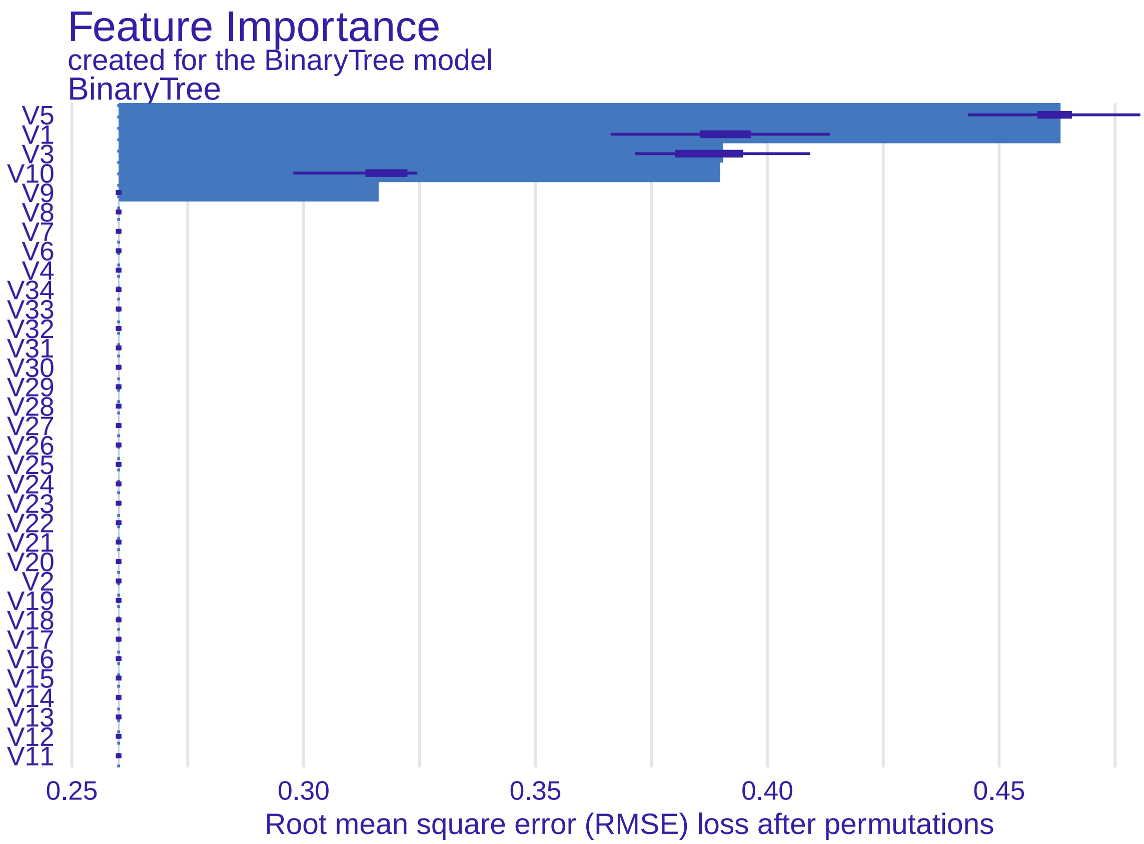
```
bd_dt_4 <-
  predict_parts(dt_explainer,
                new_observation = train_data[4, -35], type = "break_down")
shap_dt_4 <-
  predict_parts(dt_explainer,
                new_observation = train_data[4, -35], type = "shap")
plot(bd_dt_4)
```




```
plot(shap_dt_4)
```




```
# Decision Tree modelinin değişken önem sıralamasını hesaplama ve grafikleme
dt_var_imp <- DALEX::variable_importance(dt_explainer)
plot(dt_var_imp)
```




## Random Forest

```
# Veri setini matrise dönüştürme
train_x <- data.matrix(train_data[,-35])
train_y <- data.matrix(train_data[, 35])
test_x <- data.matrix(test_data[,-35])
test_y <- data.matrix(test_data[, 35])

# Random Forest modelini eğitme
rf_model <- randomForest(x = train_x,
                         y = as.factor(train_y),
                         ntree = 500)

# Random Forest modelini açıklama nesnesi ile açıklama
rf_explainer <-
  DALEX::explain(rf_model, data = train_data[, -35], y = train_data$V35)
```

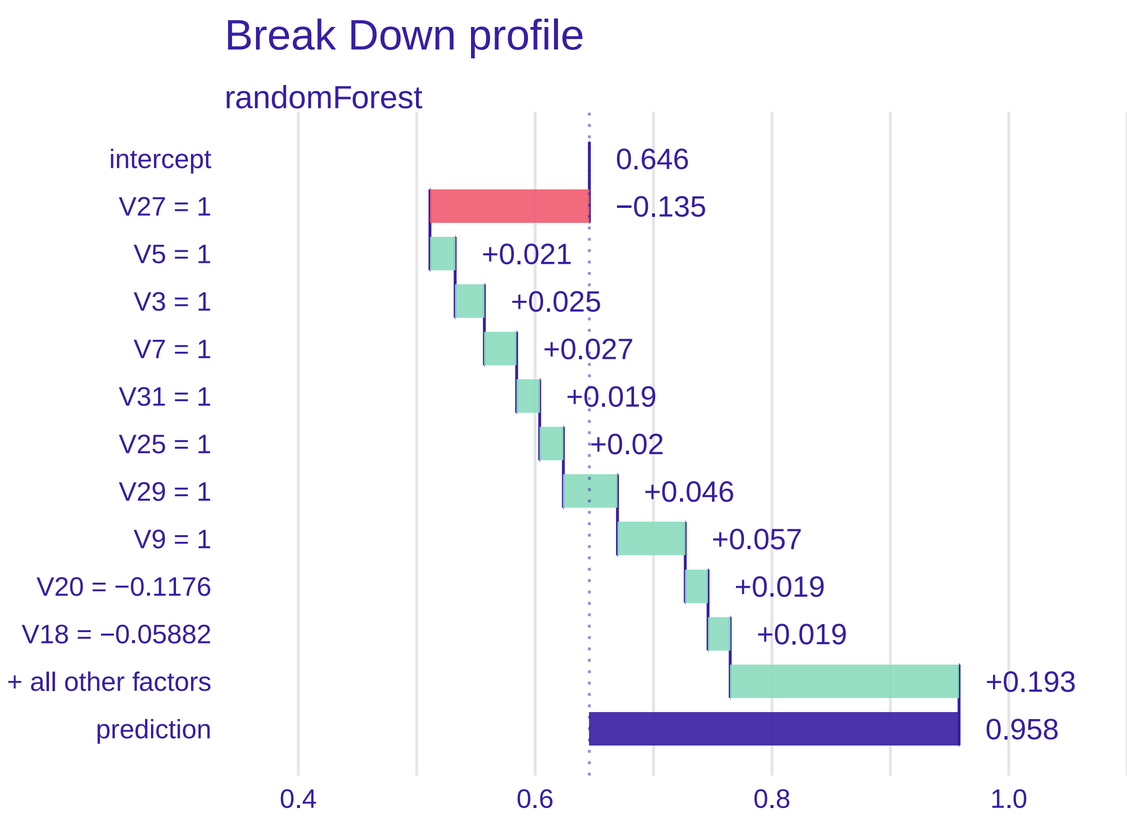
```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  280  rows  34  cols
##   -> target variable   :  280  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
```

```
##   -> model_info        :  package randomForest , ver. 4.7.1.1 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0 , mean =  0.6458143 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -0.36 , mean =  -0.006528571 , max =  0.332
##   A new explainer has been created!
```

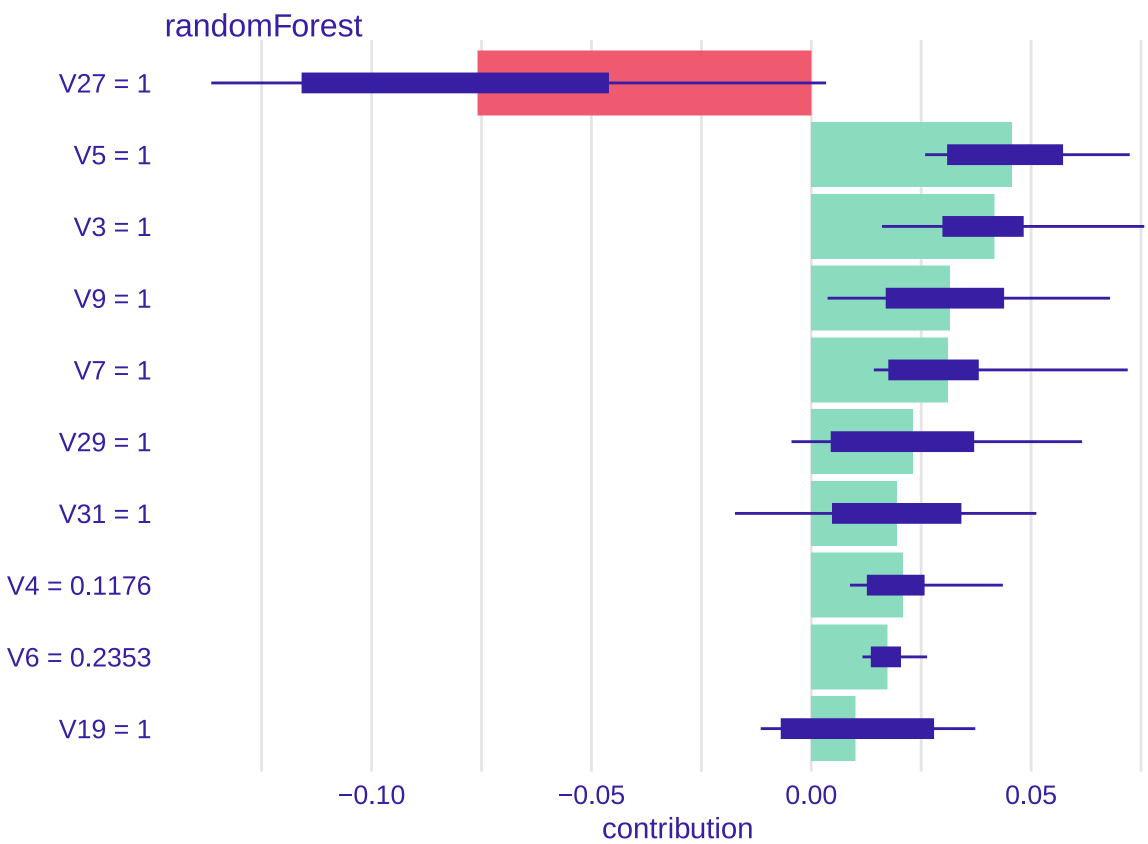
```r
# İlk gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[1, ]
```

```
##      V1 V2 V3      V4 V5      V6 V7      V8 V9     V10 V11     V12 V13     V14
## 176   1  0  1 0.11765  1 0.23529  1 0.41176  1 0.05882   1 0.23529   1 0.11765
##      V15     V16 V17      V18 V19      V20 V21     V22 V23     V24 V25      V26
## 176    1 0.47059   1 -0.05882   1 -0.11765   1 0.35294   1 0.41176   1 -0.11765
##      V27     V28 V29     V30 V31     V32 V33     V34 V35
## 176    1 0.20225   1 0.05882   1 0.35294   1 0.23529   1
```

```r
bd_rf_1 <-
  predict_parts(rf_explainer,
                new_observation = train_data[1, -35], type = "break_down")
shap_rf_1 <-
  predict_parts(rf_explainer,
                new_observation = train_data[1, -35], type = "shap")
plot(bd_rf_1)
```



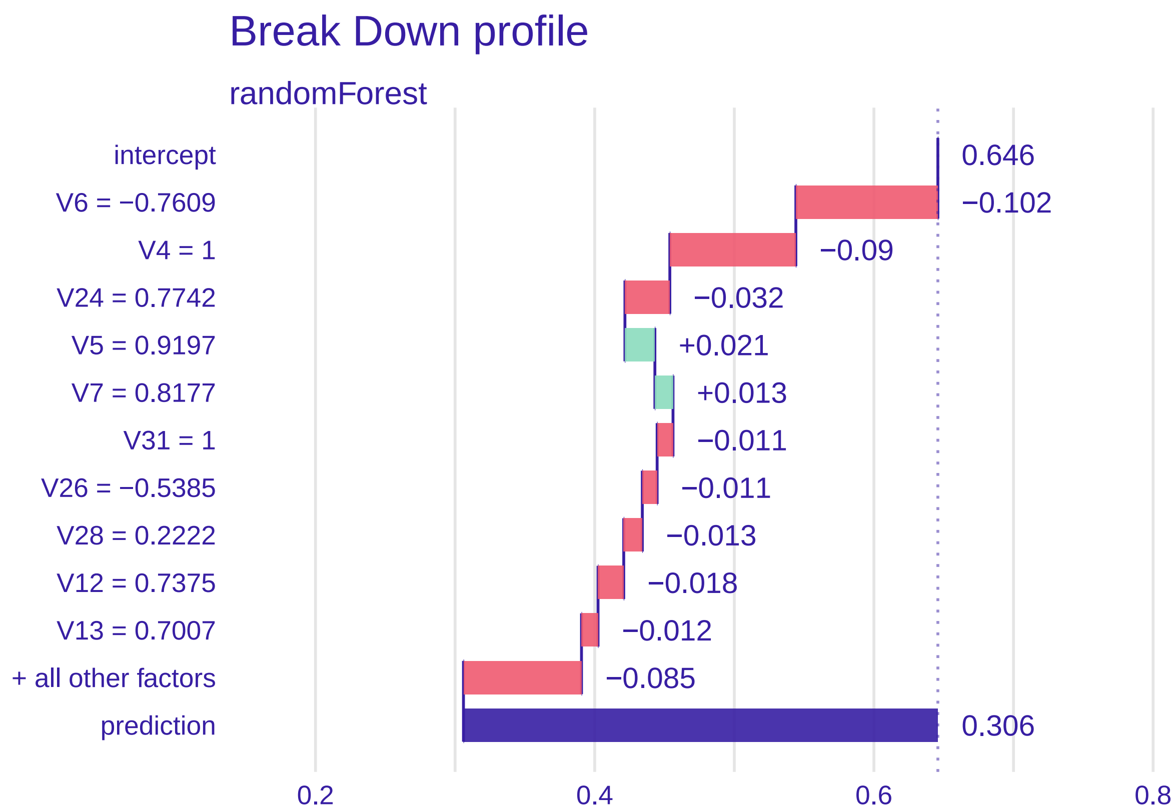
```
plot(shap_rf_1)
```



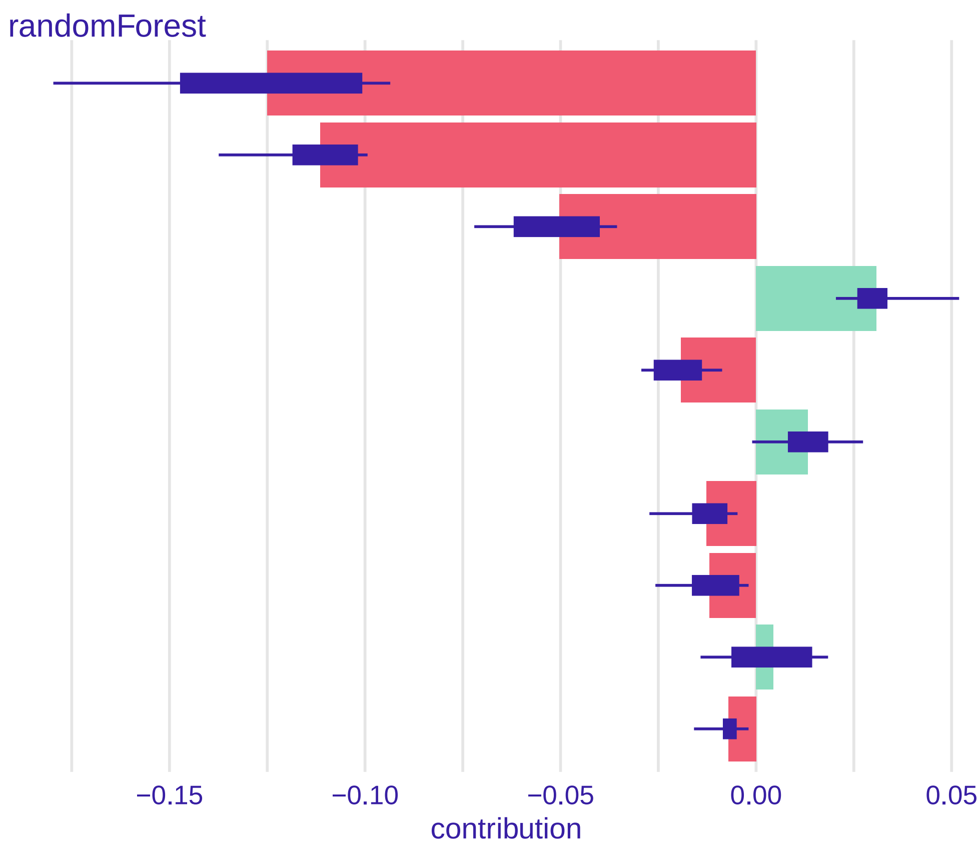

```
# İkinci gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[2, ]
```

```
##      V1 V2      V3 V4      V5       V6      V7      V8      V9     V10     V11
## 235   1  0 0.68729  1 0.91973 -0.76087 0.81773 0.04348 0.76087 0.10702 0.86789
##         V12     V13     V14    V15     V16     V17      V18     V19     V20
## 235 0.73746 0.70067 0.18227 0.7592 0.13712 0.93478 -0.25084 0.70736 0.18729
##         V21     V22     V23     V24 V25      V26     V27     V28    V29     V30
## 235 0.64883 0.24582 0.60201 0.77425   1 -0.53846 0.89262 0.22216 0.7107 0.53846
##      V31      V32     V33     V34 V35
## 235   1 -0.06522 0.56522 0.23913   0
```

```
bd_rf_2 <-
  predict_parts(rf_explainer,
                new_observation = train_data[2, -35], type = "break_down")
shap_rf_2 <-
  predict_parts(rf_explainer,
                new_observation = train_data[2, -35], type = "shap")
plot(bd_rf_2)
```



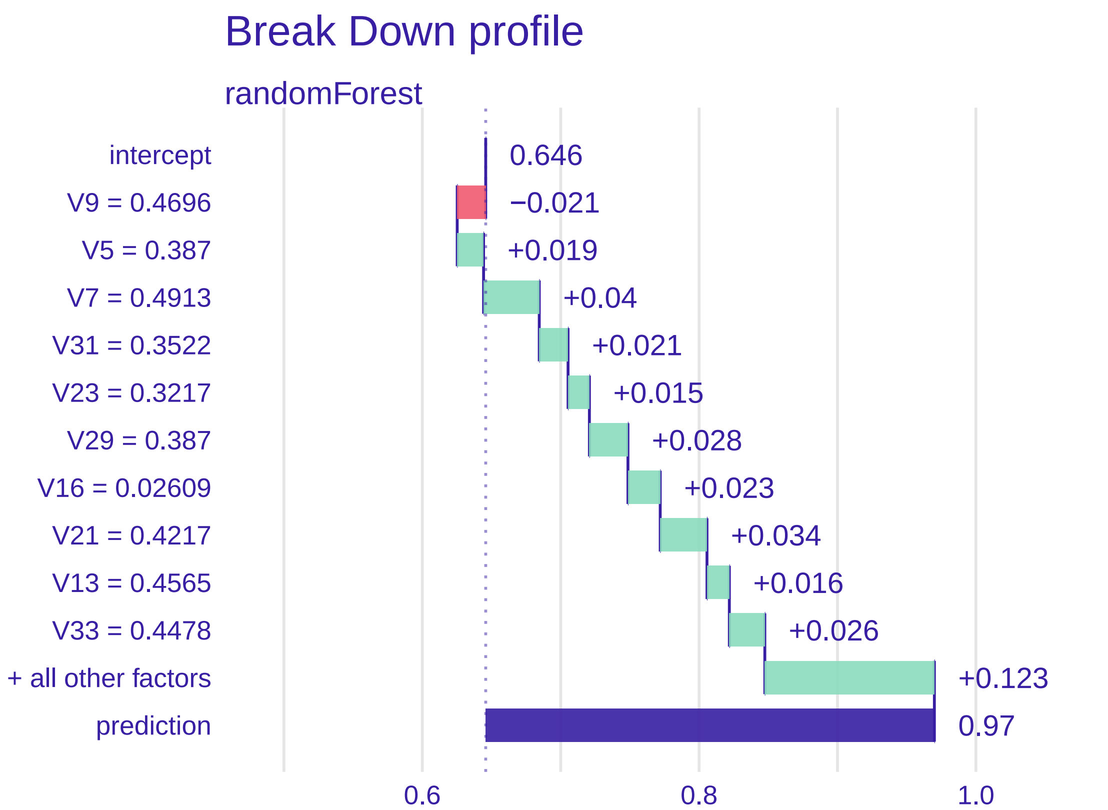

```
plot(shap_rf_2)
```



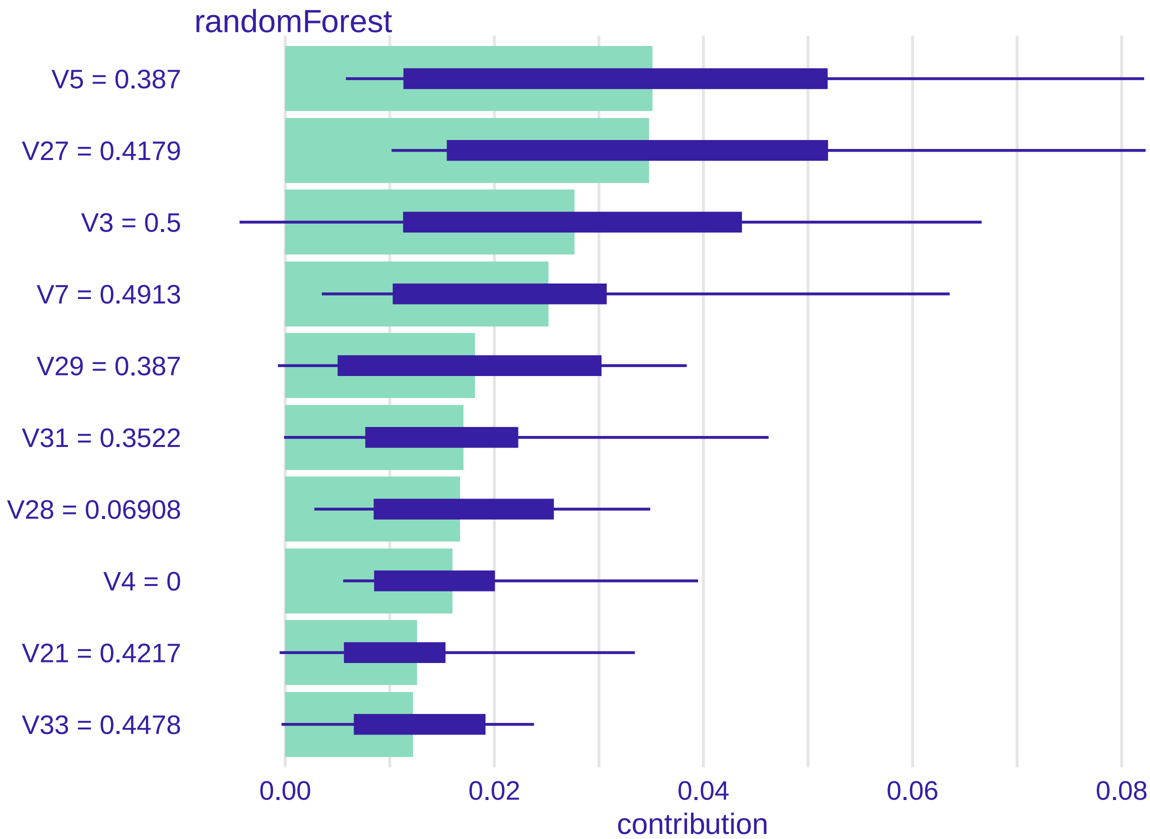

```r
# Üçüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[3, ]
```

```
##     V1 V2  V3 V4      V5      V6     V7      V8      V9      V10     V11
## 261  1  0 0.5  0 0.38696 0.10435 0.4913 0.06522 0.46957 -0.03913 0.35652
##          V12     V13     V14     V15     V16     V17     V18     V19     V20
## 261 -0.12609 0.45652 0.04783 0.50435 0.02609 0.35652 0.19565 0.42174 0.14783
##         V21      V22     V23      V24     V25     V26     V27     V28     V29
## 261 0.42174 -0.02609 0.32174 -0.11304 0.47391 -0.0087 0.41789 0.06908 0.38696
##         V30     V31     V32     V33     V34 V35
## 261 0.03913 0.35217 0.14783 0.44783 0.17391   1
```

```r
bd_rf_3 <-
  predict_parts(rf_explainer,
                new_observation = train_data[3, -35], type = "break_down")
shap_rf_3 <-
  predict_parts(rf_explainer,
                new_observation = train_data[3, -35], type = "shap")
plot(bd_rf_3)
```



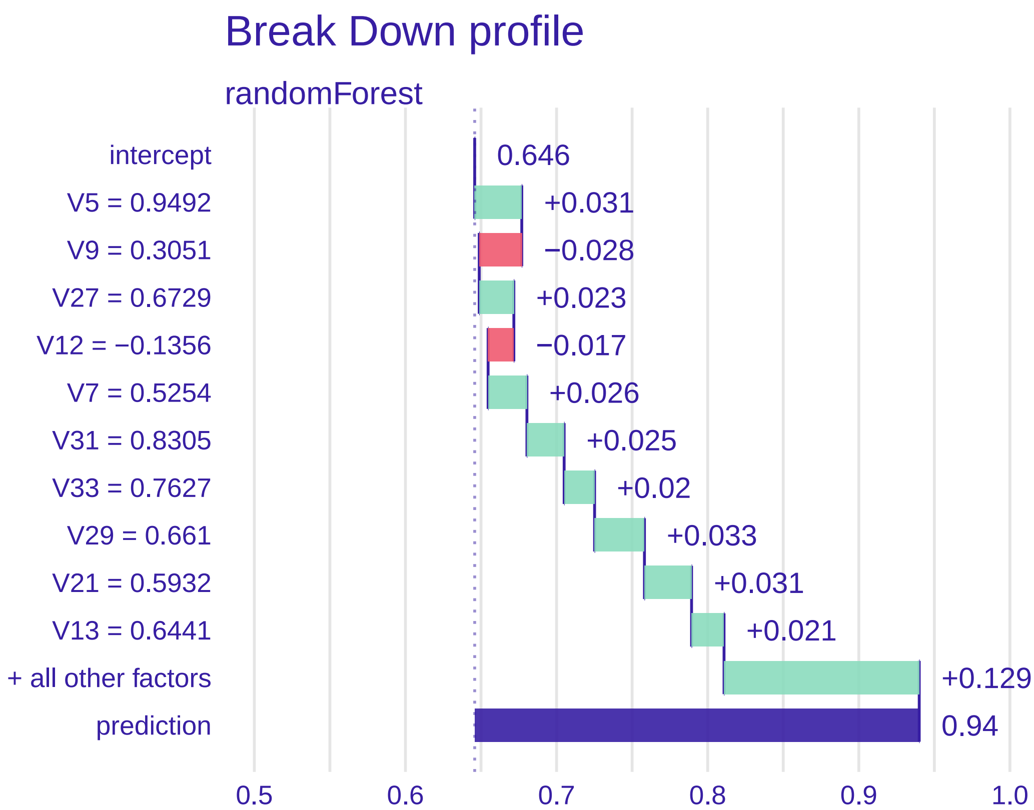

```
plot(shap_rf_3)
```



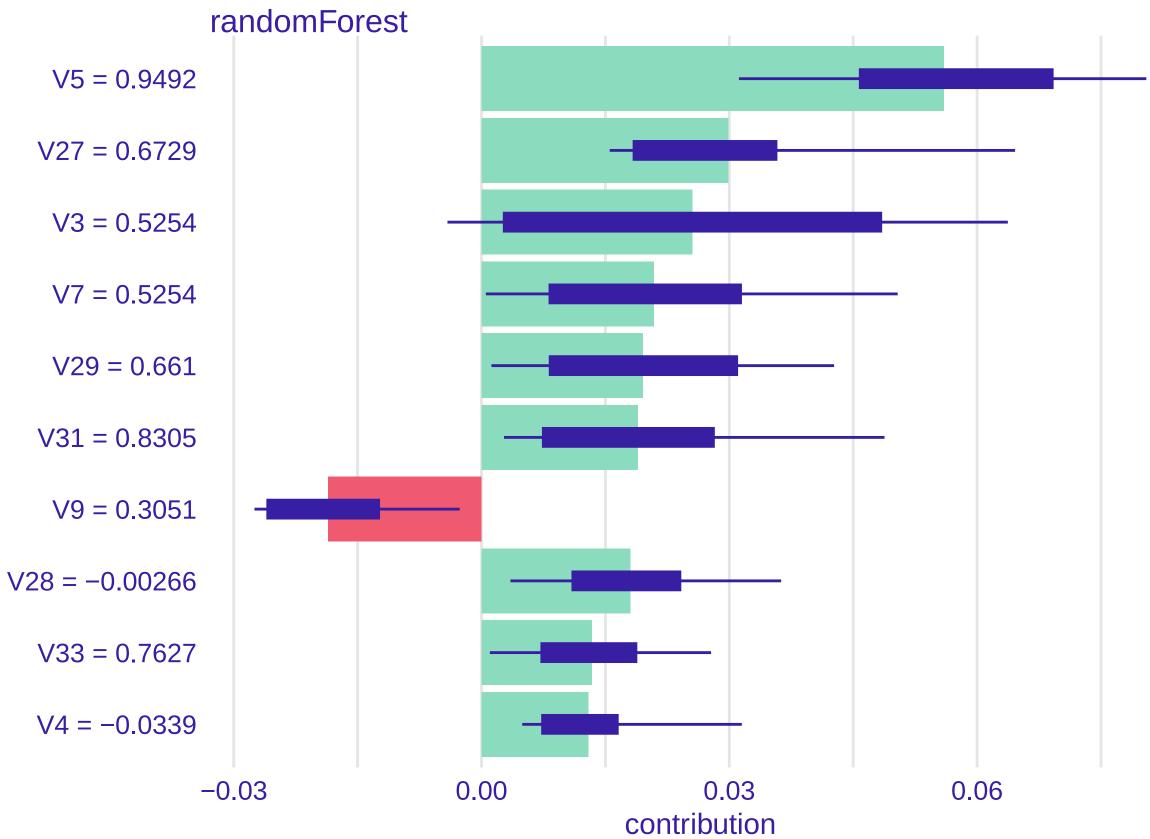

```r
# Dördüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[4, ]
```

```
##      V1 V2       V3      V4      V5      V6      V7       V8      V9      V10
## 306   1  0 0.52542 -0.0339 0.94915 0.08475 0.52542 -0.16949 0.30508 -0.01695
##          V11      V12     V13     V14     V15      V16     V17     V18     V19
## 306 0.50847 -0.13559 0.64407 0.28814 0.83051 -0.35593 0.54237 0.01695 0.55932
##         V20     V21     V22     V23     V24     V25     V26     V27      V28
## 306 0.0339 0.59322 0.30508 0.86441 0.05085 0.40678 0.15254 0.67287 -0.00266
##          V29      V30     V31      V32     V33      V34 V35
## 306 0.66102 -0.0339 0.83051 -0.15254 0.76271 -0.10169   1
```

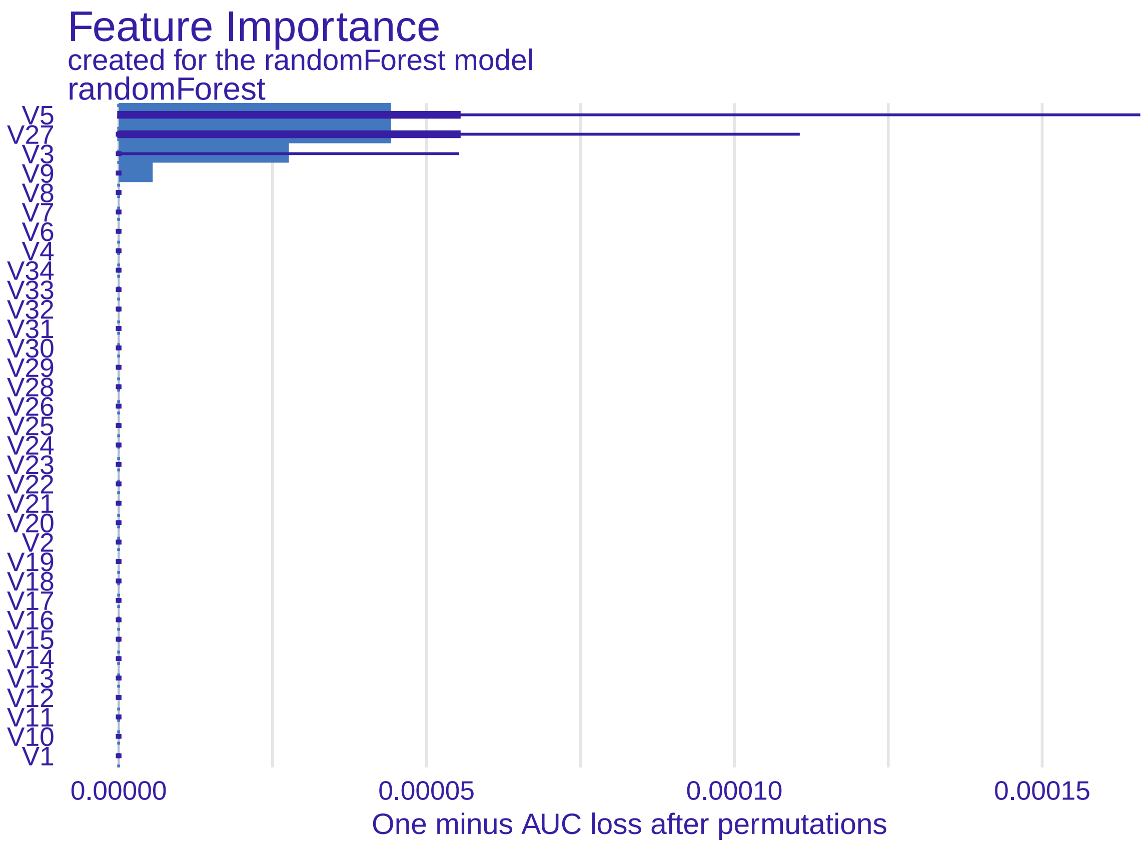
```r
bd_rf_4 <-
  predict_parts(rf_explainer,
              new_observation = train_data[4, -35], type = "break_down")
shap_rf_4 <-
  predict_parts(rf_explainer,
              new_observation = train_data[4, -35], type = "shap")
plot(bd_rf_4)
```




```
plot(shap_rf_4)
```




```
# Random Forest modelinin değişken önem sıralamasını hesaplama ve grafikleme
rf_var_imp <- DALEX::variable_importance(rf_explainer)
plot(rf_var_imp)
```




## SVM

```
# SVM modelini oluşturma
svm_model <-
  svm(
    V35 ~ .,
    data = train_data,
    type = 'C-classification',
    scale = FALSE,
    probability = TRUE
  )

# SVM modelini açıklama nesnesi ile açıklama
svm_explainer <-
  DALEX::explain(svm_model, data = train_data[, -35], y = train_data$V35)
```

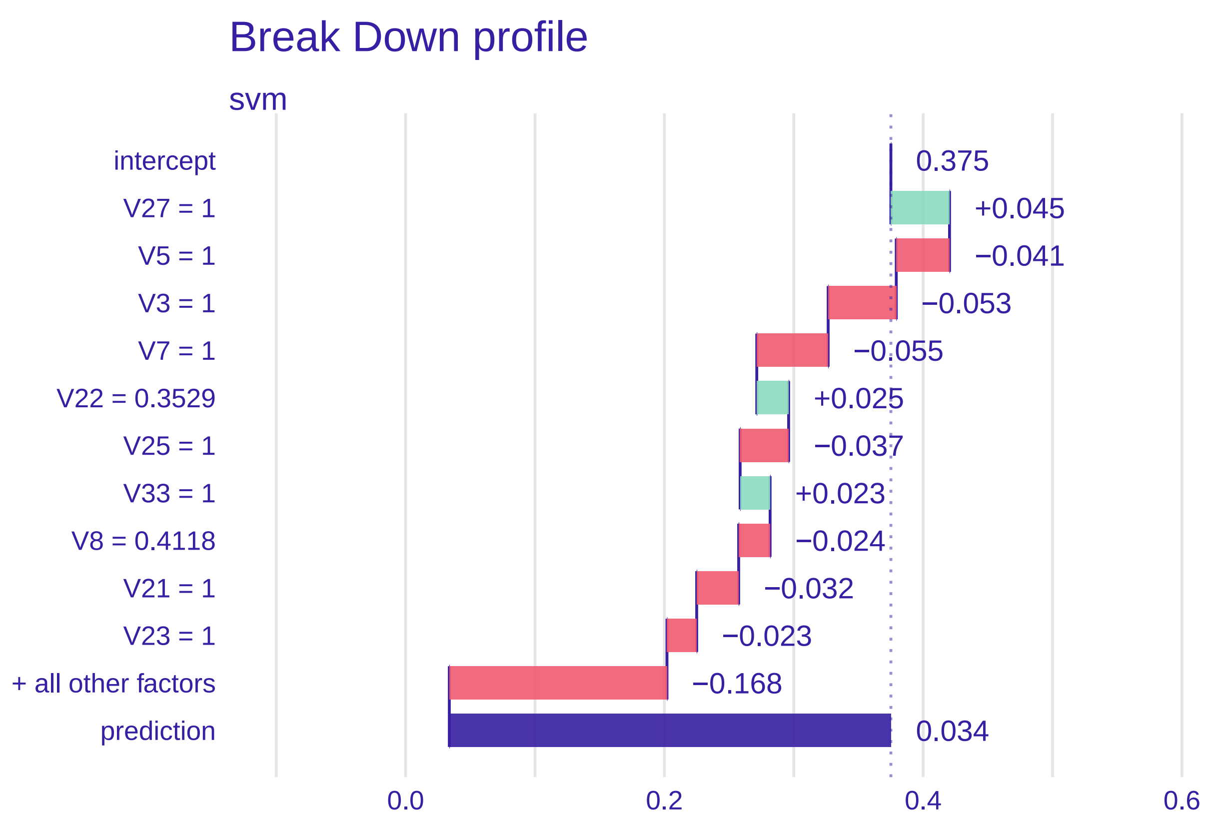
```
## Preparation of a new explainer is initiated
##   -> model label       :  svm  (  default  )
##   -> data              :  280  rows  34  cols
##   -> target variable   :  280  values
##   -> predict function  :  yhat.svm  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package e1071 , ver. 1.7.11 , task classification (  default  )
```

```
##    -> predicted values  :  numerical, min =  0.005215066 , mean =  0.3750405 , max =  0.9999992
##    -> residual function :  difference between y and yhat (  default  )
##    -> residuals         :  numerical, min =  -0.9999992 , mean =  0.2642452 , max =  0.9947849
##    A new explainer has been created!
```

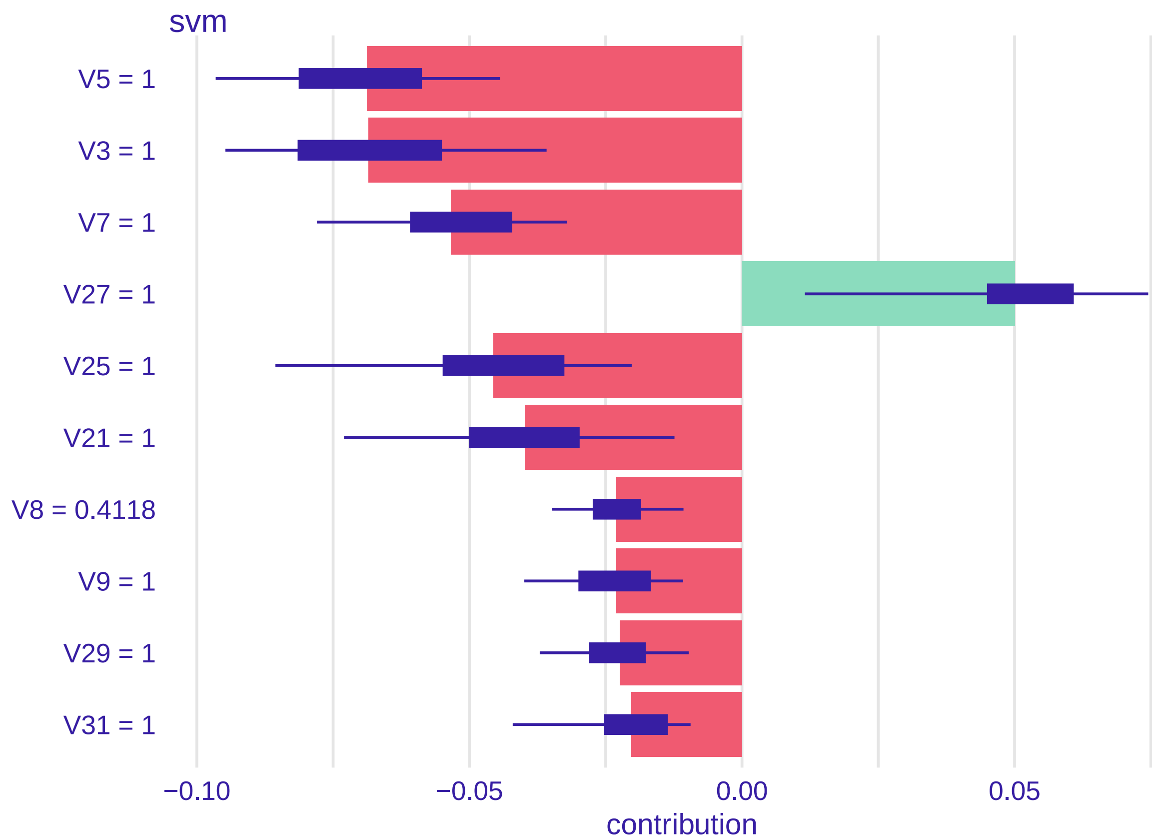
```r
# İlk gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[1, ]
```

```
##      V1  V2  V3      V4  V5      V6  V7      V8  V9     V10  V11     V12  V13     V14
## 176   1   0   1 0.11765   1 0.23529   1 0.41176   1 0.05882    1 0.23529    1 0.11765
##      V15     V16  V17      V18  V19      V20  V21     V22  V23     V24  V25      V26
## 176    1 0.47059    1 -0.05882    1 -0.11765    1 0.35294    1 0.41176    1 -0.11765
##      V27     V28  V29     V30  V31     V32  V33     V34  V35
## 176    1 0.20225    1 0.05882    1 0.35294    1 0.23529    1
```

```r
bd_svm_1 <-
  predict_parts(svm_explainer,
                new_observation = train_data[1, -35], type = "break_down")
shap_svm_1 <-
  predict_parts(svm_explainer,
                new_observation = train_data[1, -35], type = "shap")
plot(bd_svm_1)
```



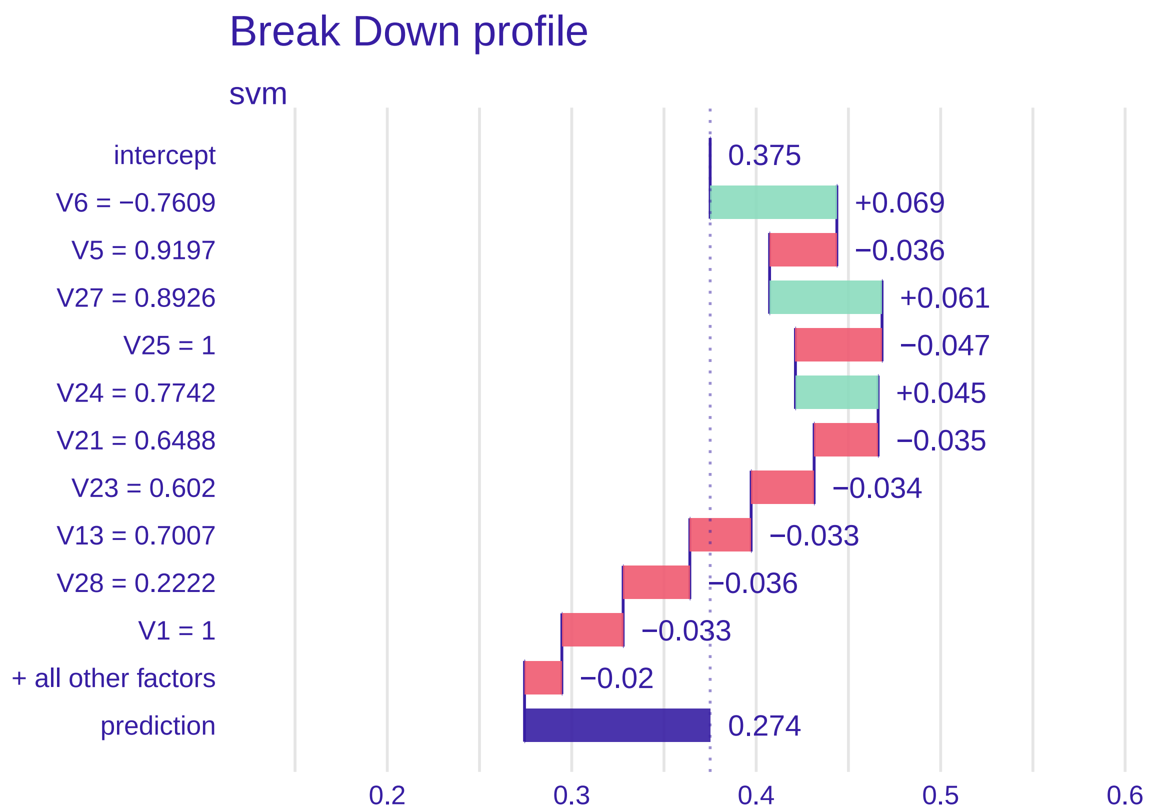
```r
plot(shap_svm_1)
```



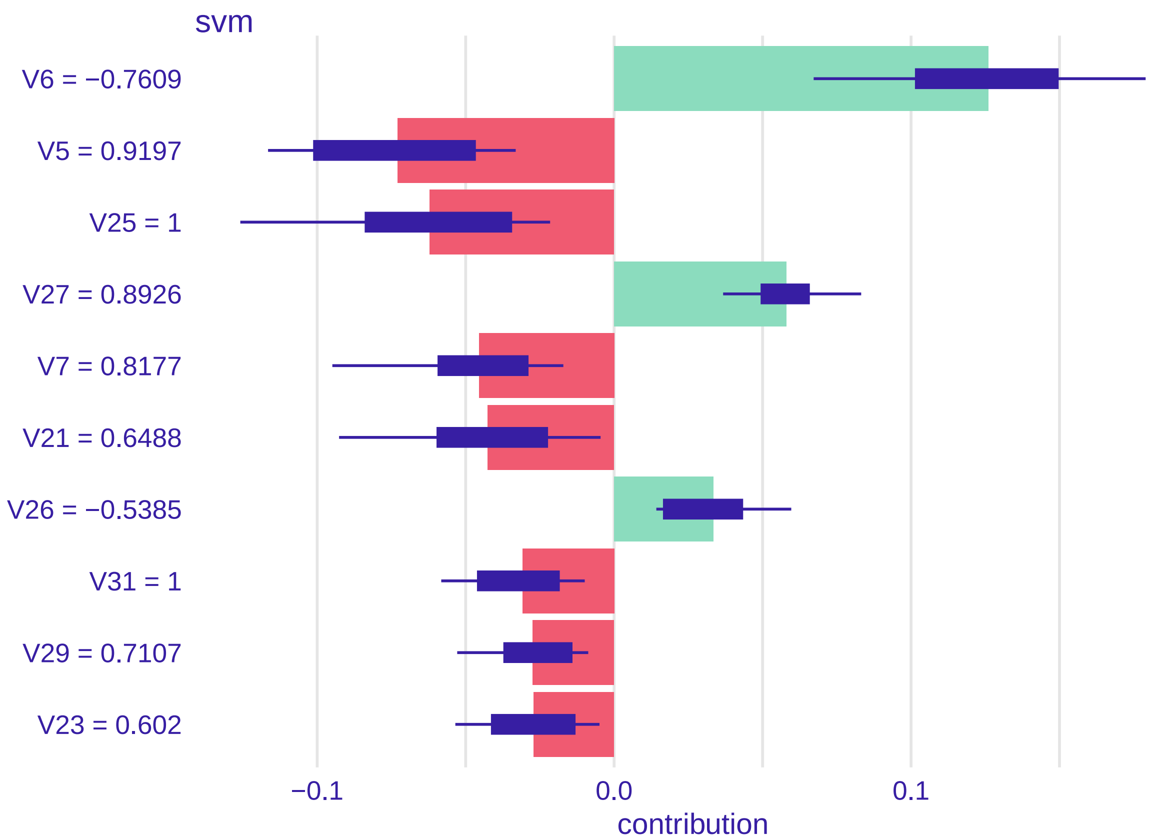

```r
# İkinci gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[2, ]
```

```
##      V1 V2      V3 V4      V5       V6      V7      V8      V9     V10     V11
## 235   1  0 0.68729  1 0.91973 -0.76087 0.81773 0.04348 0.76087 0.10702 0.86789
##         V12     V13     V14    V15     V16     V17      V18     V19     V20
## 235 0.73746 0.70067 0.18227 0.7592 0.13712 0.93478 -0.25084 0.70736 0.18729
##         V21     V22     V23     V24 V25      V26     V27     V28    V29     V30
## 235 0.64883 0.24582 0.60201 0.77425   1 -0.53846 0.89262 0.22216 0.7107 0.53846
##     V31      V32     V33     V34 V35
## 235   1 -0.06522 0.56522 0.23913   0
```

```r
bd_svm_2 <-
  predict_parts(svm_explainer,
                new_observation = train_data[2, -35], type = "break_down")
shap_svm_2 <-
  predict_parts(svm_explainer,
                new_observation = train_data[2, -35], type = "shap")
plot(bd_svm_2)
```



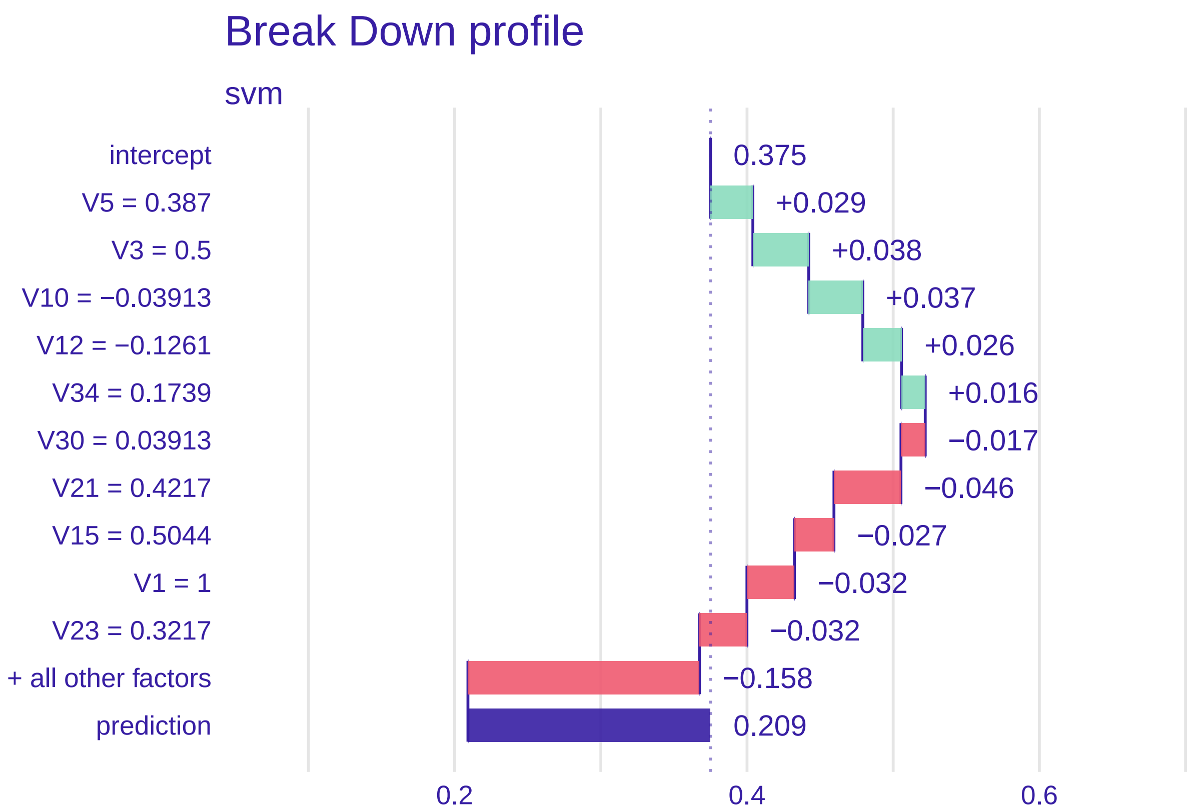

```
plot(shap_svm_2)
```



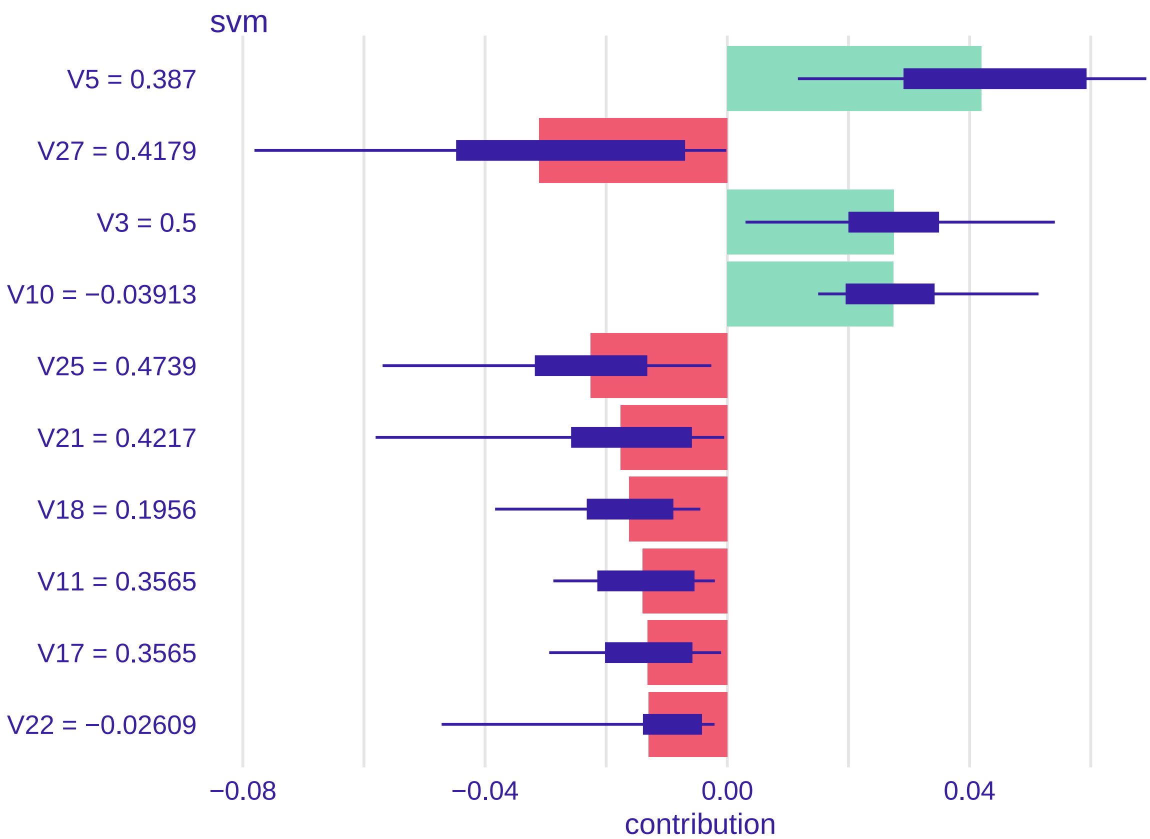

```
# Üçüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[3, ]
```

```
##     V1 V2  V3 V4       V5      V6     V7      V8      V9      V10     V11
## 261  1  0 0.5  0 0.38696 0.10435 0.4913 0.06522 0.46957 -0.03913 0.35652
##          V12     V13     V14     V15     V16     V17     V18     V19     V20
## 261 -0.12609 0.45652 0.04783 0.50435 0.02609 0.35652 0.19565 0.42174 0.14783
##         V21      V22     V23      V24     V25     V26     V27     V28     V29
## 261 0.42174 -0.02609 0.32174 -0.11304 0.47391 -0.0087 0.41789 0.06908 0.38696
##         V30     V31     V32     V33     V34 V35
## 261 0.03913 0.35217 0.14783 0.44783 0.17391   1
```

```
bd_svm_3 <-
  predict_parts(svm_explainer,
                new_observation = train_data[3, -35], type = "break_down")
shap_svm_3 <-
  predict_parts(svm_explainer,
                new_observation = train_data[3, -35], type = "shap")
plot(bd_svm_3)
```



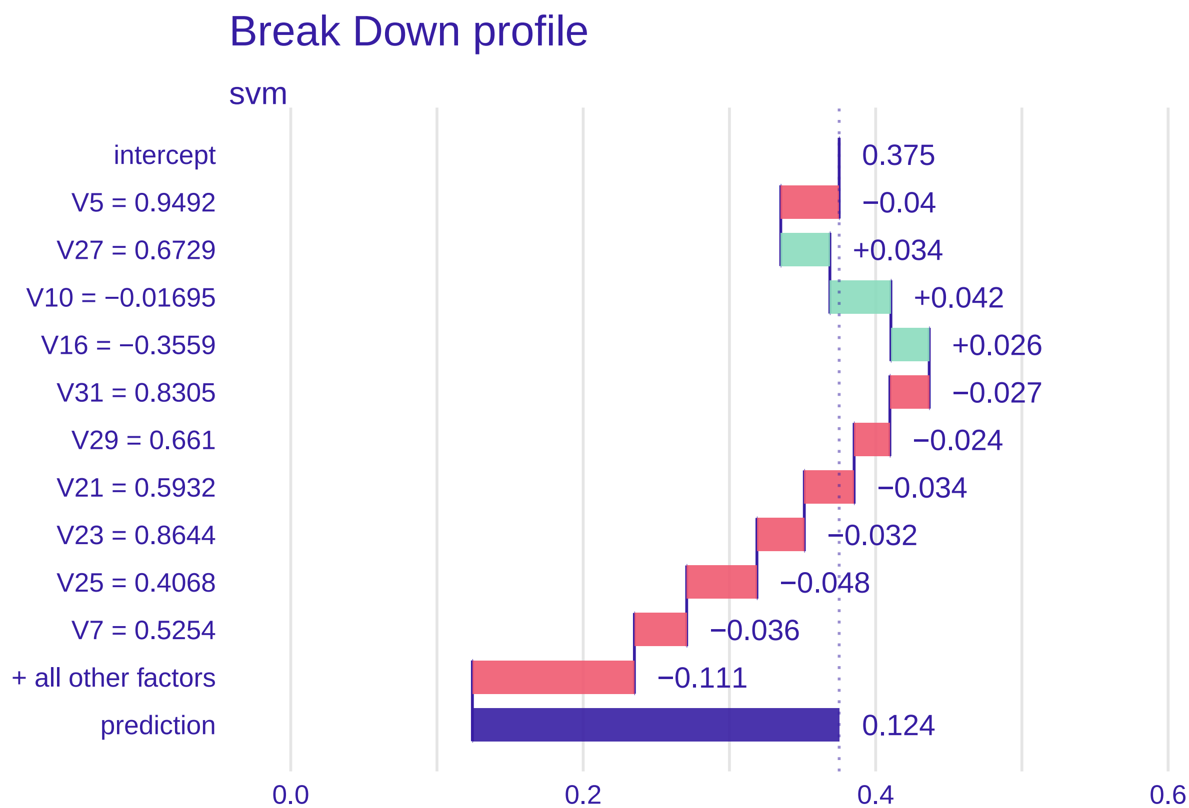

```
plot(shap_svm_3)
```



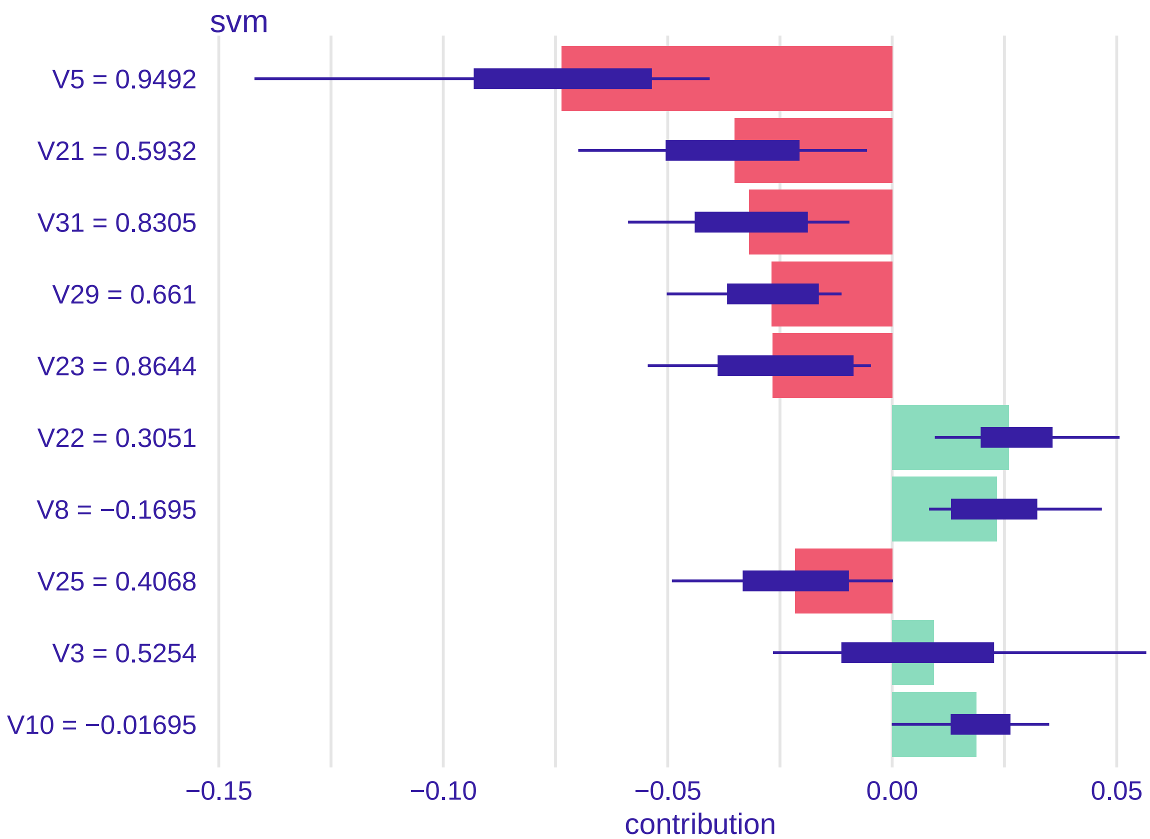

```r
# Dördüncü gözlem için Break Down (bd) ve SHAP değerleri hesaplama ve grafikleme
train_data[4, ]
```

```
##      V1 V2       V3       V4       V5       V6       V7        V8       V9       V10
## 306   1  0 0.52542 -0.0339 0.94915 0.08475 0.52542 -0.16949 0.30508 -0.01695
##          V11       V12      V13      V14      V15       V16      V17      V18      V19
## 306 0.50847 -0.13559 0.64407 0.28814 0.83051 -0.35593 0.54237 0.01695 0.55932
##         V20      V21      V22      V23      V24      V25      V26      V27      V28
## 306 0.0339 0.59322 0.30508 0.86441 0.05085 0.40678 0.15254 0.67287 -0.00266
##          V29      V30      V31       V32      V33       V34 V35
## 306 0.66102 -0.0339 0.83051 -0.15254 0.76271 -0.10169   1
```

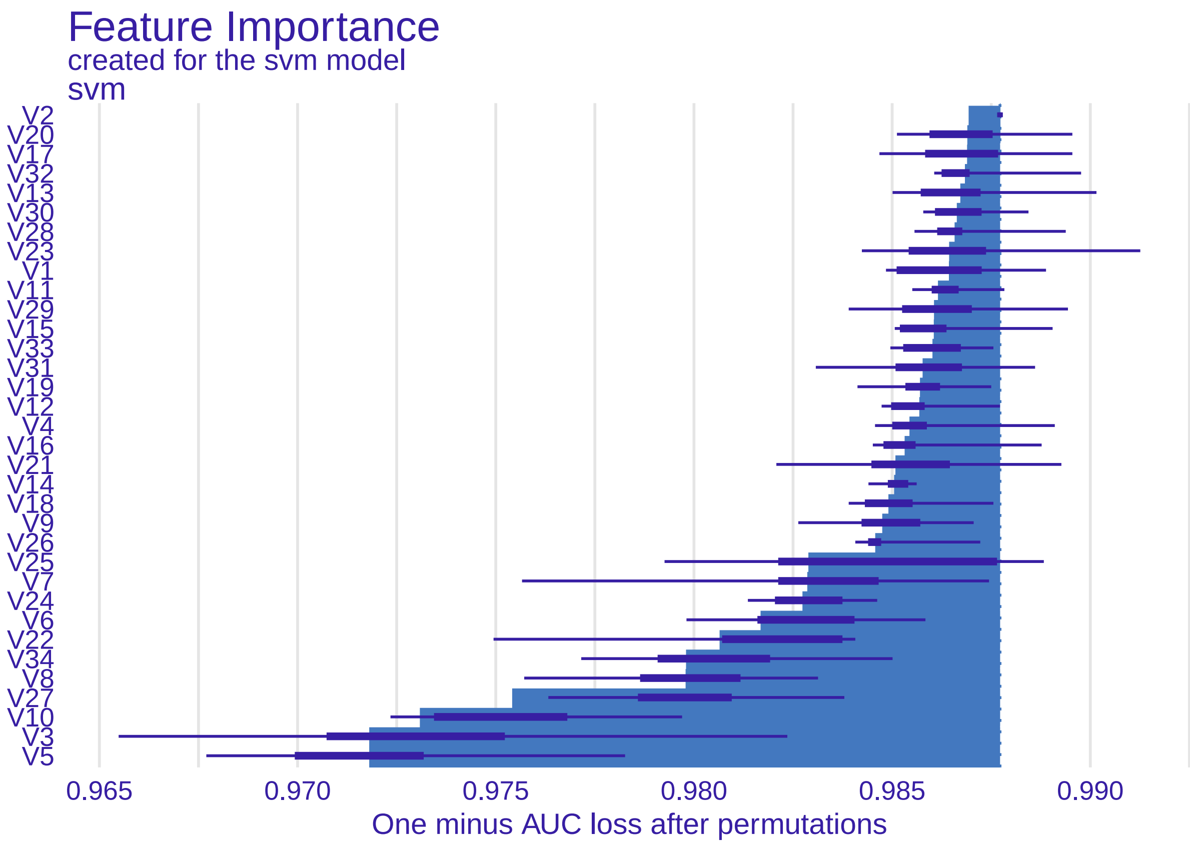
```r
bd_svm_4 <-
  predict_parts(svm_explainer,
                new_observation = train_data[4, -35], type = "break_down")
shap_svm_4 <-
  predict_parts(svm_explainer,
                new_observation = train_data[4, -35], type = "shap")
plot(bd_svm_4)
```




```
plot(shap_svm_4)
```




```
# SVM modelinin değişken önem sıralamasını hesaplama ve grafikleme
svm_var_imp <- DALEX::variable_importance(svm_explainer)
plot(svm_var_imp)
```




# Yorumlar

## forestfire

### XGboost BD & Shap

```
train_data_ff[1,]
```

```
##     X Y month day FFMC   DMC    DC  ISI temp RH wind rain area
## 415 5 4     8   7 93.6 235.1 723.1 10.1 24.1 50    4    0    0
```

```
train_data_ff[2,]
```

```
##     X Y month day FFMC   DMC    DC ISI temp RH wind rain area
## 463 1 4     9   7   91 276.3 825.1 7.1 14.5 76  7.6    0 3.71
```

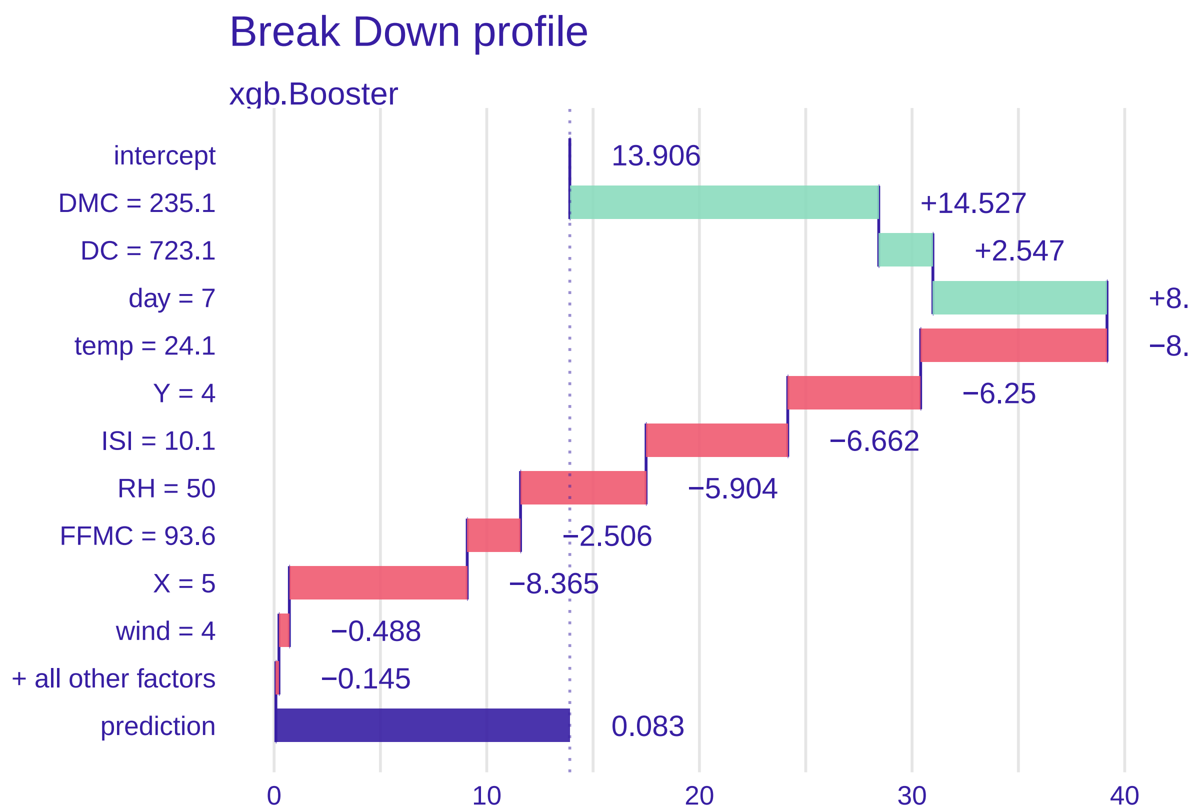
```
train_data_ff[3,]
```

```
##     X Y month day FFMC  DMC    DC ISI temp RH wind rain area
## 179 2 5     9   3 90.1 82.9 735.7 6.2 18.3 45  2.2    0 4.88
```

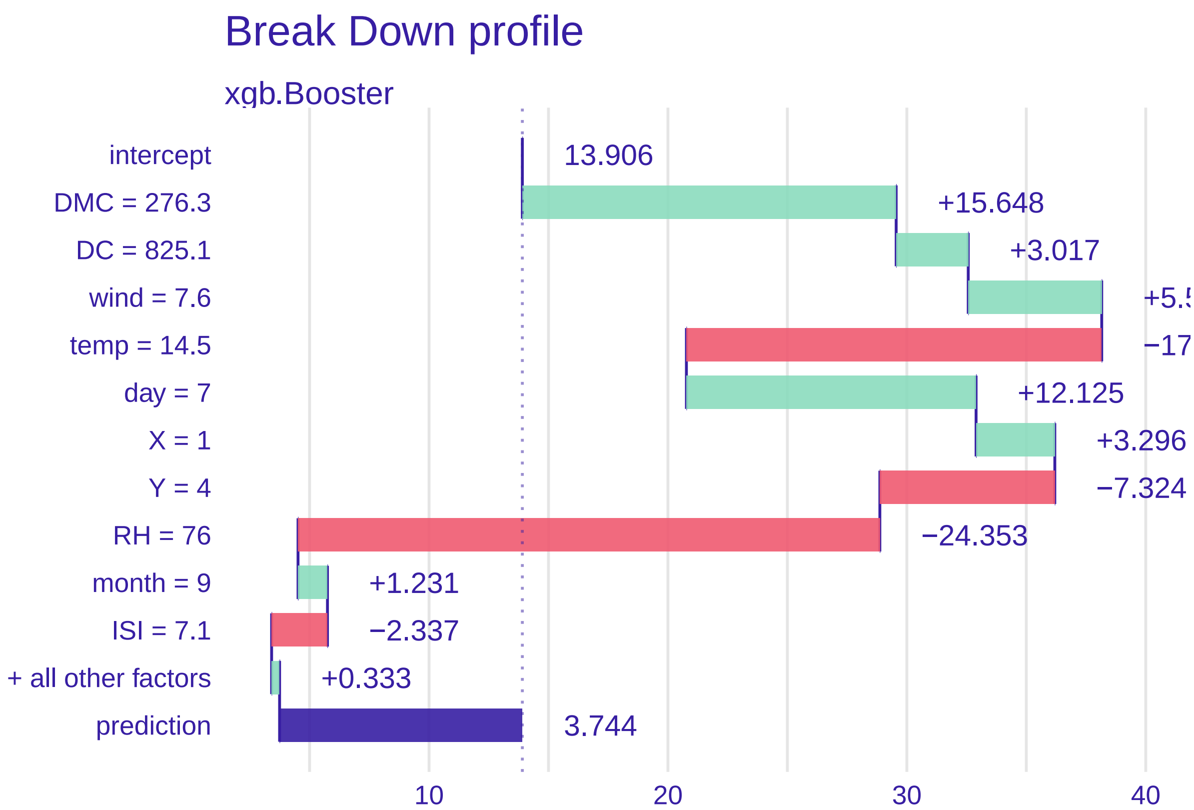
```
train_data_ff[4,]
```

```
##       X Y month day FFMC   DMC     DC ISI temp RH wind rain  area
## 195 2 2      8   2 94.8 108.3 647.1  17 24.6 22  4.5    0 10.01
```

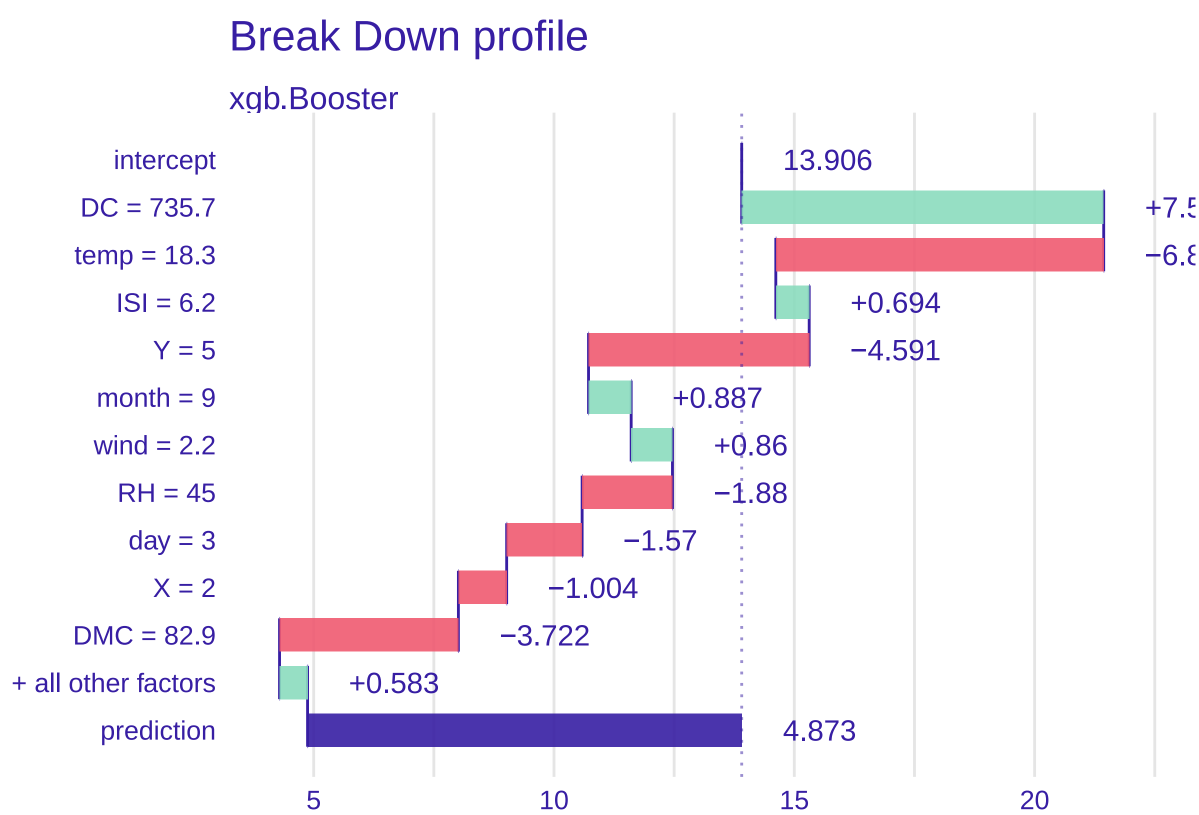
```
plot(bd_xgb_1)
```




```
plot(bd_xgb_2)
```

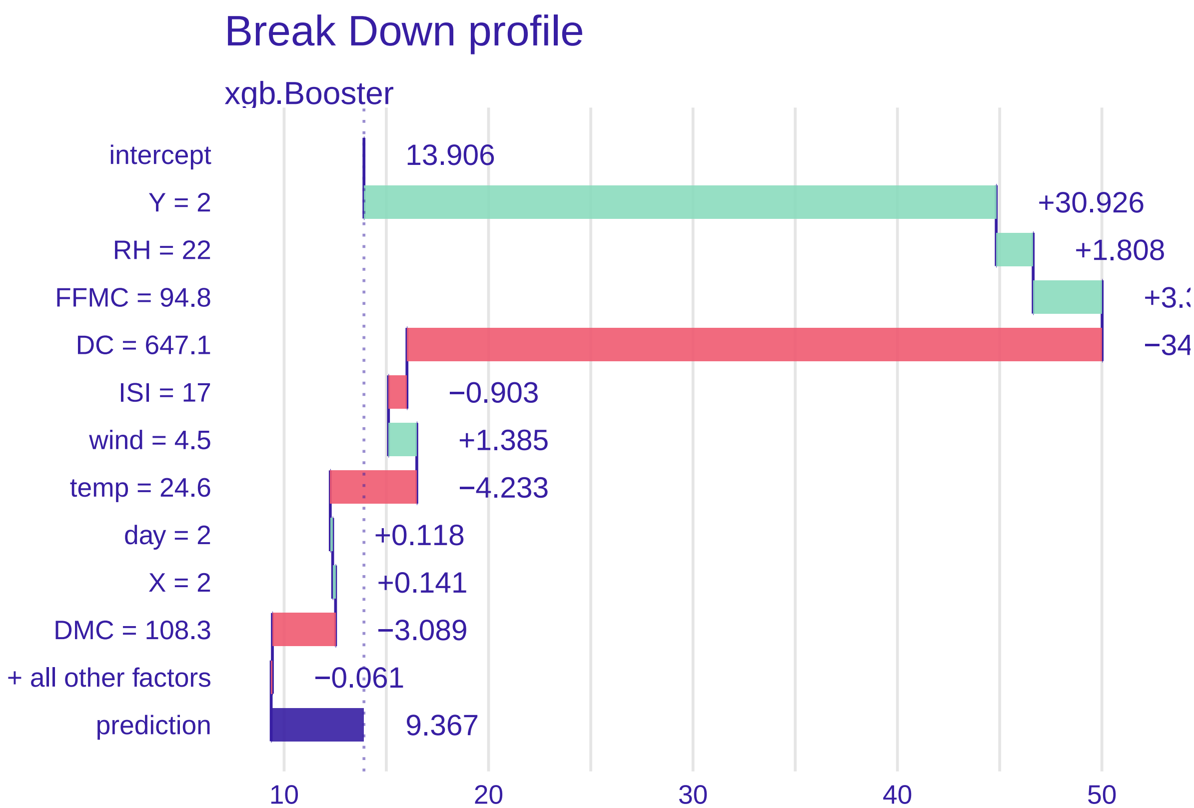

Break Down profile
xgb.Booster
intercept
DMC = 276.3
DC = 825.1
wind = 7.6
temp = 14.5
day = 7
X = 1
Y = 4
RH = 76
month = 9
ISI = 7.1
+ all other factors
prediction
13.906
+15.648
+3.017
+12.125
+3.296
−7.324
−24.353
+1.231
−2.337
+0.333
3.744
10
20
30
40

```
plot(bd_xgb_3)
```

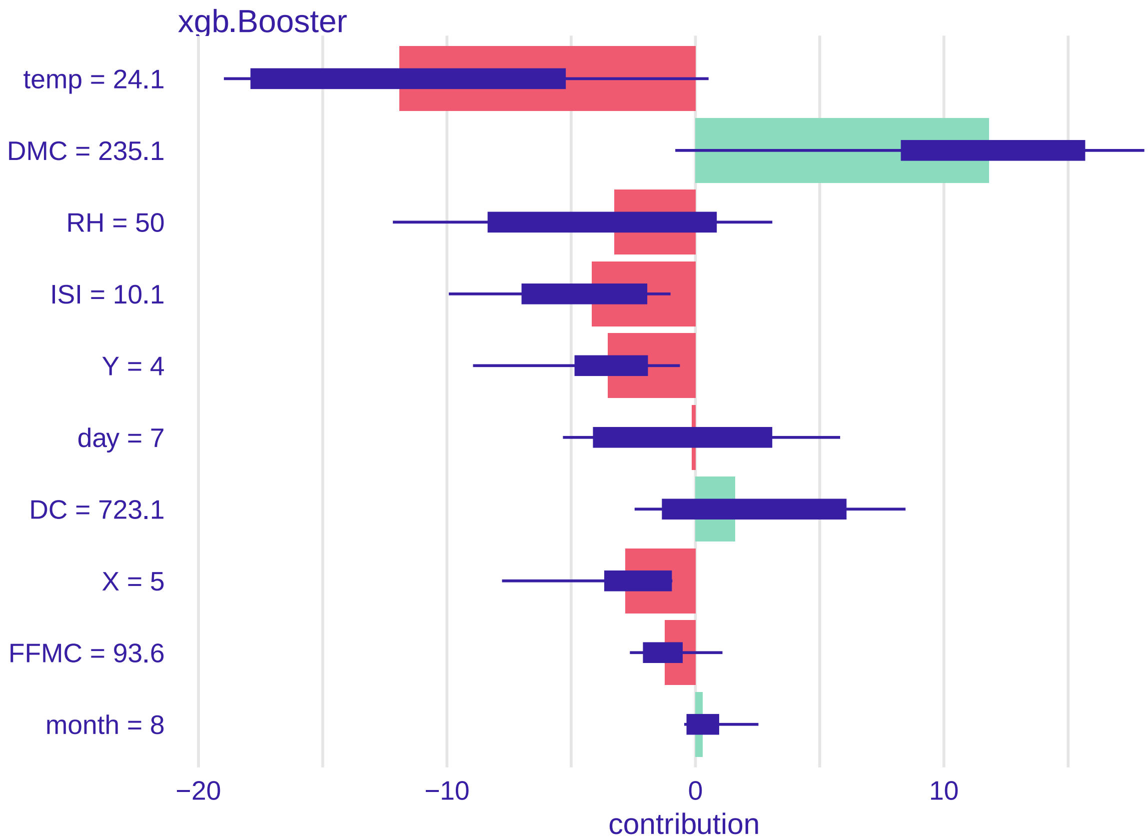


Break Down profile
xgb.Booster
intercept
DC = 735.7
temp = 18.3
ISI = 6.2
Y = 5
month = 9
wind = 2.2
RH = 45
day = 3
X = 2
DMC = 82.9
+ all other factors
prediction
13.906
+0.694
−4.591
+0.887
+0.86
−1.88
−1.57
−1.004
−3.722
+0.583
4.873
5
10
15
20

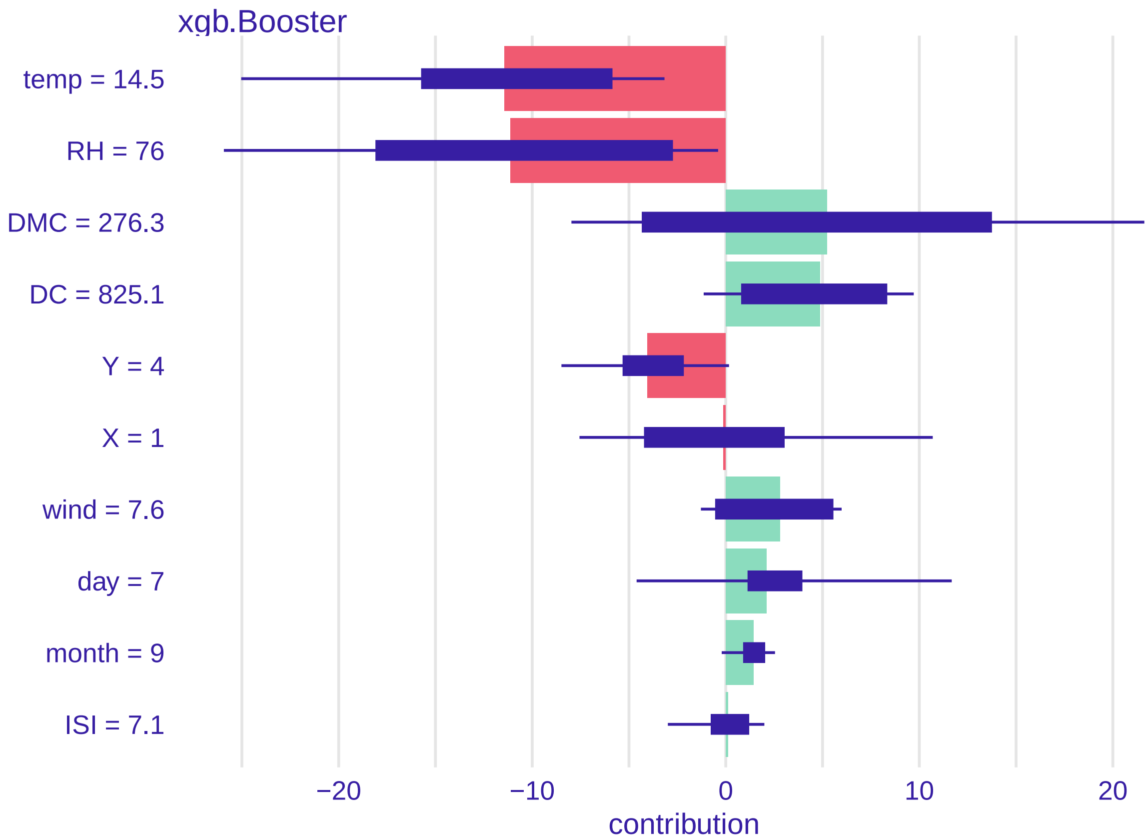

```
plot(bd_xgb_4)
```



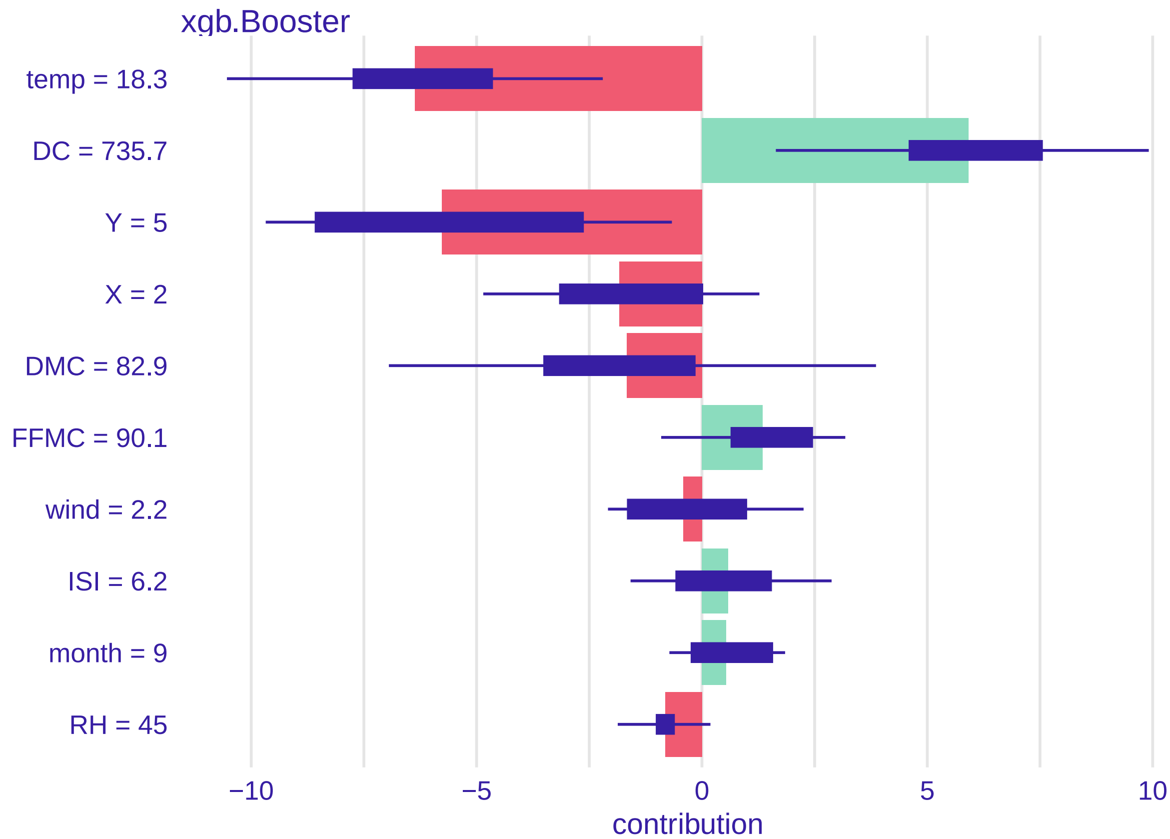

```
plot(shap_xgb_1)
```



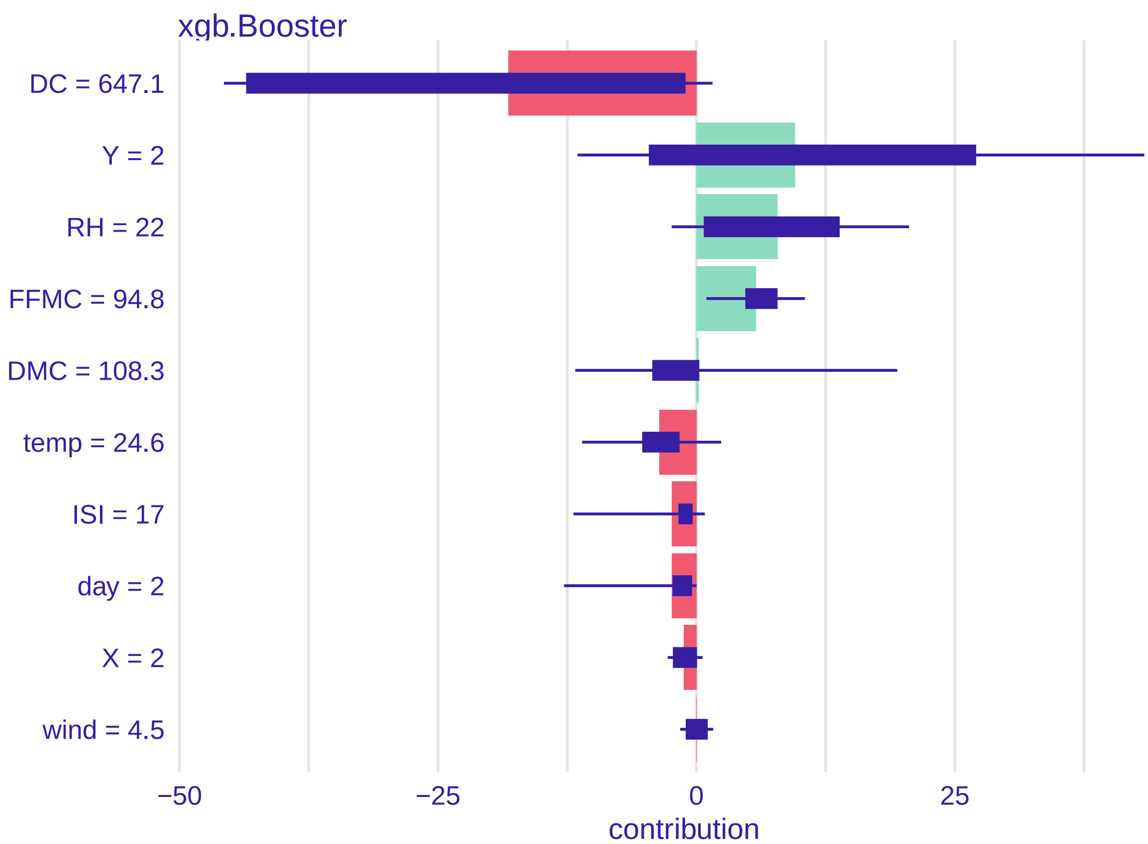

```
plot(shap_xgb_2)
```




```
plot(shap_xgb_3)
```




```
plot(shap_xgb_4)
```




**Break Down** İlk grafikte (1. gözlem), DMC (Duff Moisture Code) ve DC (Drought Code) değerlerinin model tahmini üzerinde büyük pozitif etkisi olduğunu görüyoruz. Bu iki değişken, orman yangını riskini değerlendirirken orman tabanı ve organik katmanın nem durumunu gösterir ve bu modelde yangın alanını tahmin etmede en belirleyici faktörlerden biri olarak ortaya çıkıyor.

İkinci grafikte (2. gözlem), yine DMC ve DC değerlerinin model tahmini üzerinde büyük pozitif etkisi var, ancak bu sefer rüzgar hızının (wind) da önemli bir etkisi olduğunu gözlemliyoruz. Yüksek rüzgar hızı yangının daha büyük alanlara yayılma riskini artırabilir.

Üçüncü grafikte (3. gözlem), DC'nin yine büyük bir pozitif etki yaptığı görülüyor, ancak DMC'nin etkisi bu sefer negatif. Bunun yanı sıra, göreceli nem (RH) ve rüzgar hızı (wind) gibi diğer değişkenlerin de küçük pozitif katkıları var.

Dördüncü grafikte (4. gözlem), model tahmini üzerinde en büyük pozitif etkiyi yine DMC ve DC'nin yaptığı görülüyor. Ancak bu sefer, ISI (Initial Spread Index) ve sıcaklık (temp) gibi diğer değişkenlerin de model tahminine büyük etkileri olduğu gözlemleniyor.

### Shap

1. Grafik (1. gözlem): Sıcaklık (temp) ve DMC en büyük negatif etkiye sahipken, DC ve ay (month) en büyük pozitif etkiye sahip. ISI ve göreceli nem (RH) gibi diğer özelliklerin de küçük katkıları var.

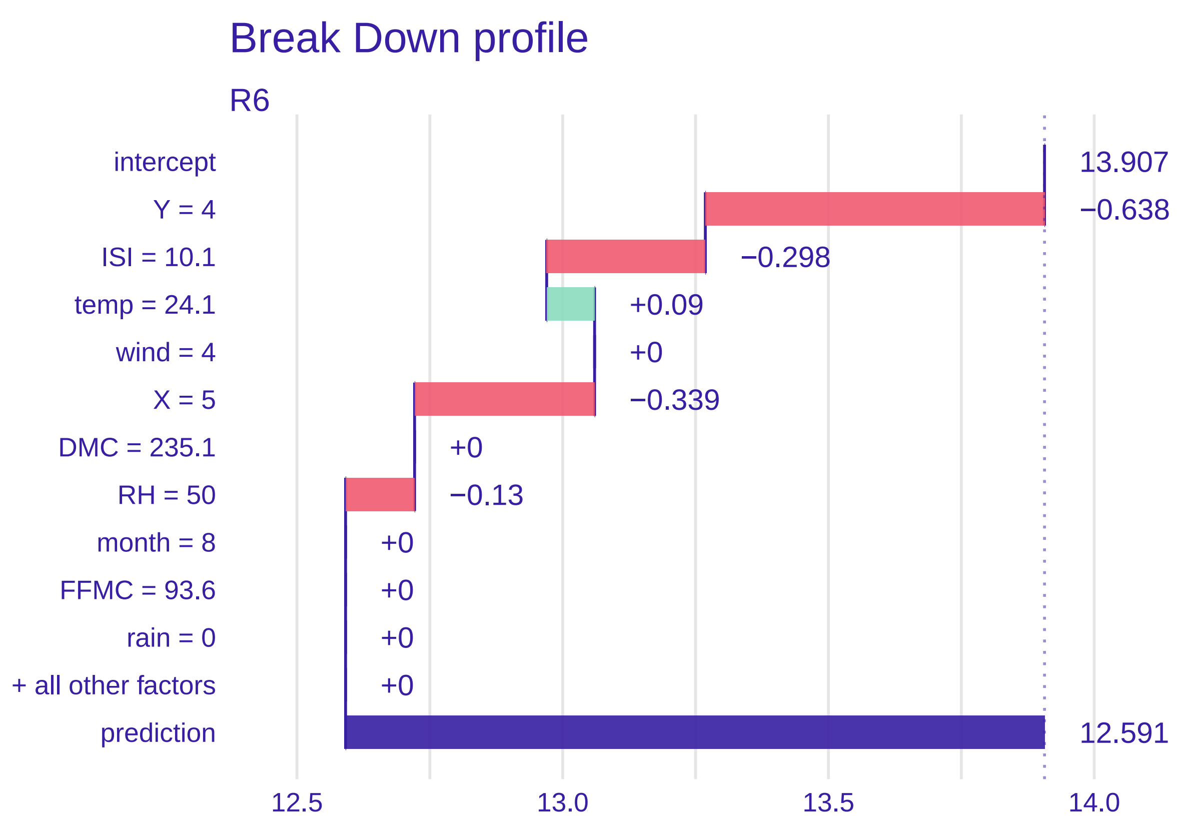
2. Grafik (2. gözlem): Bu grafikte, DMC'nin yüksek pozitif bir etkisi varken, RH (göreceli nem) ciddi bir negatif katkıya sahip. Diğer özelliklerin katkıları karışık; DC ve ay (month) pozitif, ISI ve rüzgar hızı (wind) negatif katkılar sağlıyor.

3. Grafik (3. gözlem): DMC ve ISI'nin küçük pozitif katkıları varken, DC ve ay (month) büyük pozitif etkilere sahip. Sıcaklık (temp) ve RH (göreceli nem) negatif katkılar sağlıyor.
4. Grafik (4. gözlem): Bu grafikte, DC ve ay (month) büyük pozitif etkiler gösterirken, temp ve ISI büyük negatif etkiler gösteriyor. DMC ve rüzgar hızının (wind) küçük negatif etkileri var.

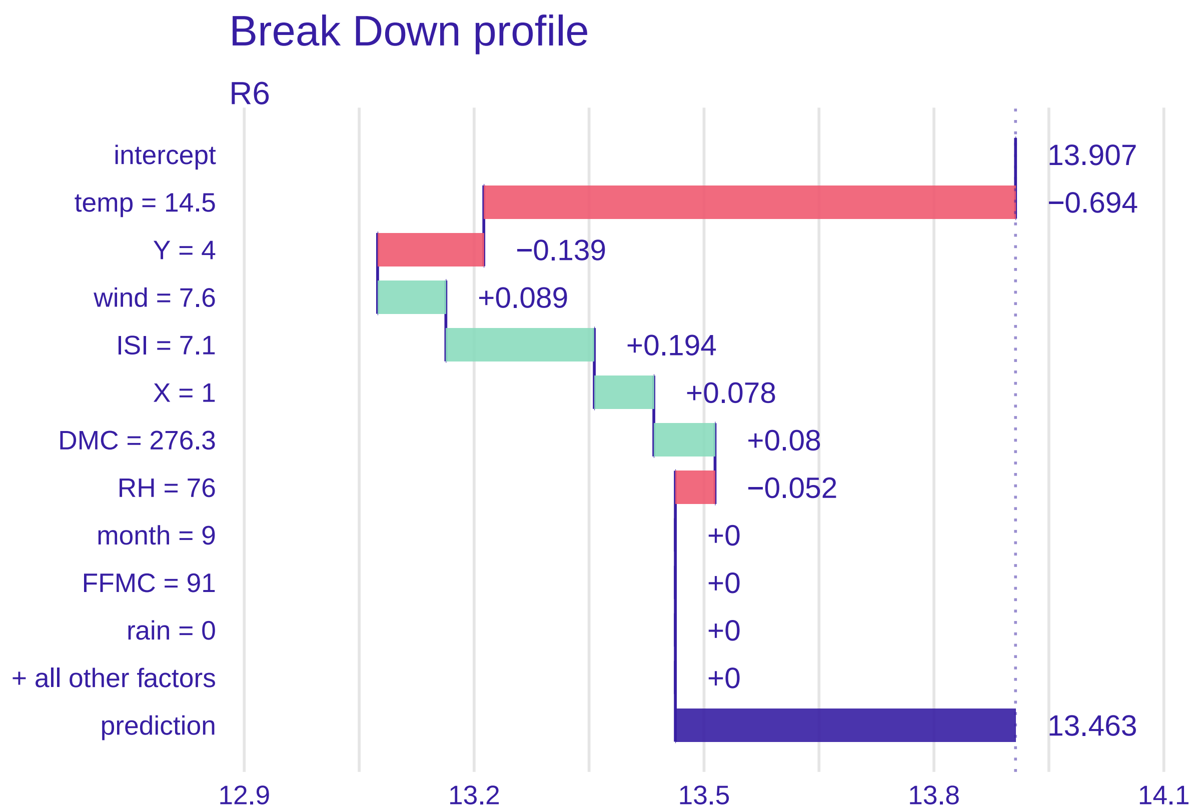
XGBoost modeli tarafından üretilen SHAP ve Break Down analizleri, orman yangını risk tahminlerinde DMC ve DC'nin özellikle belirleyici olduğunu göstermektedir. Her iki analiz türü de, sıcaklık ve rüzgar hızının tahminleri önemli ölçüde etkilediğini, ancak bu etkilerin gözlem noktalarına göre değişkenlik gösterdiğini ortaya koymaktadır. Göreceli nemin genellikle yangın riskini düşüren bir faktör olarak öne çıktığı, mevsimsellik etkisinin ise ay değişkeninde açıkça görüldüğü anlaşılmaktadır. Bu analizler, modelin hangi özelliklere daha fazla ağırlık verdiğini ve bu özelliklerin yangın büyüklüğü ve yayılımı üzerindeki potansiyel etkilerini derinlemesine anlamak için kritik bilgiler sağlar.

**lighgbm BD & Shap**

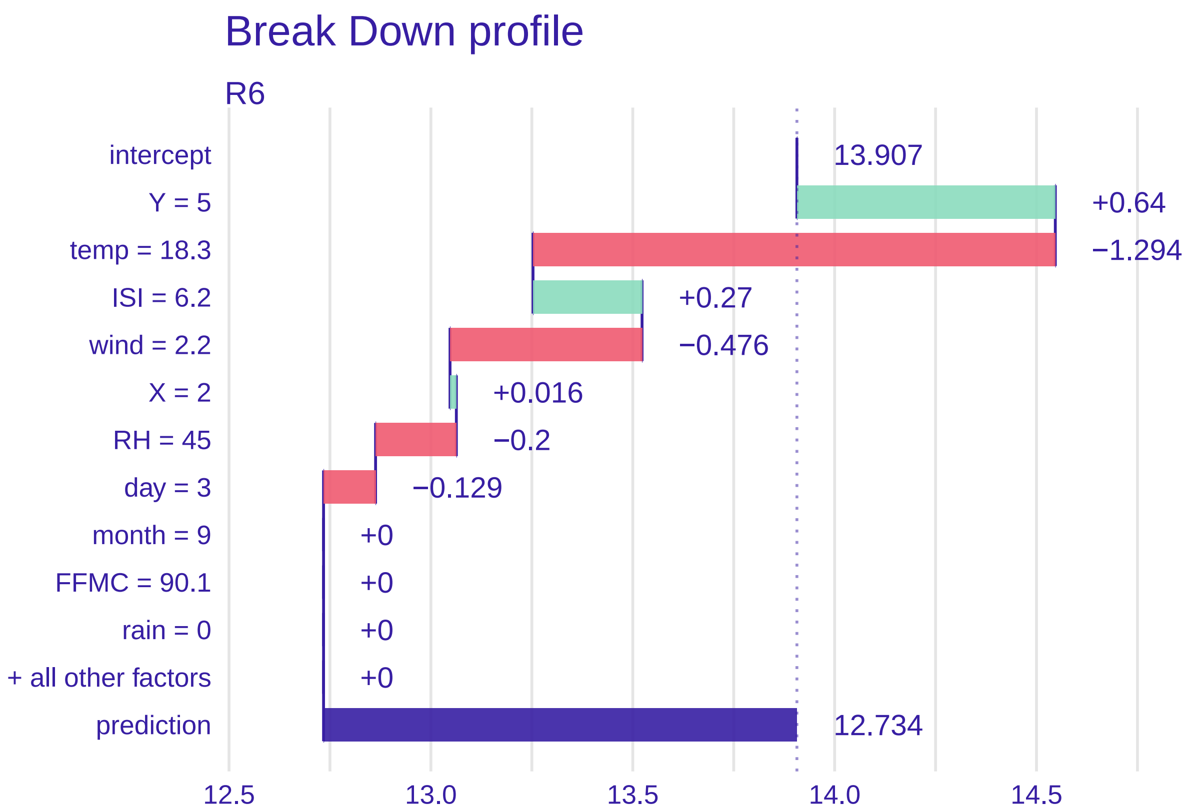
```
plot(bd_lgb_1)
```



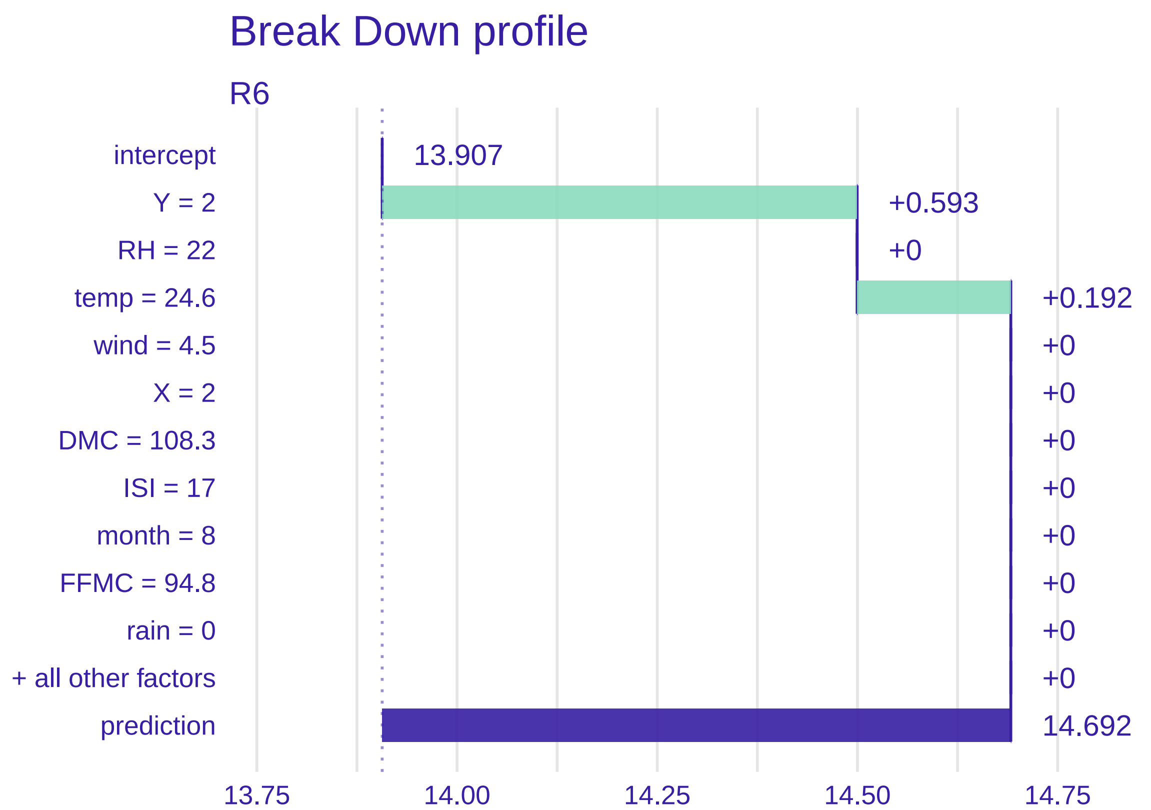

```
plot(bd_lgb_2)
```


Break Down profile
R6
intercept
temp = 14.5
Y = 4
wind = 7.6
ISI = 7.1
X = 1
DMC = 276.3
RH = 76
month = 9
FFMC = 91
rain = 0
+ all other factors
prediction
13.907
−0.694
−0.139
+0.089
+0.194
+0.078
+0.08
−0.052
+0
+0
+0
+0
13.463
12.9
13.2
13.5
13.8
14.1

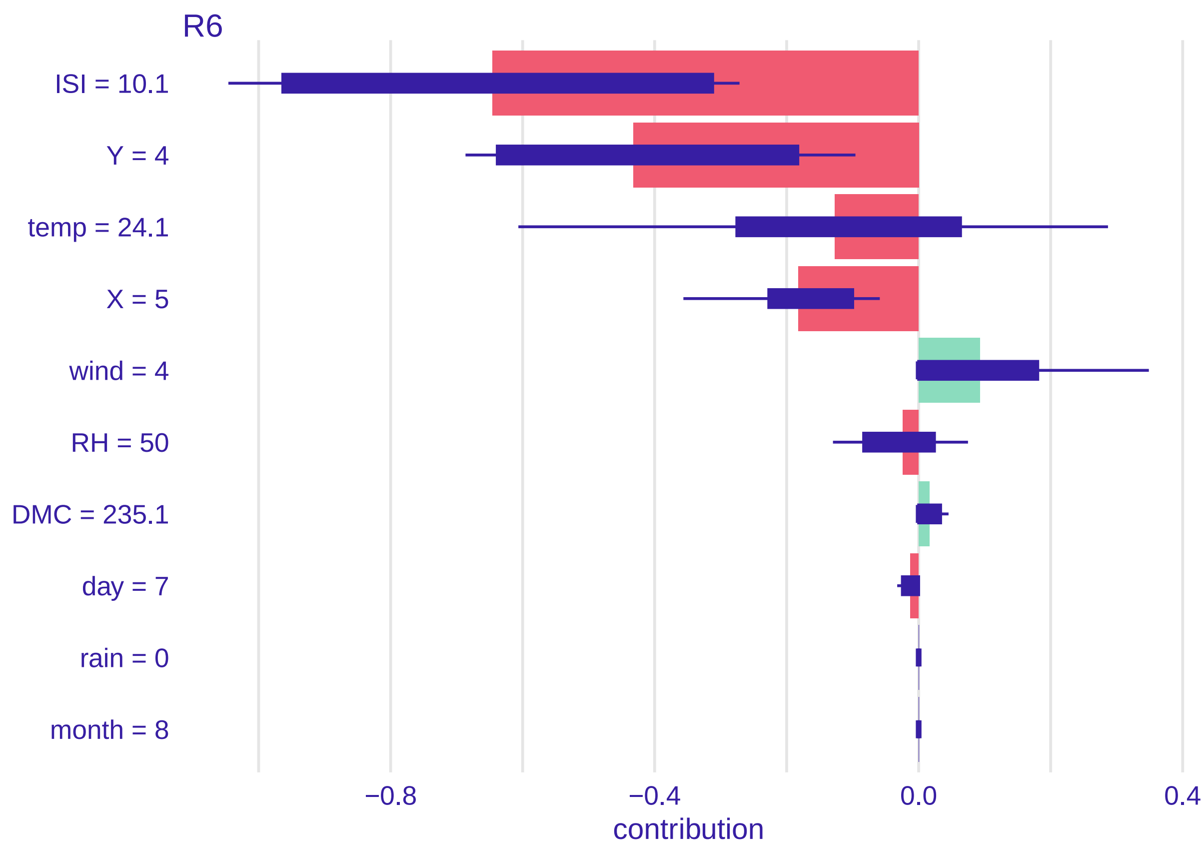
```
plot(bd_lgb_3)
```


Break Down profile
R6
intercept
Y = 5
temp = 18.3
ISI = 6.2
wind = 2.2
X = 2
RH = 45
day = 3
month = 9
FFMC = 90.1
rain = 0
+ all other factors
prediction
13.907
+0.64
−1.294
+0.27
−0.476
+0.016
−0.2
−0.129
+0
+0
+0
+0
12.734
12.5
13.0
13.5
14.0
14.5

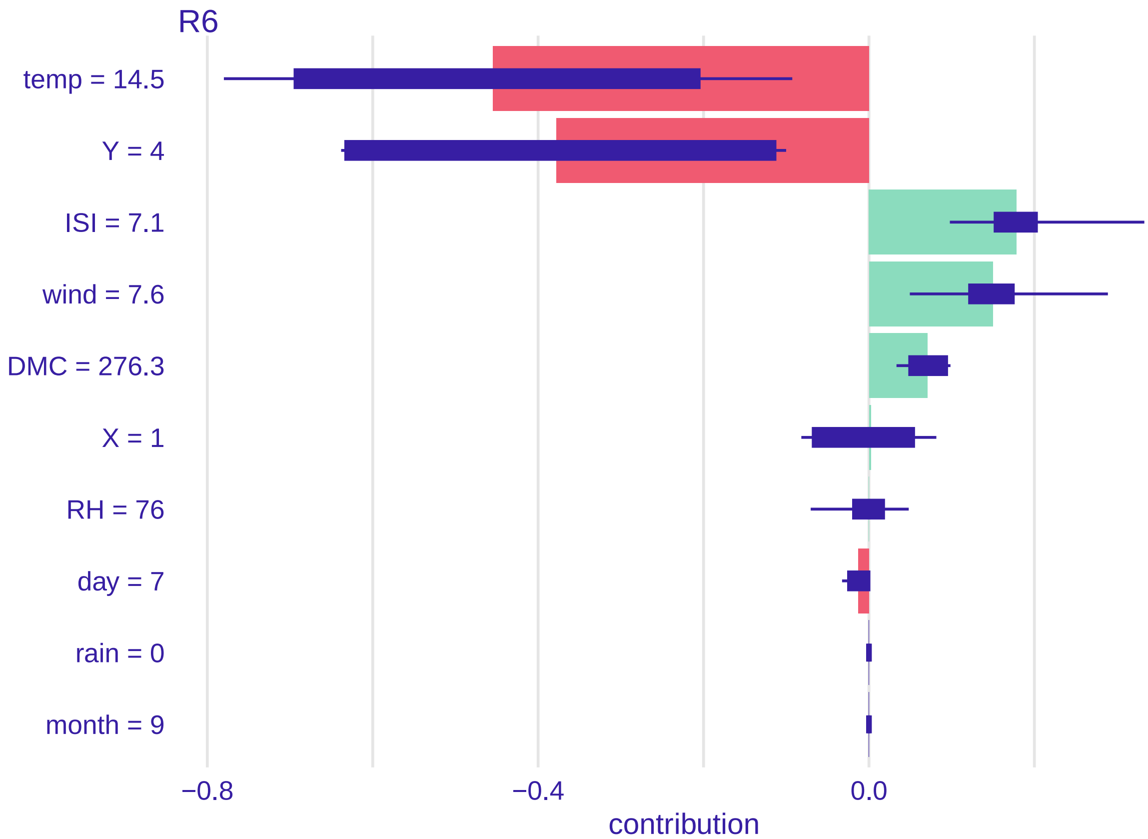
```
plot(bd_lgb_4)
```



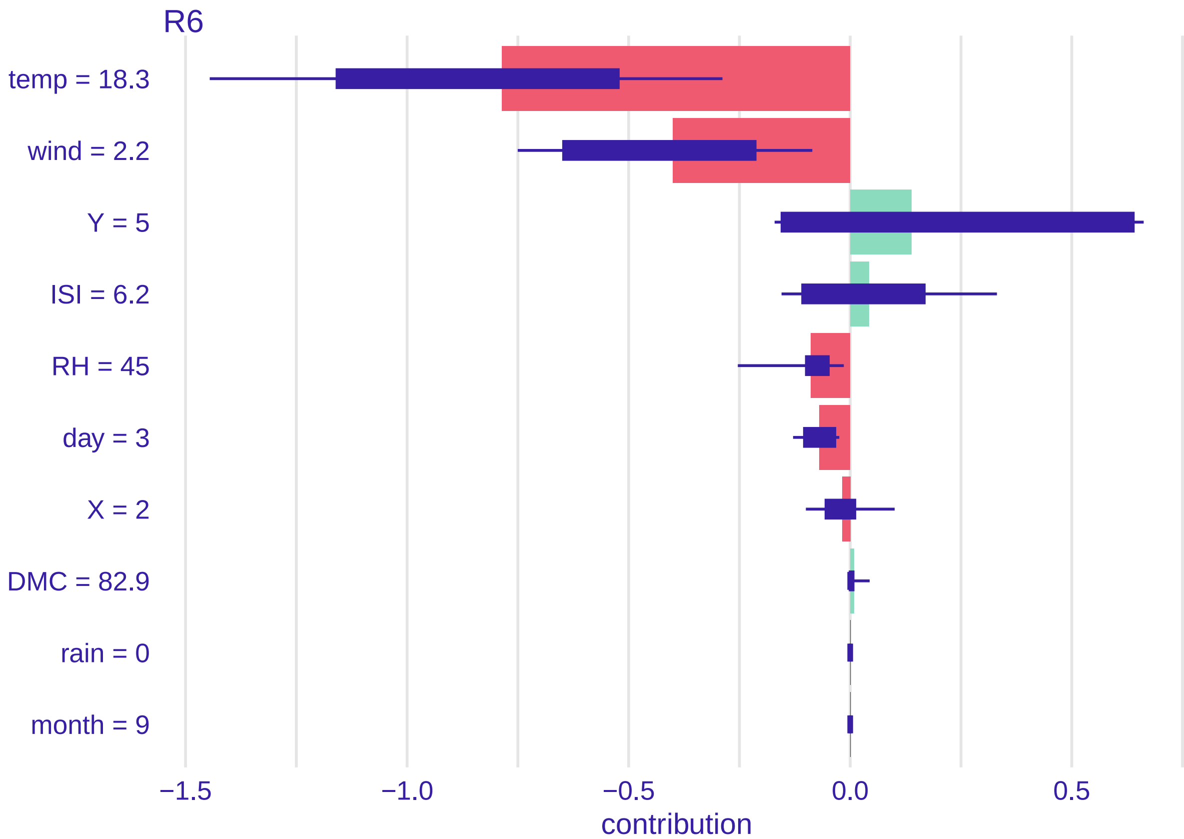

```
plot(shap_lgb_1)
```



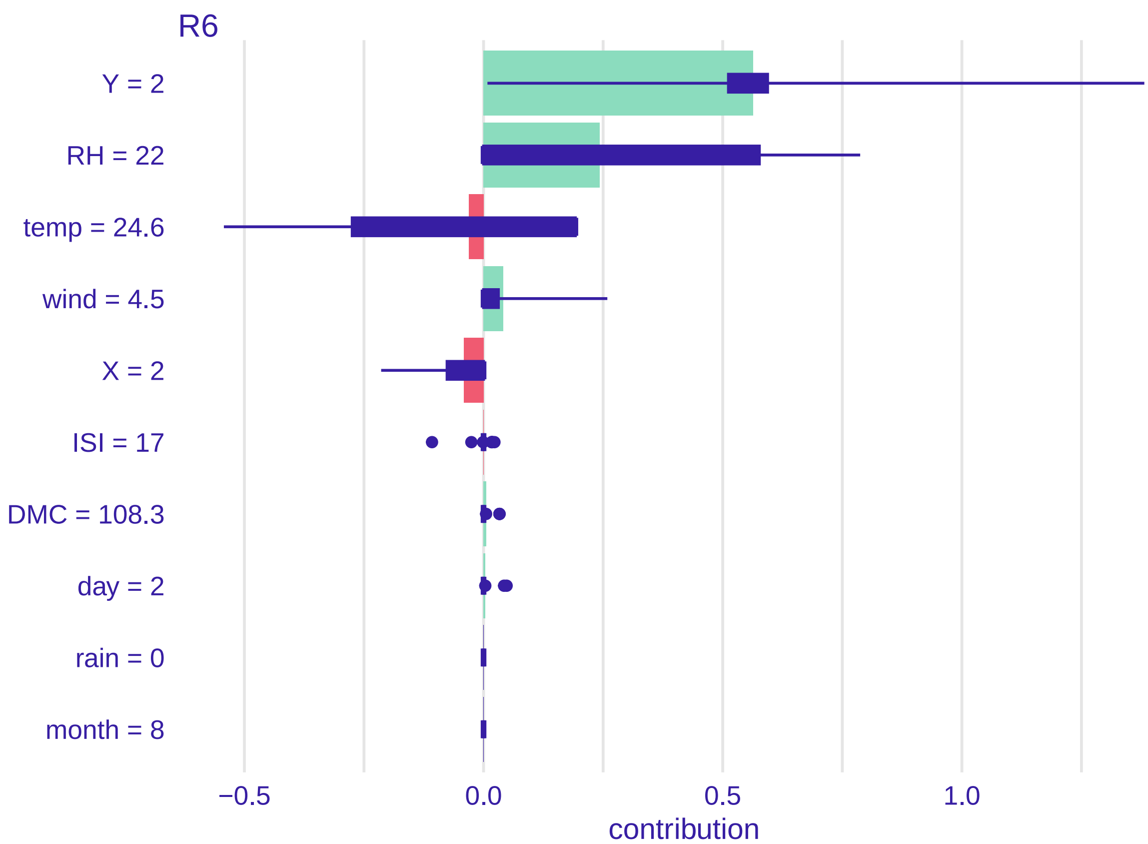

```
plot(shap_lgb_2)
```




```
plot(shap_lgb_3)
```

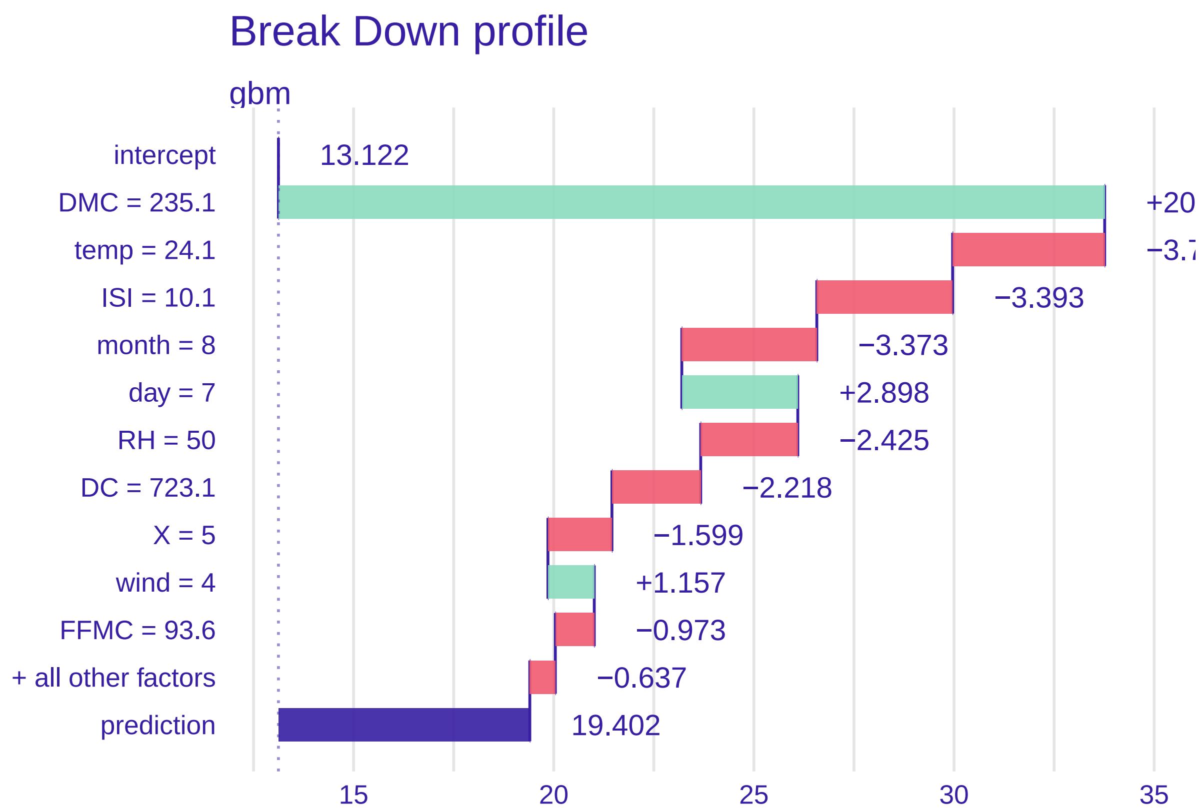



```
plot(shap_lgb_4)
```

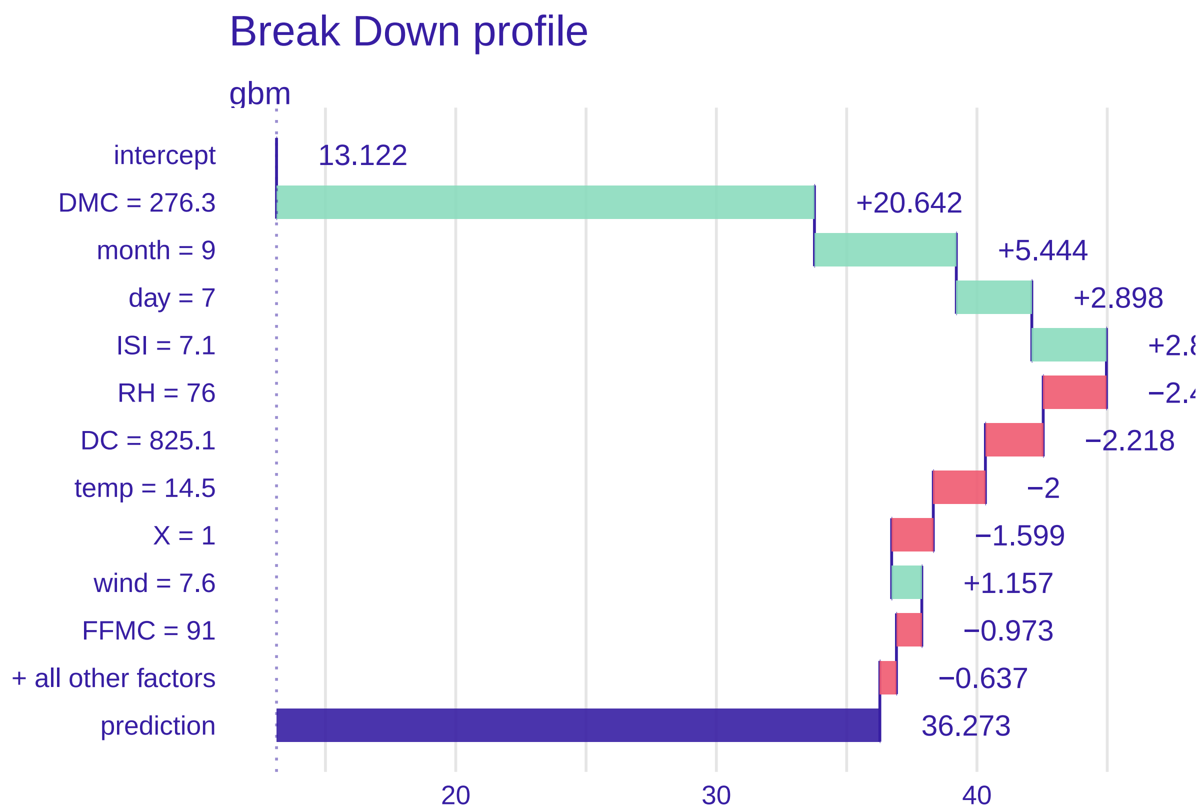



**gbm BD & Shap**

```
plot(bd_gbm_1)
```

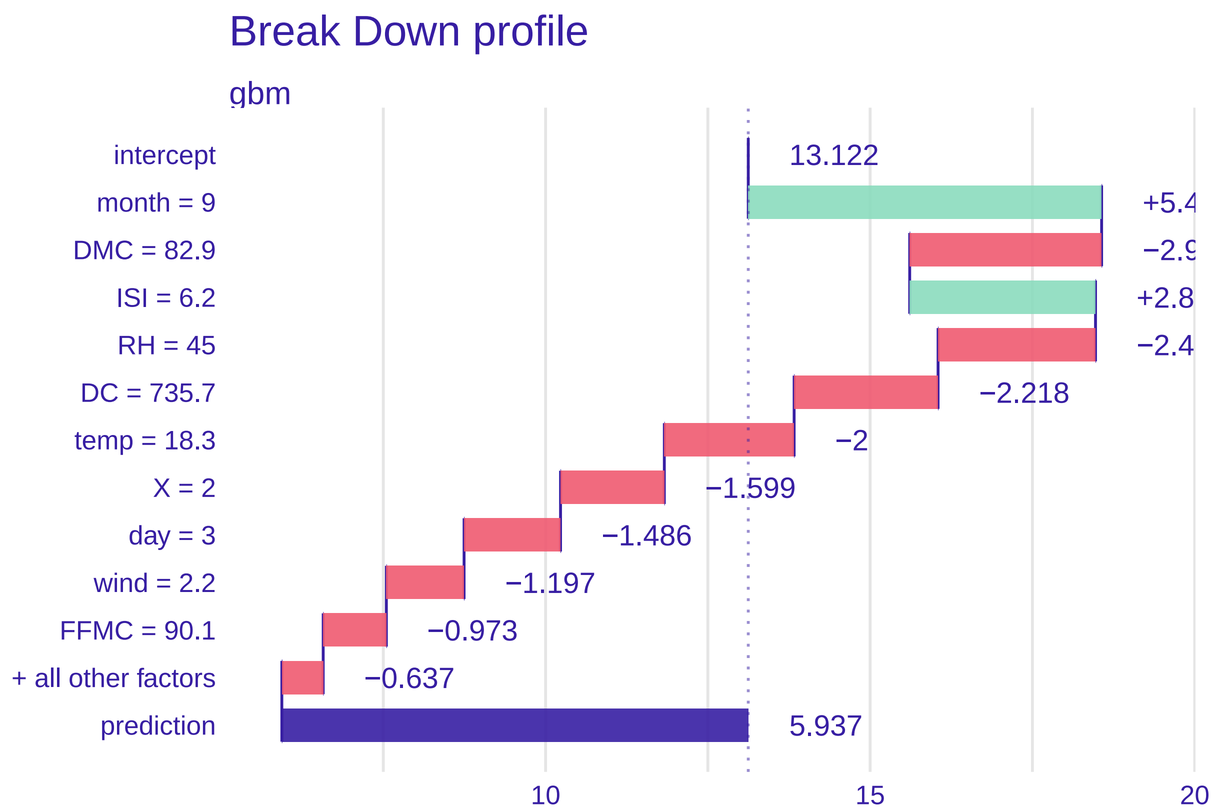

Break Down profile
gbm
intercept
DMC = 235.1
temp = 24.1
ISI = 10.1
month = 8
day = 7
RH = 50
DC = 723.1
X = 5
wind = 4
FFMC = 93.6
+ all other factors
prediction
13.122
+20
−3.393
−3.373
+2.898
−2.425
−2.218
−1.599
+1.157
−0.973
−0.637
19.402
15
20
25
30
35

```
plot(bd_gbm_2)
```

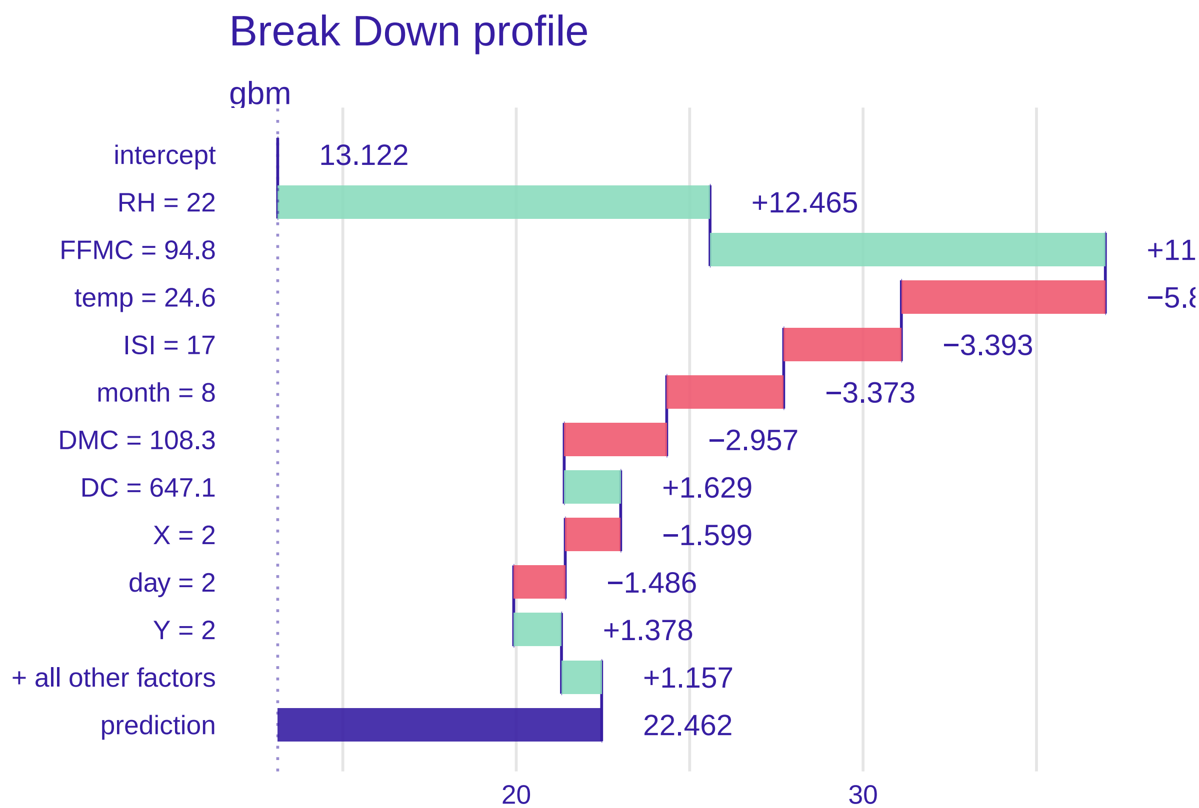


Break Down profile
gbm
intercept
DMC = 276.3
month = 9
day = 7
ISI = 7.1
RH = 76
DC = 825.1
temp = 14.5
X = 1
wind = 7.6
FFMC = 91
+ all other factors
prediction
13.122
+20.642
+5.444
+2.898
−2.218
−2
−1.599
+1.157
−0.973
−0.637
36.273
20
30
40

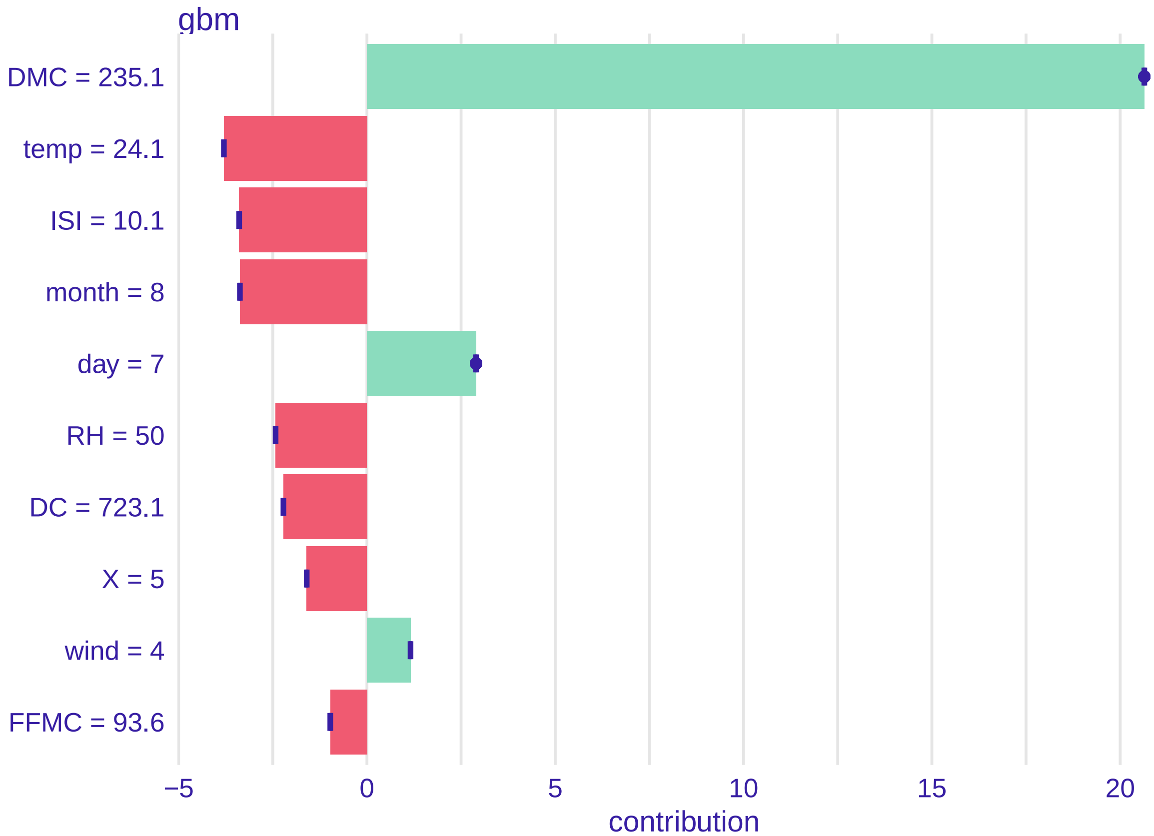

```
plot(bd_gbm_3)
```


Break Down profile
gbm
intercept
month = 9
DMC = 82.9
ISI = 6.2
RH = 45
DC = 735.7
temp = 18.3
X = 2
day = 3
wind = 2.2
FFMC = 90.1
+ all other factors
prediction
13.122
−2.218
−2
−1.599
−1.486
−1.197
−0.973
−0.637
5.937
10
15
20

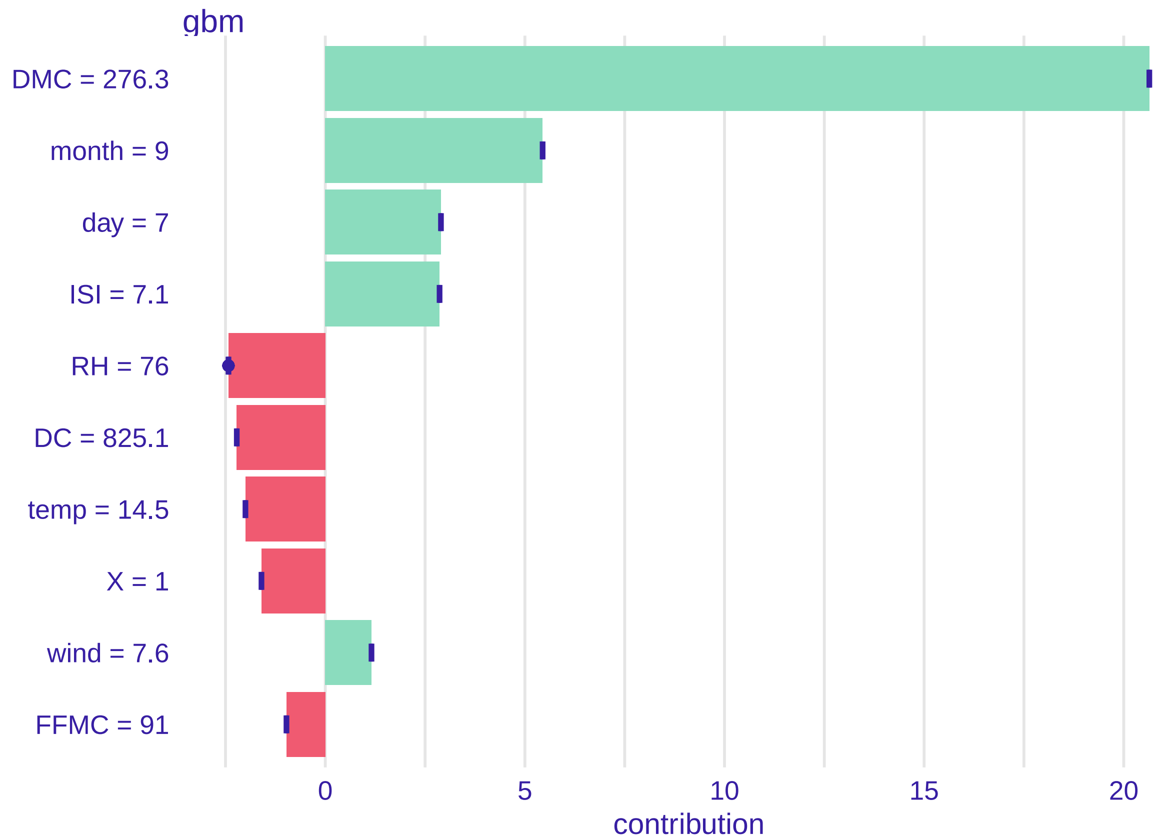
```
plot(bd_gbm_4)
```



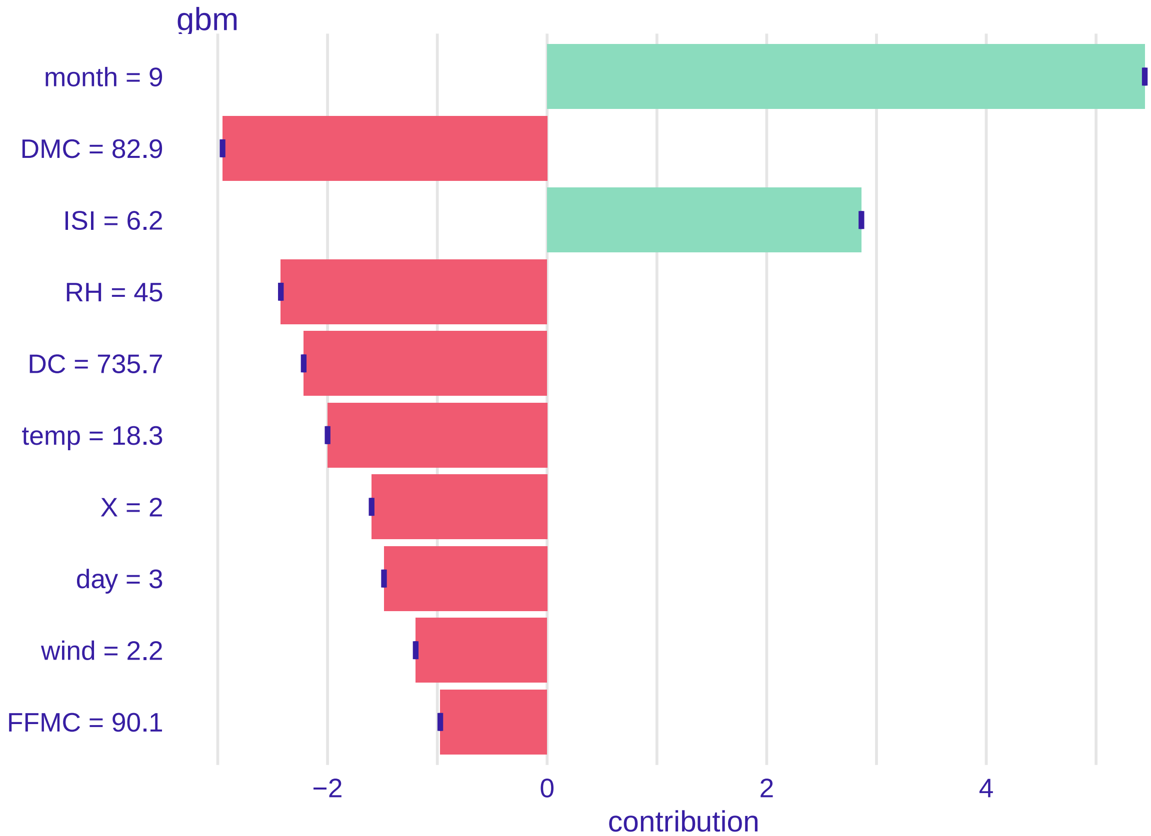

```
plot(shap_gb_1)
```



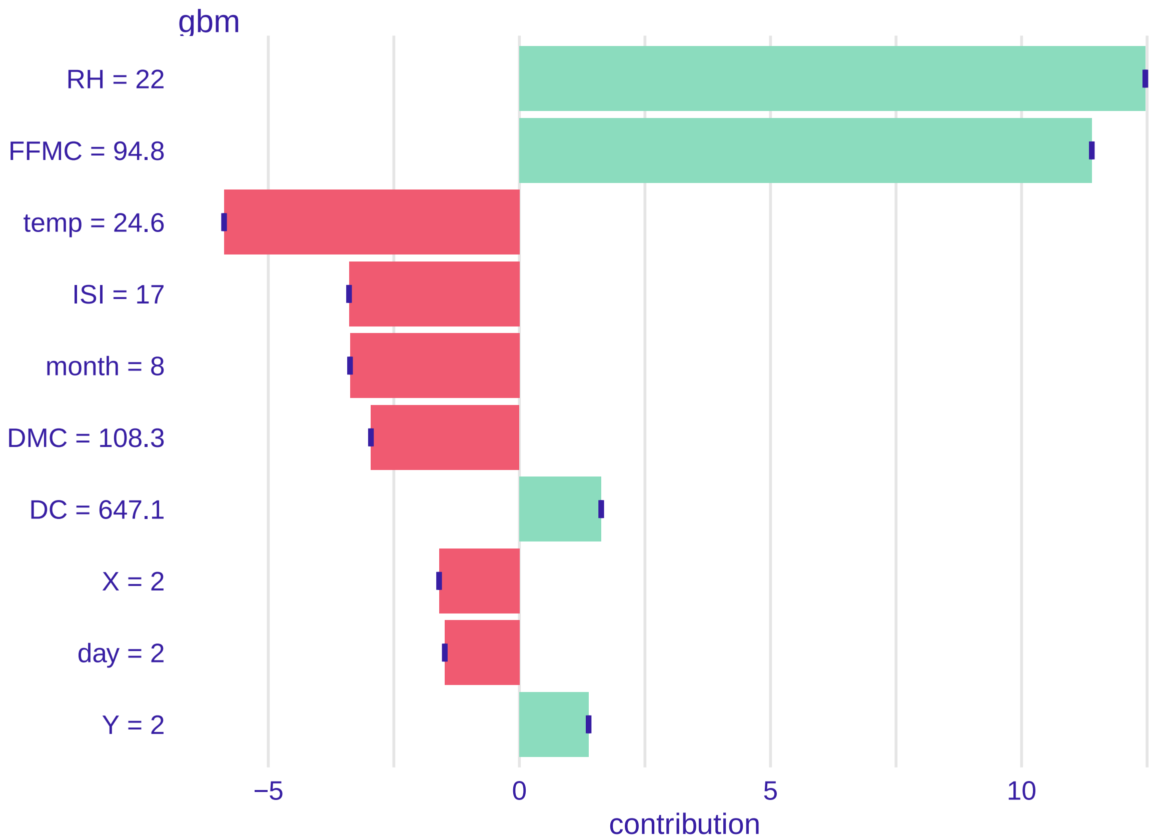

```
plot(shap_gb_2)
```




```
plot(shap_gb_3)
```

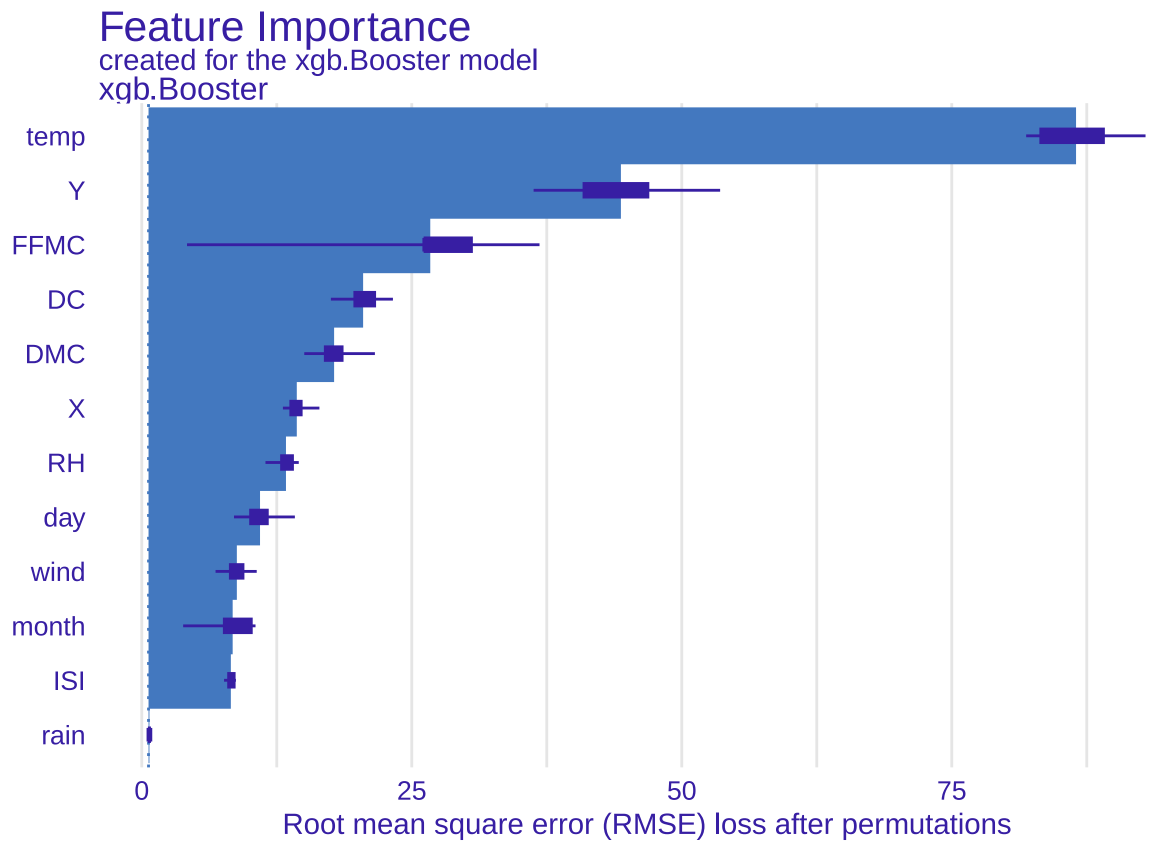



```
plot(shap_gb_4)
```

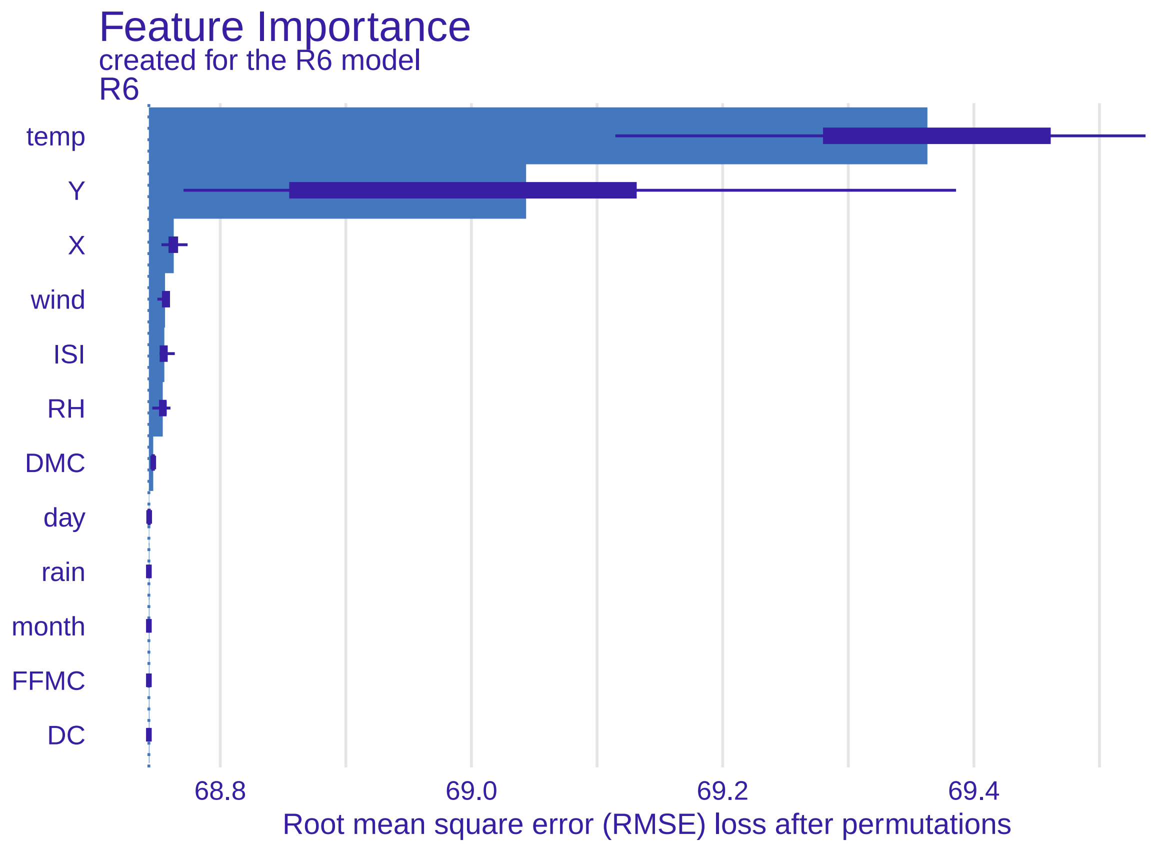



**Variable Importance**

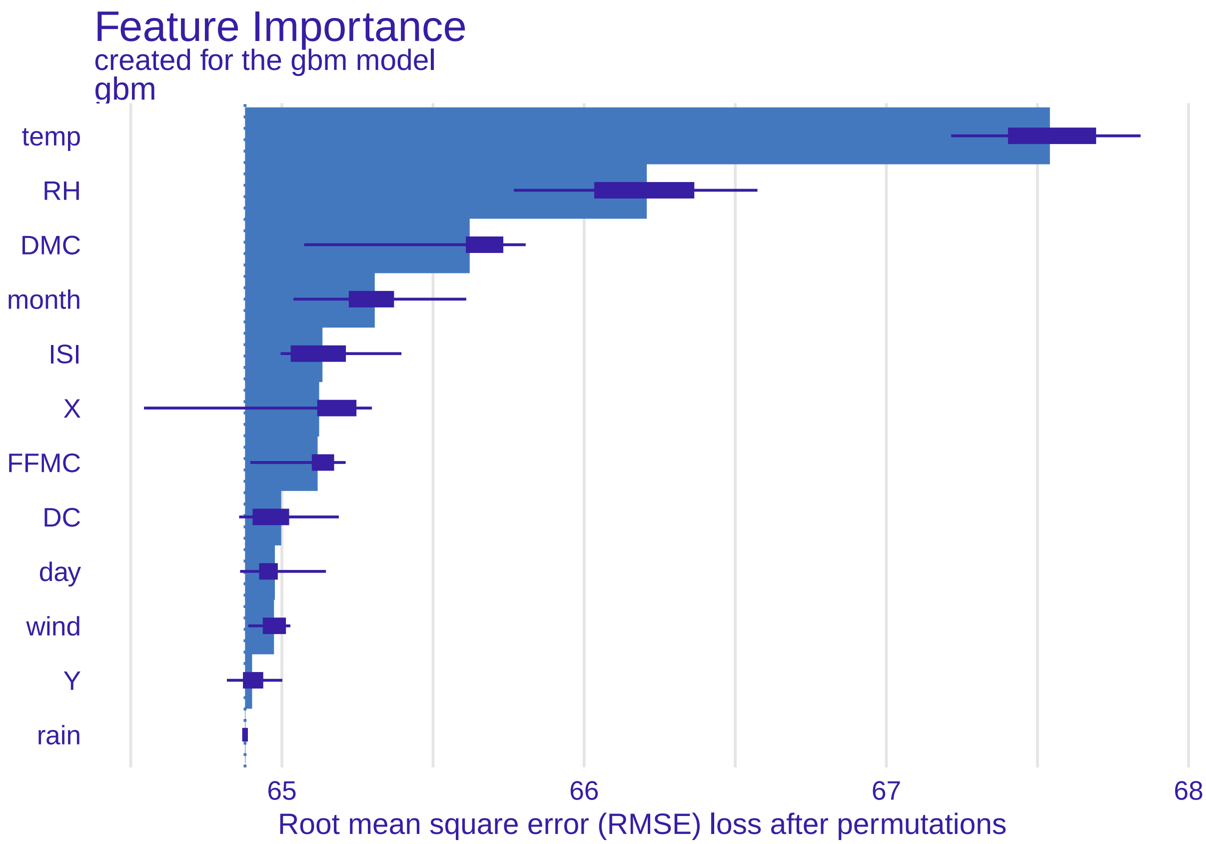
```
plot(xgb_var_imp)
```




```
plot(lgb_var_imp)
```




```
plot(gbm_var_imp)
```




**XGBoost Modeli:**

- 'temp' değişkeni, bu model için açıkça en önemli özellik olarak öne çıkıyor, diğer tüm değişkenlerden çok daha büyük bir öneme sahip.
- 'Y' ve 'FFMC' değişkenleri de önemli özellikler olarak belirginleşiyor, ancak 'temp' değişkeninin yaklaşık yarısı kadar etkiye sahipler.
- 'DMC', 'DC', 'RH', ve diğer değişkenler daha az öneme sahip ancak yine de modelin tahminleri üzerinde bir etkiye sahip oldukları görülüyor.

**LGB Modeli:**

- Bu modelde de 'temp' değişkeni en önemli özellik olarak belirlenmiş, ancak XGBoost modeline göre daha az belirgin bir farkla.
- 'Y' ve 'RH' özellikleri de bu modelde önemli olarak sıralanmış.
- 'X', 'ISI', 'wind' ve diğer değişkenlerin önemi göreceli olarak düşük.

**GBM Modeli:**

- 'temp' değişkeni, GBM modelinde de en önemli özellik olarak karşımıza çıkıyor ve açık bir farkla diğerlerinden öne çıkıyor.
- 'DMC', 'RH' ve 'month' de bu model için önemli özellikler.
- 'DC', 'FFMC', 'day' ve diğer değişkenler daha az önemli ancak hala model tahminlerine katkı sağlıyorlar.

**Genel Yorum:**

- Tüm modellerde 'temp' değişkeni en önemli özellik olarak sıralanmıştır, bu durum bu değişkenin tahminler üzerindeki güçlü ve tutarlı etkisini gösterir.
- Modelden modele değişken öneminde bazı farklılıklar olsa da, bazı değişkenler (örneğin 'temp', 'Y', 'RH') genel olarak önemli özellikler olarak sıralanmaktadır.

## ionosphpere

### decision tree BD & Shap

```
train_data[1,]
```

```
##     V1 V2 V3      V4 V5      V6 V7      V8 V9     V10 V11     V12 V13     V14
## 176  1  0  1 0.11765  1 0.23529  1 0.41176  1 0.05882   1 0.23529   1 0.11765
##     V15     V16 V17      V18 V19      V20 V21     V22 V23     V24 V25      V26
## 176   1 0.47059   1 -0.05882   1 -0.11765   1 0.35294   1 0.41176   1 -0.11765
##     V27     V28 V29     V30 V31     V32 V33     V34 V35
## 176   1 0.20225   1 0.05882   1 0.35294   1 0.23529   1
```

```
train_data[2,]
```

```
##     V1 V2      V3 V4      V5       V6      V7      V8      V9     V10     V11
## 235  1  0 0.68729  1 0.91973 -0.76087 0.81773 0.04348 0.76087 0.10702 0.86789
##         V12     V13     V14    V15     V16     V17      V18     V19     V20
## 235 0.73746 0.70067 0.18227 0.7592 0.13712 0.93478 -0.25084 0.70736 0.18729
##         V21     V22     V23     V24 V25      V26     V27     V28    V29     V30
## 235 0.64883 0.24582 0.60201 0.77425   1 -0.53846 0.89262 0.22216 0.7107 0.53846
##     V31      V32     V33     V34 V35
## 235   1 -0.06522 0.56522 0.23913   0
```

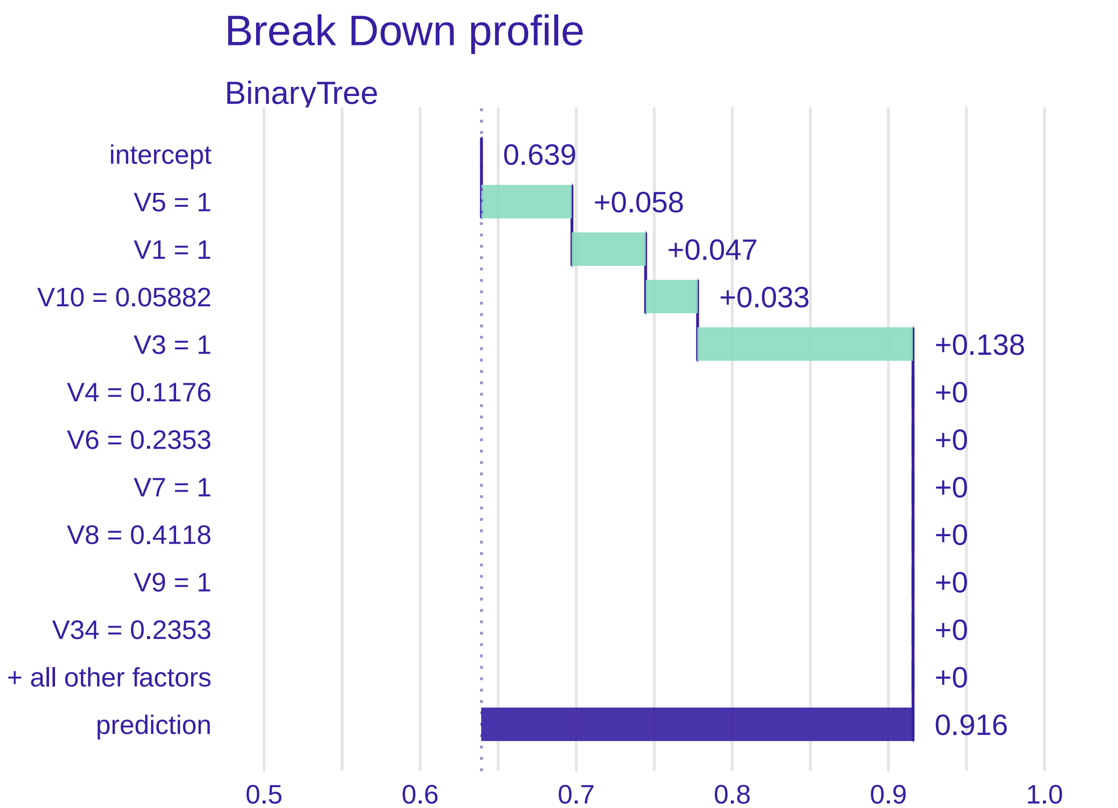
```
train_data[3,]
```

```
##     V1 V2  V3 V4      V5      V6     V7      V8      V9      V10     V11
## 261  1  0 0.5  0 0.38696 0.10435 0.4913 0.06522 0.46957 -0.03913 0.35652
##          V12     V13     V14     V15     V16     V17     V18     V19     V20
## 261 -0.12609 0.45652 0.04783 0.50435 0.02609 0.35652 0.19565 0.42174 0.14783
##         V21      V22     V23      V24     V25     V26     V27     V28     V29
## 261 0.42174 -0.02609 0.32174 -0.11304 0.47391 -0.0087 0.41789 0.06908 0.38696
##         V30     V31     V32     V33     V34 V35
## 261 0.03913 0.35217 0.14783 0.44783 0.17391   1
```

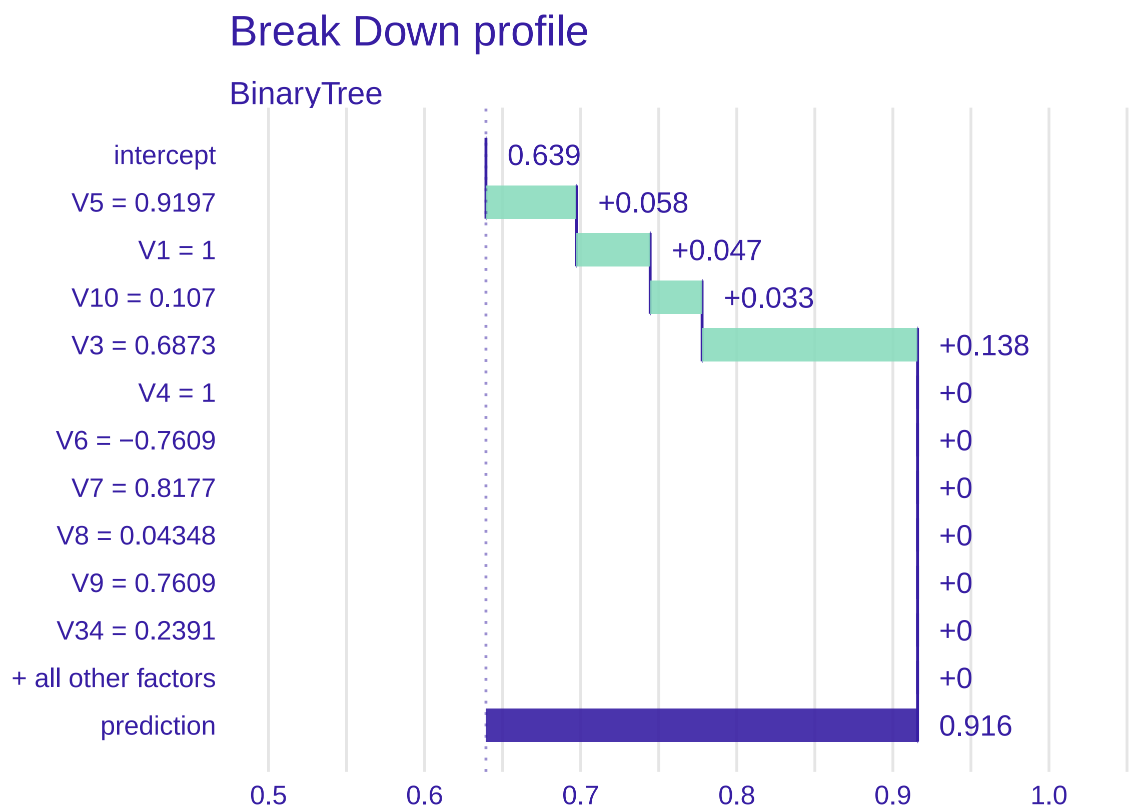
```
train_data[4,]
```

```
##     V1 V2      V3      V4      V5      V6      V7       V8      V9      V10
## 306  1  0 0.52542 -0.0339 0.94915 0.08475 0.52542 -0.16949 0.30508 -0.01695
##         V11      V12     V13     V14     V15      V16     V17     V18     V19
## 306 0.50847 -0.13559 0.64407 0.28814 0.83051 -0.35593 0.54237 0.01695 0.55932
##        V20     V21     V22     V23     V24     V25     V26     V27      V28
## 306 0.0339 0.59322 0.30508 0.86441 0.05085 0.40678 0.15254 0.67287 -0.00266
##         V29     V30     V31      V32     V33      V34 V35
## 306 0.66102 -0.0339 0.83051 -0.15254 0.76271 -0.10169   1
```

```
plot(bd_dt_1)
```

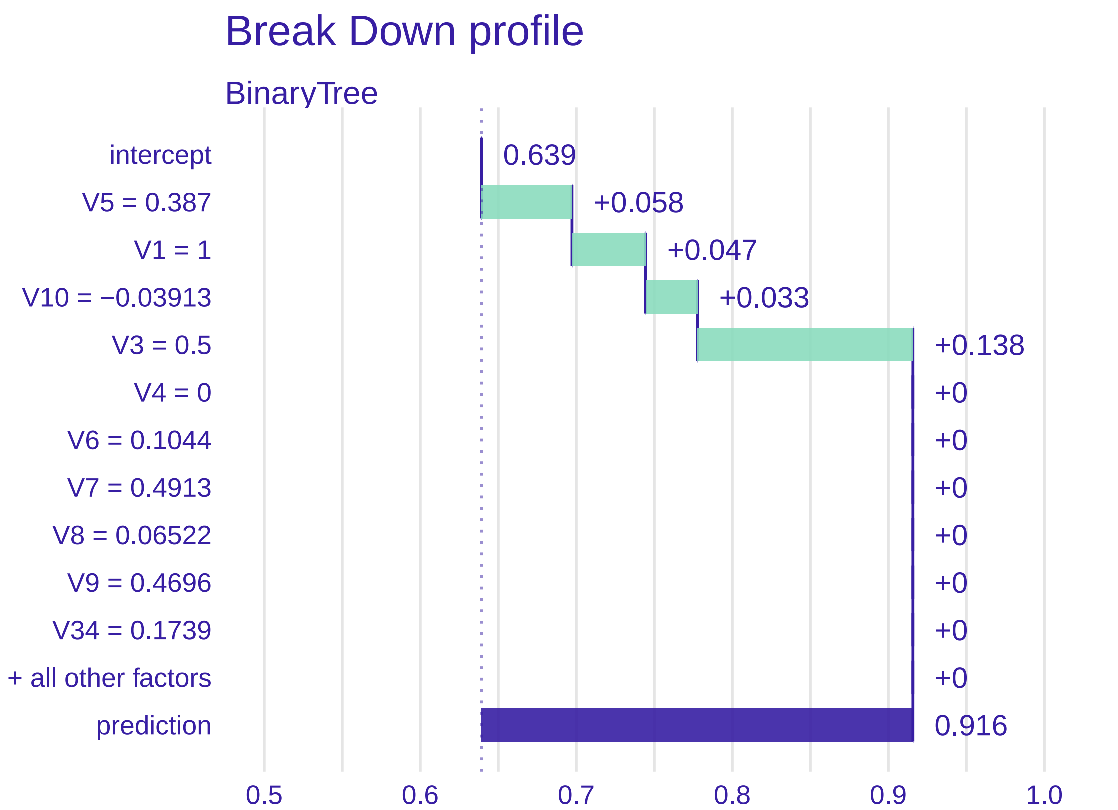



```
plot(bd_dt_2)
```

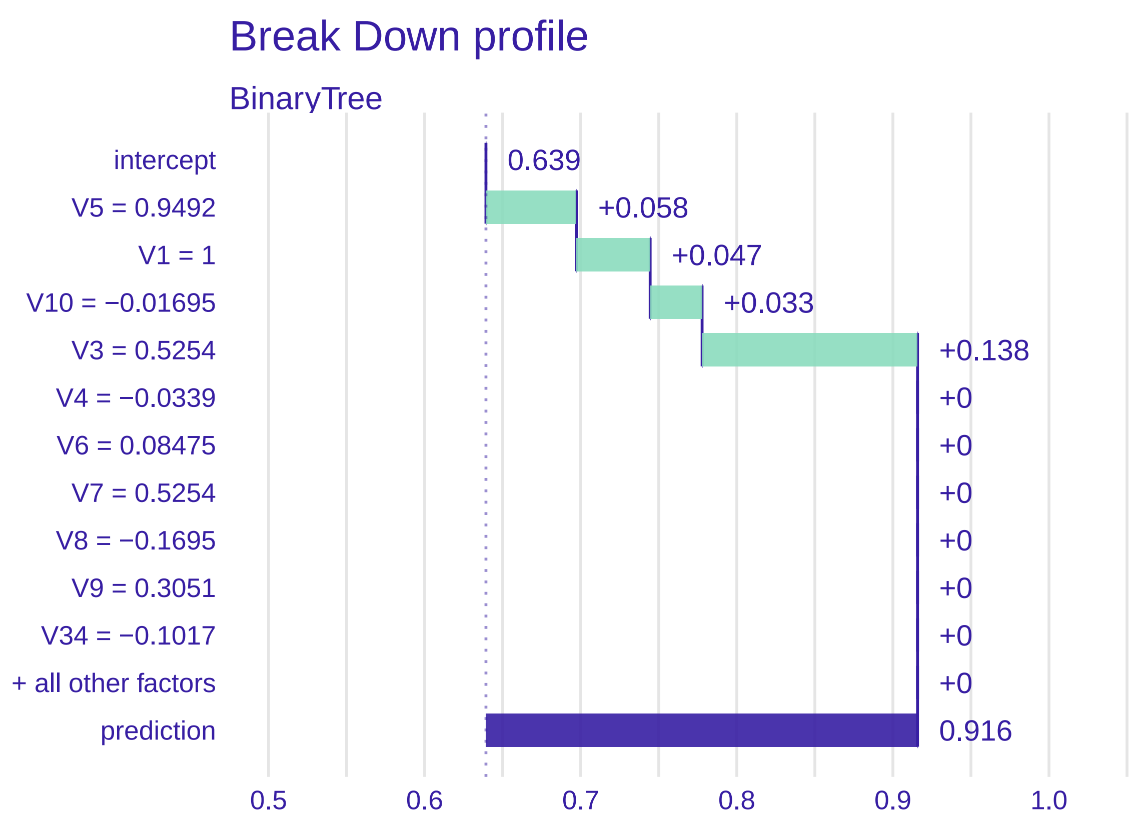


Break Down profile
BinaryTree
intercept
V5 = 0.9197
V1 = 1
V10 = 0.107
V3 = 0.6873
V4 = 1
V6 = −0.7609
V7 = 0.8177
V8 = 0.04348
V9 = 0.7609
V34 = 0.2391
+ all other factors
prediction
0.639
+0.058
+0.047
+0.033
+0.138
+0
+0
+0
+0
+0
+0
+0
0.916
0.5
0.6
0.7
0.8
0.9
1.0

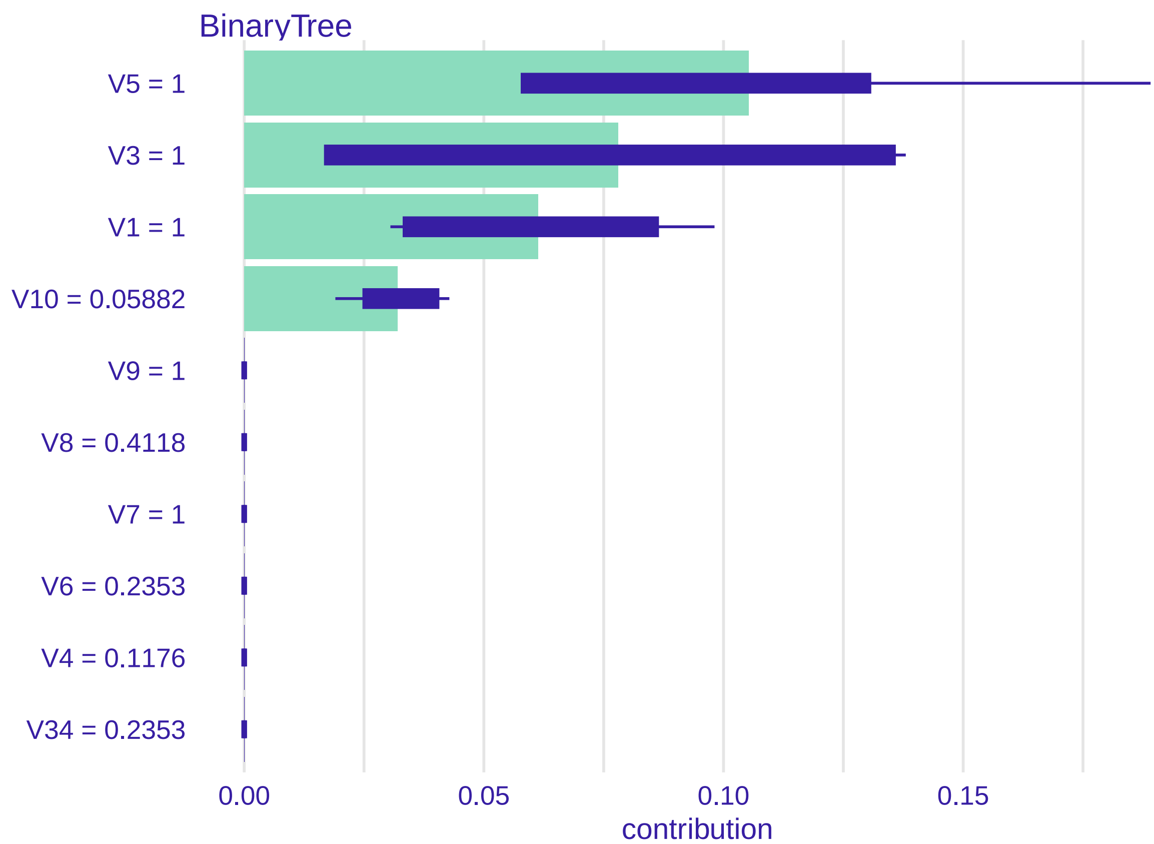

```
plot(bd_dt_3)
```


Break Down profile
BinaryTree
intercept
V5 = 0.387
V1 = 1
V10 = −0.03913
V3 = 0.5
V4 = 0
V6 = 0.1044
V7 = 0.4913
V8 = 0.06522
V9 = 0.4696
V34 = 0.1739
+ all other factors
prediction
0.639
+0.058
+0.047
+0.033
+0.138
+0
+0
+0
+0
+0
+0
+0
0.916
0.5
0.6
0.7
0.8
0.9
1.0

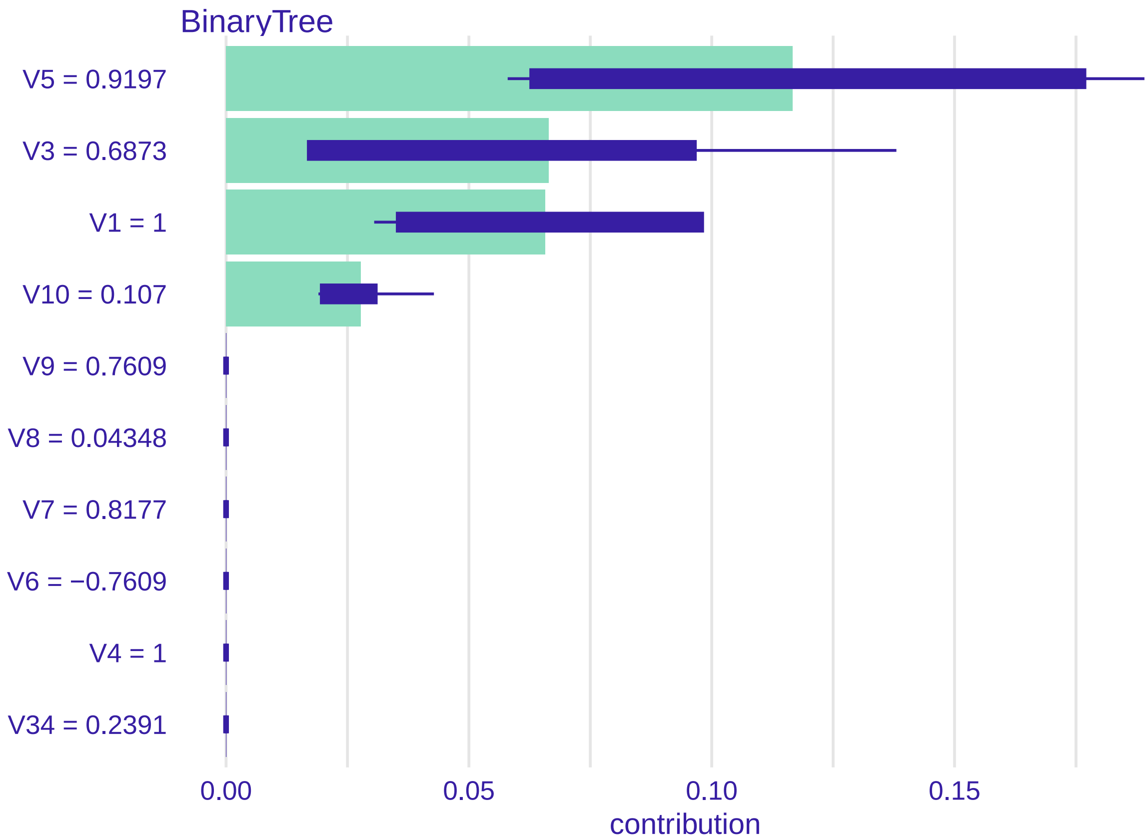
```
plot(bd_dt_4)
```



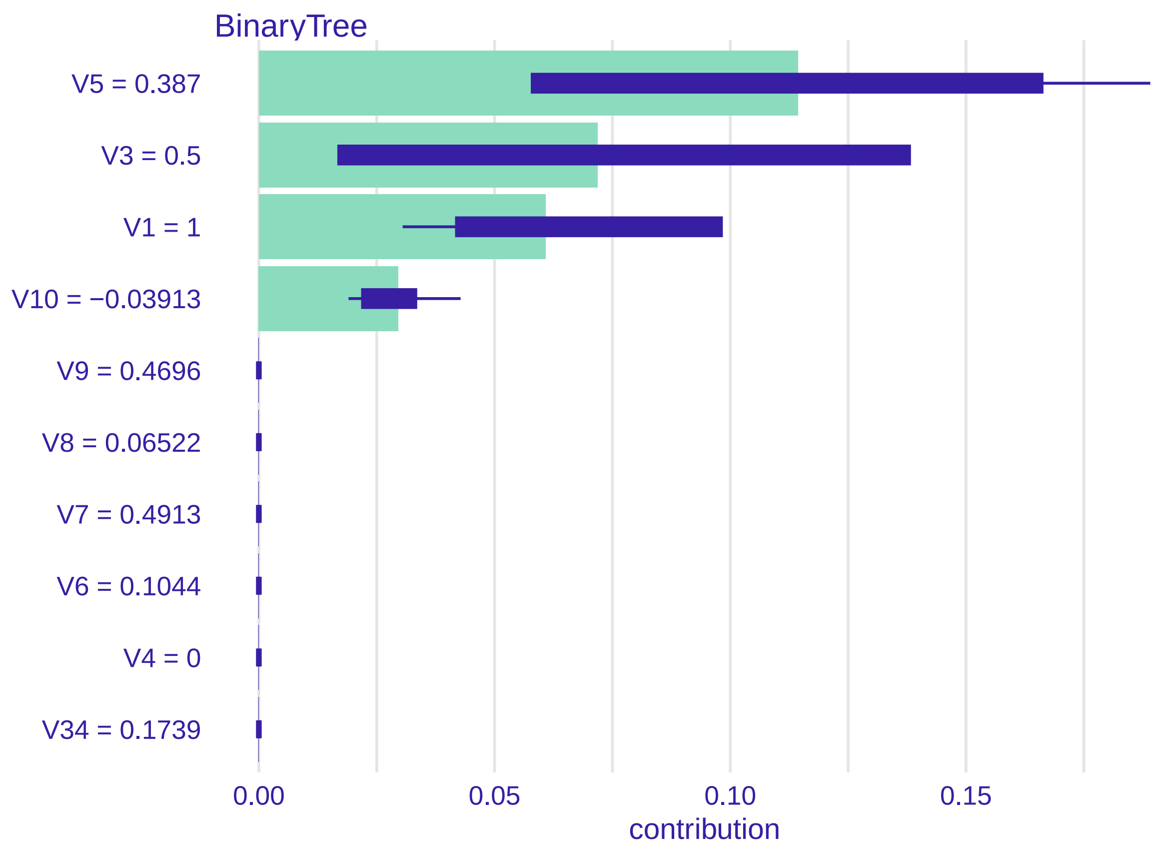

```
plot(shap_dt_1)
```



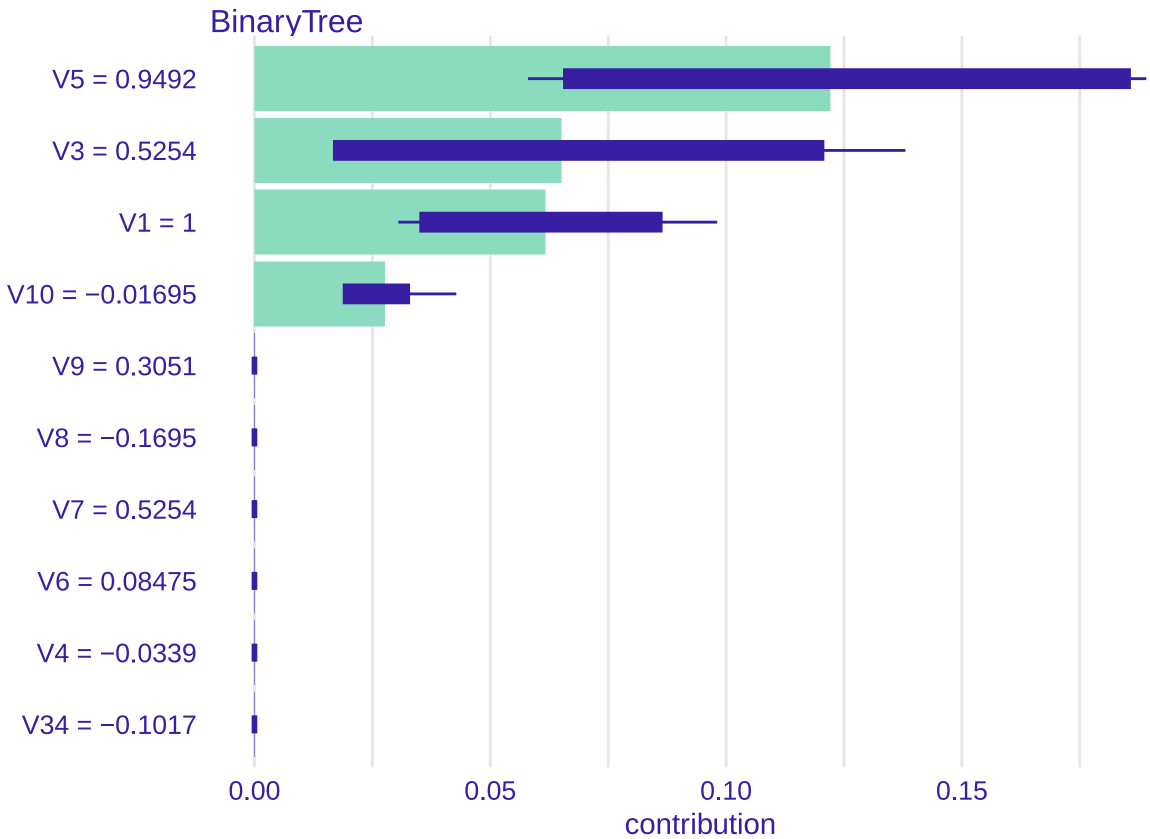

```
plot(shap_dt_2)
```




```
plot(shap_dt_3)
```

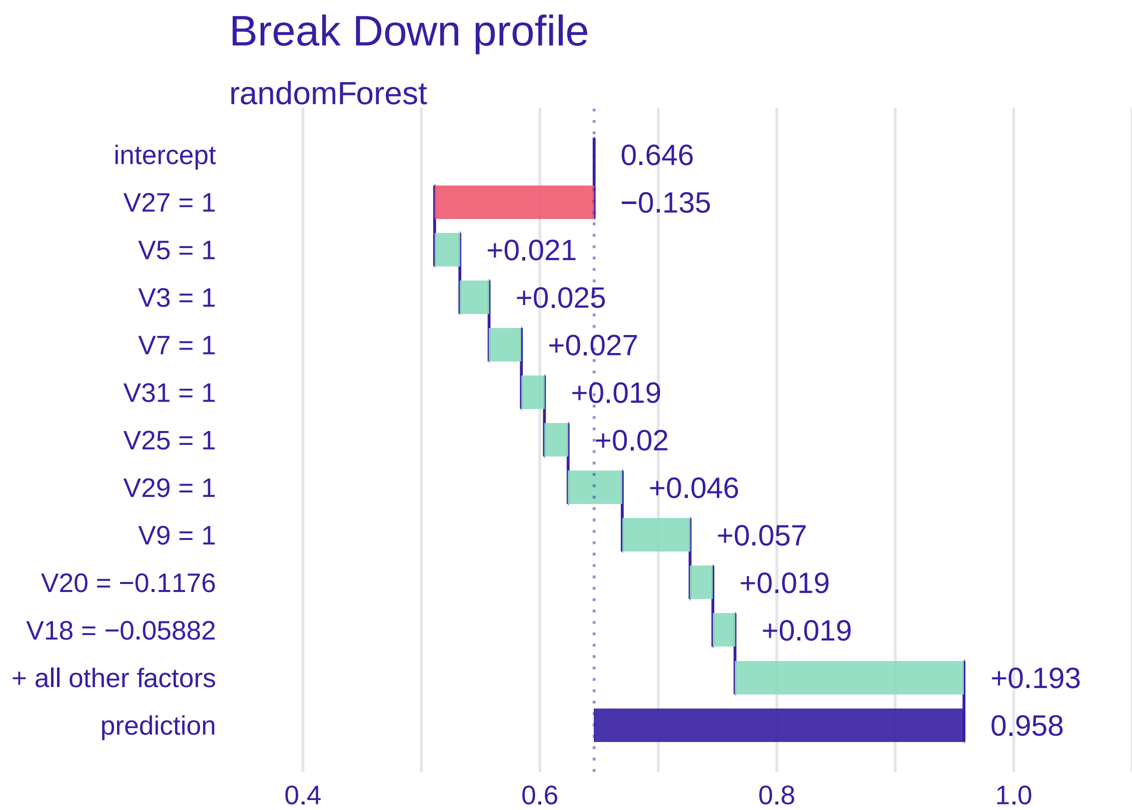



```
plot(shap_dt_4)
```

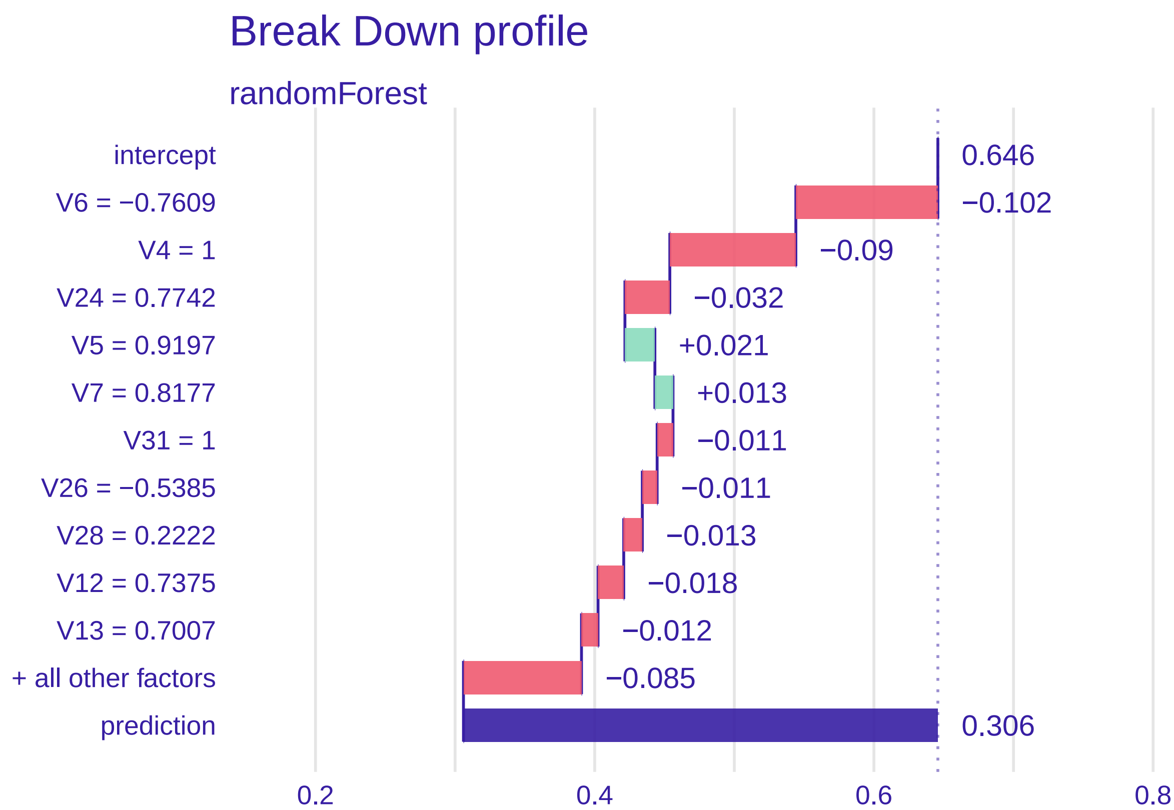



**random forest BD & Shap**

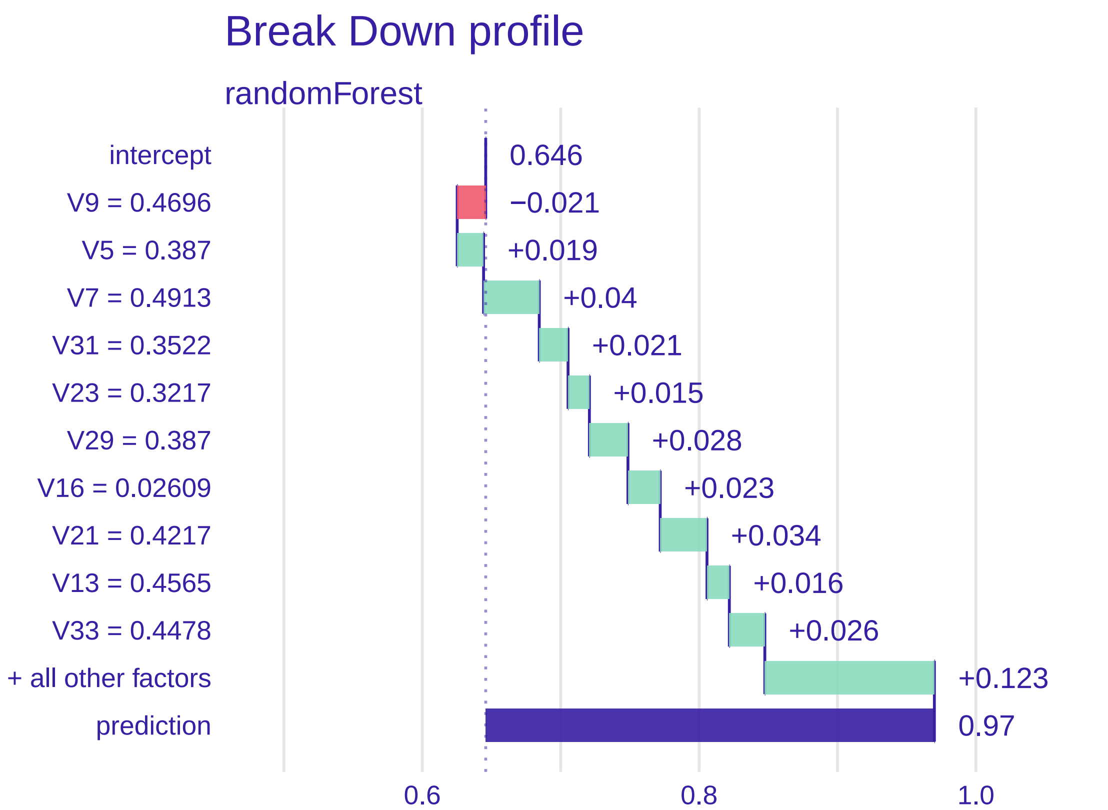
```
plot(bd_rf_1)
```


Break Down profile
randomForest
intercept
V27 = 1
V5 = 1
V3 = 1
V7 = 1
V31 = 1
V25 = 1
V29 = 1
V9 = 1
V20 = −0.1176
V18 = −0.05882
+ all other factors
prediction
0.646
−0.135
+0.021
+0.025
+0.027
+0.019
+0.02
+0.046
+0.057
+0.019
+0.019
+0.193
0.958
0.4
0.6
0.8
1.0

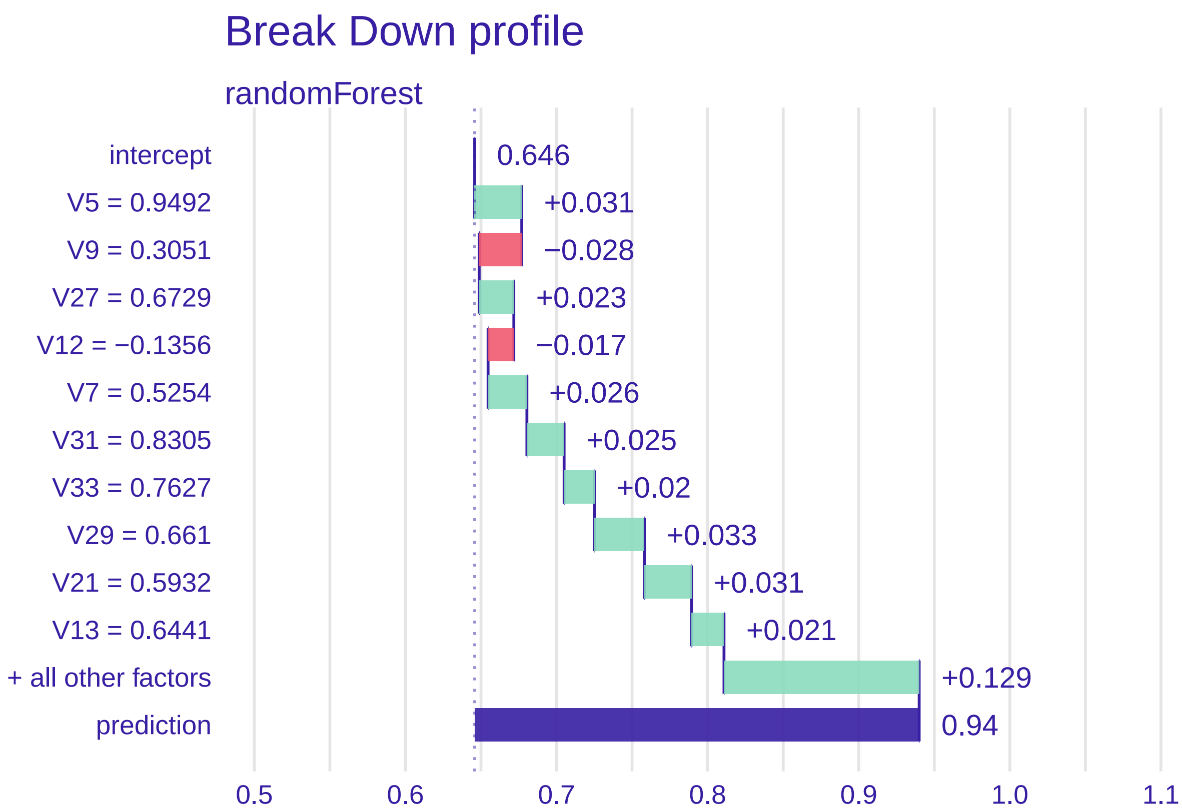
```
plot(bd_rf_2)
```


Break Down profile
randomForest
intercept
V6 = −0.7609
V4 = 1
V24 = 0.7742
V5 = 0.9197
V7 = 0.8177
V31 = 1
V26 = −0.5385
V28 = 0.2222
V12 = 0.7375
V13 = 0.7007
+ all other factors
prediction
0.646
−0.102
−0.09
−0.032
+0.021
+0.013
−0.011
−0.011
−0.013
−0.018
−0.012
−0.085
0.306
0.2
0.4
0.6
0.8

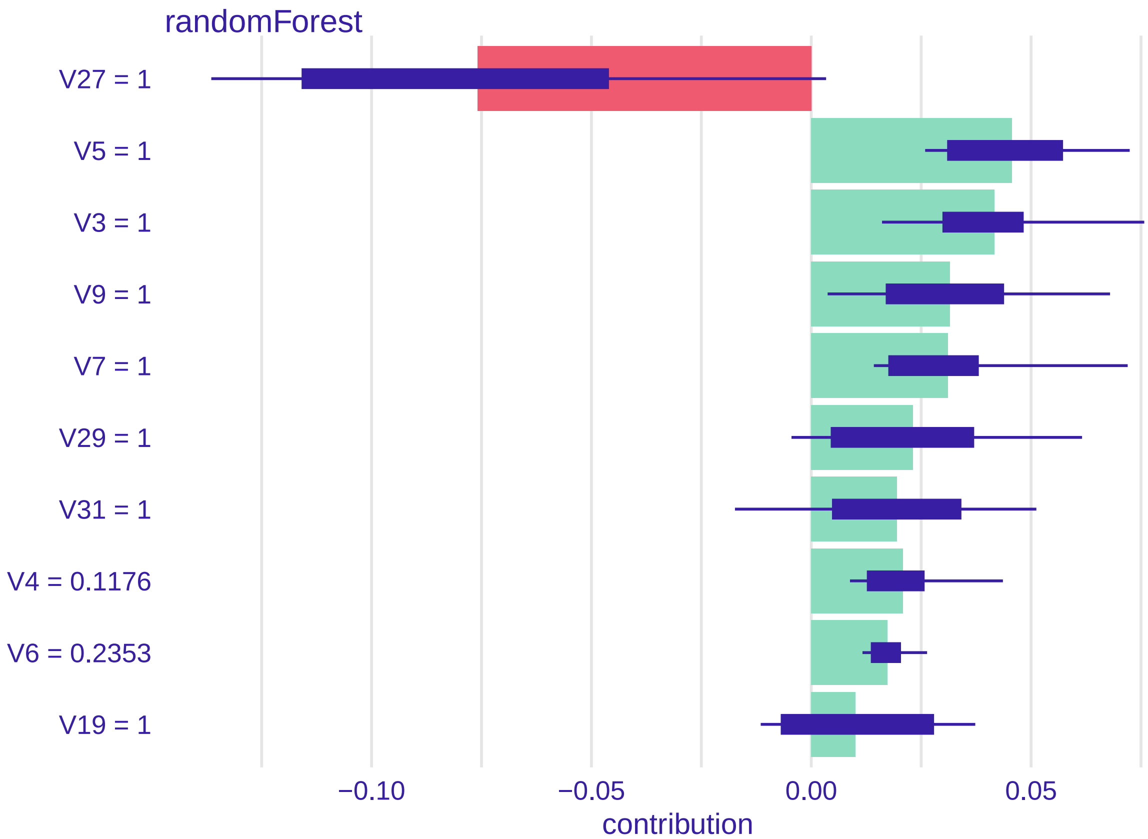
```
plot(bd_rf_3)
```


Break Down profile
randomForest
intercept
V9 = 0.4696
V5 = 0.387
V7 = 0.4913
V31 = 0.3522
V23 = 0.3217
V29 = 0.387
V16 = 0.02609
V21 = 0.4217
V13 = 0.4565
V33 = 0.4478
+ all other factors
prediction
0.646
−0.021
+0.019
+0.04
+0.021
+0.015
+0.028
+0.023
+0.034
+0.016
+0.026
+0.123
0.97
0.6
0.8
1.0

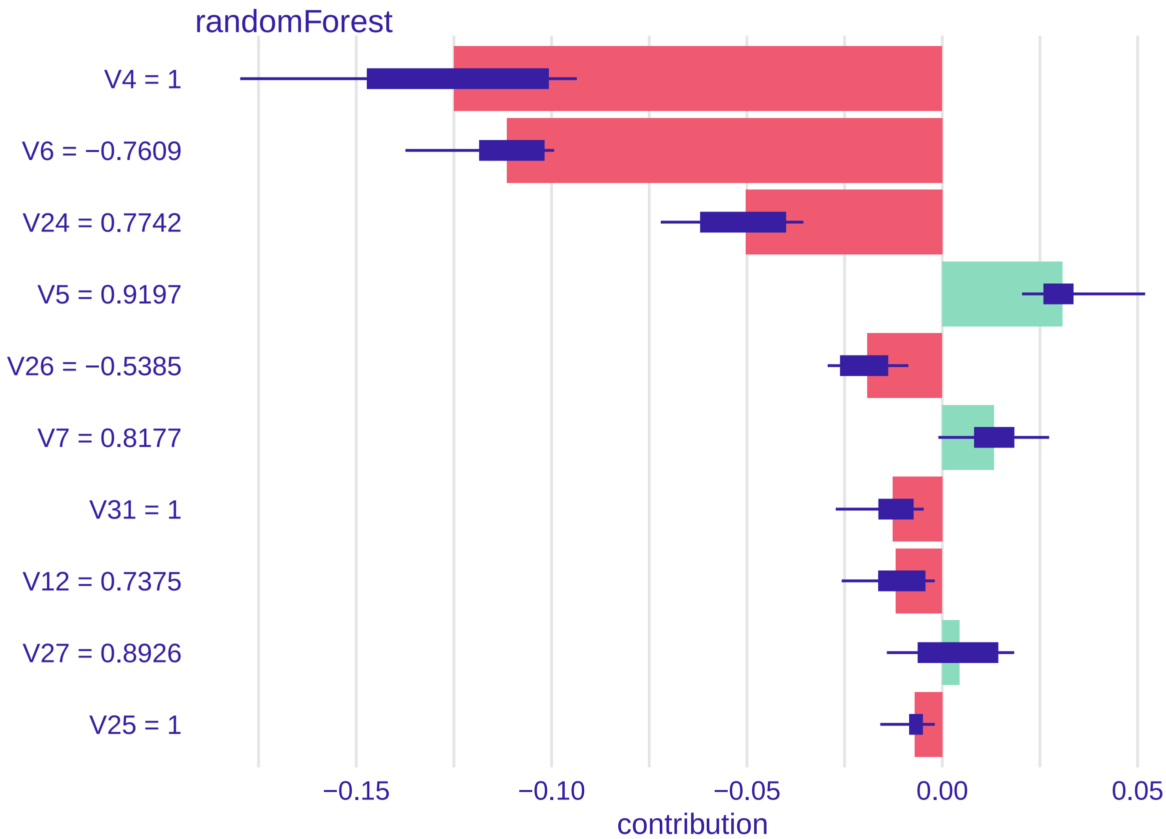
```
plot(bd_rf_4)
```



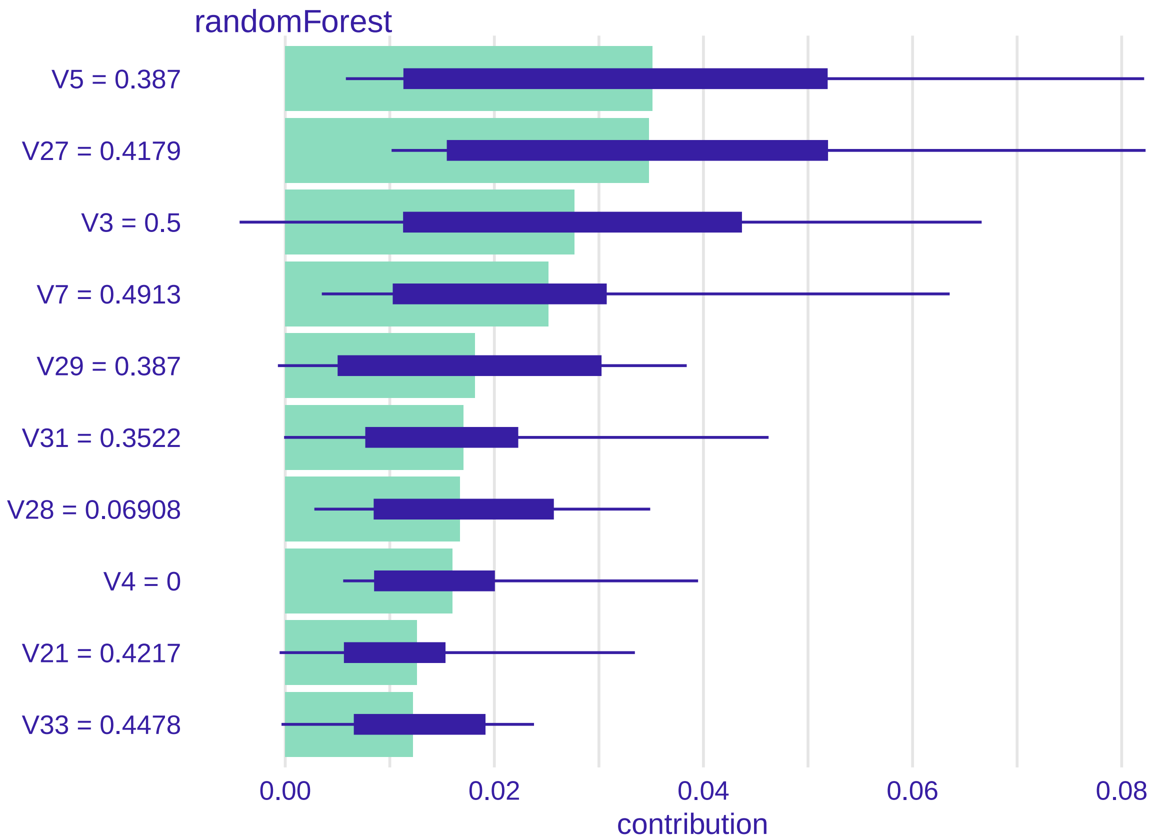

```
plot(shap_rf_1)
```



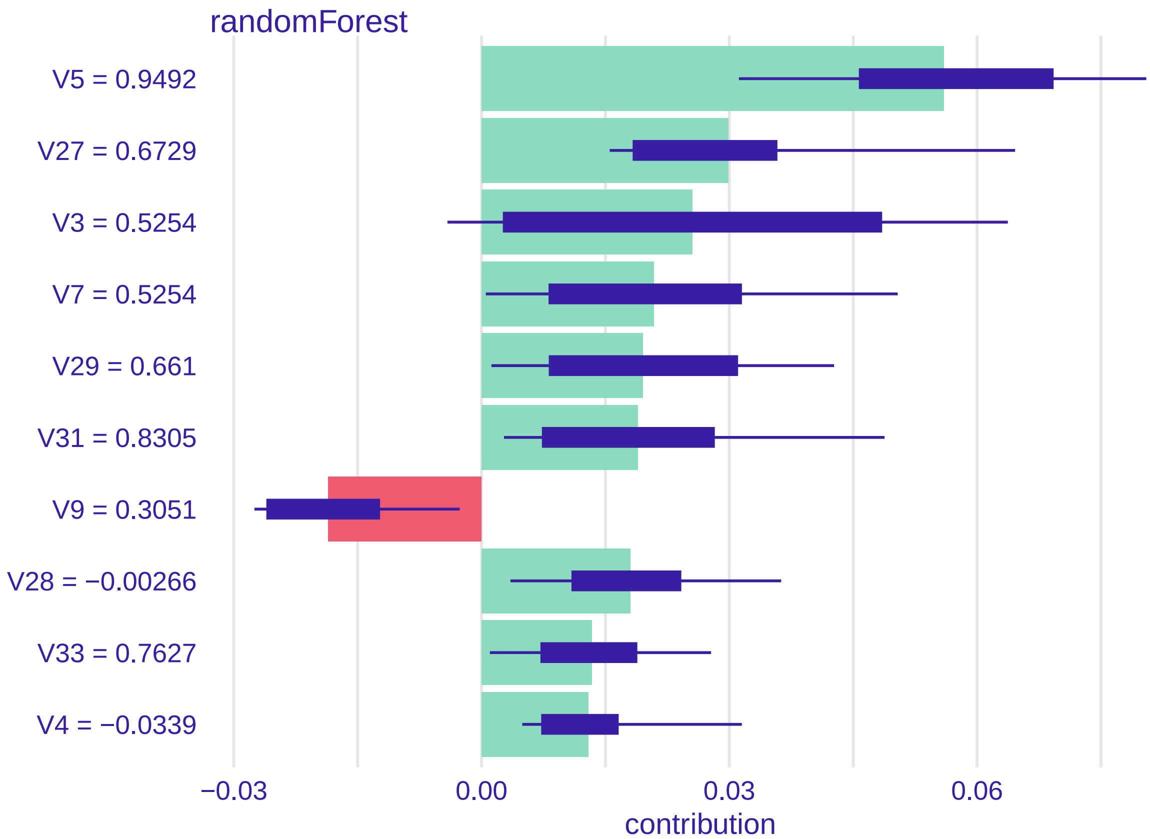

```
plot(shap_rf_2)
```




```
plot(shap_rf_3)
```

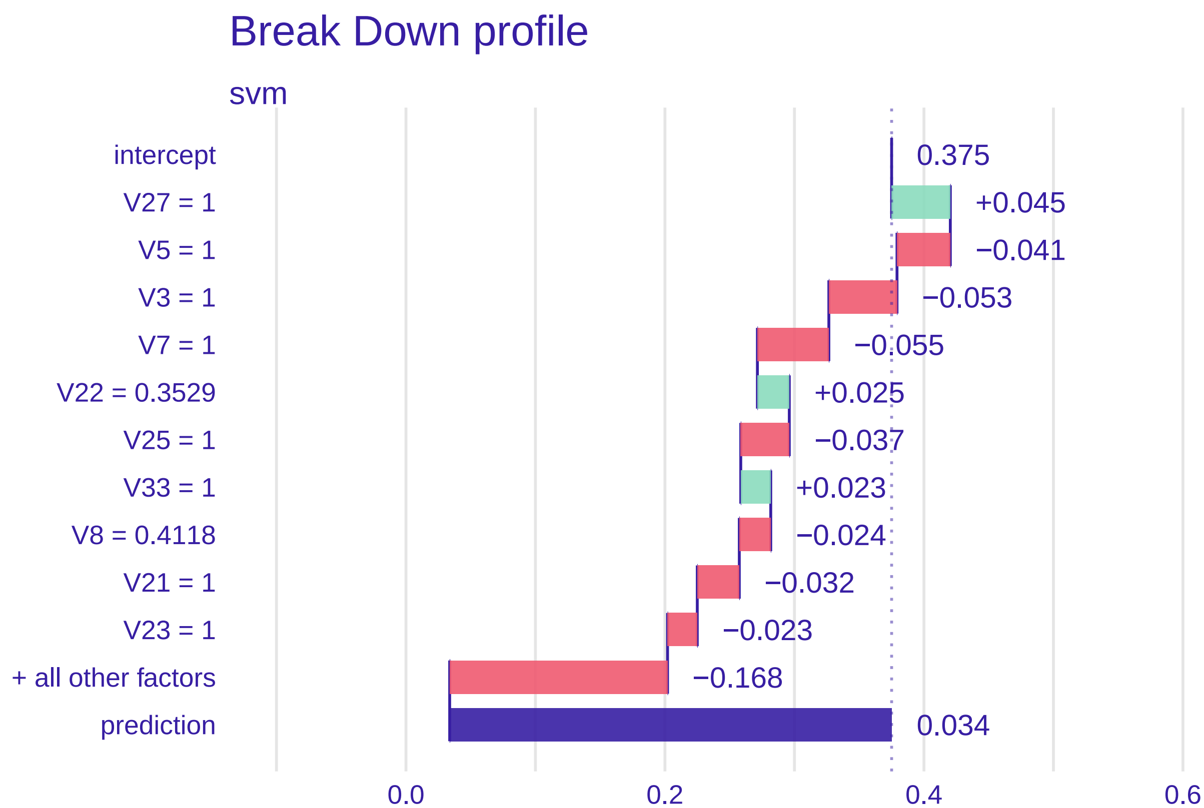



```
plot(shap_rf_4)
```

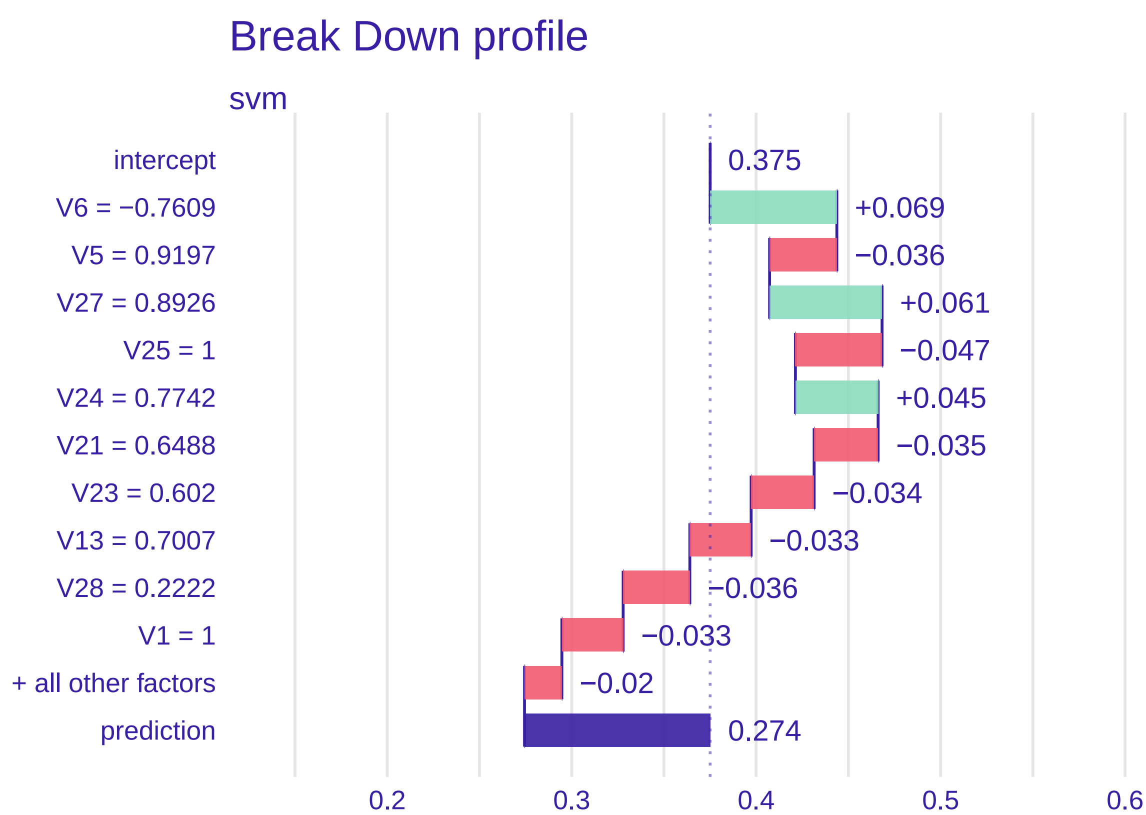



#### svm BD & Shap

```
plot(bd_svm_1)
```

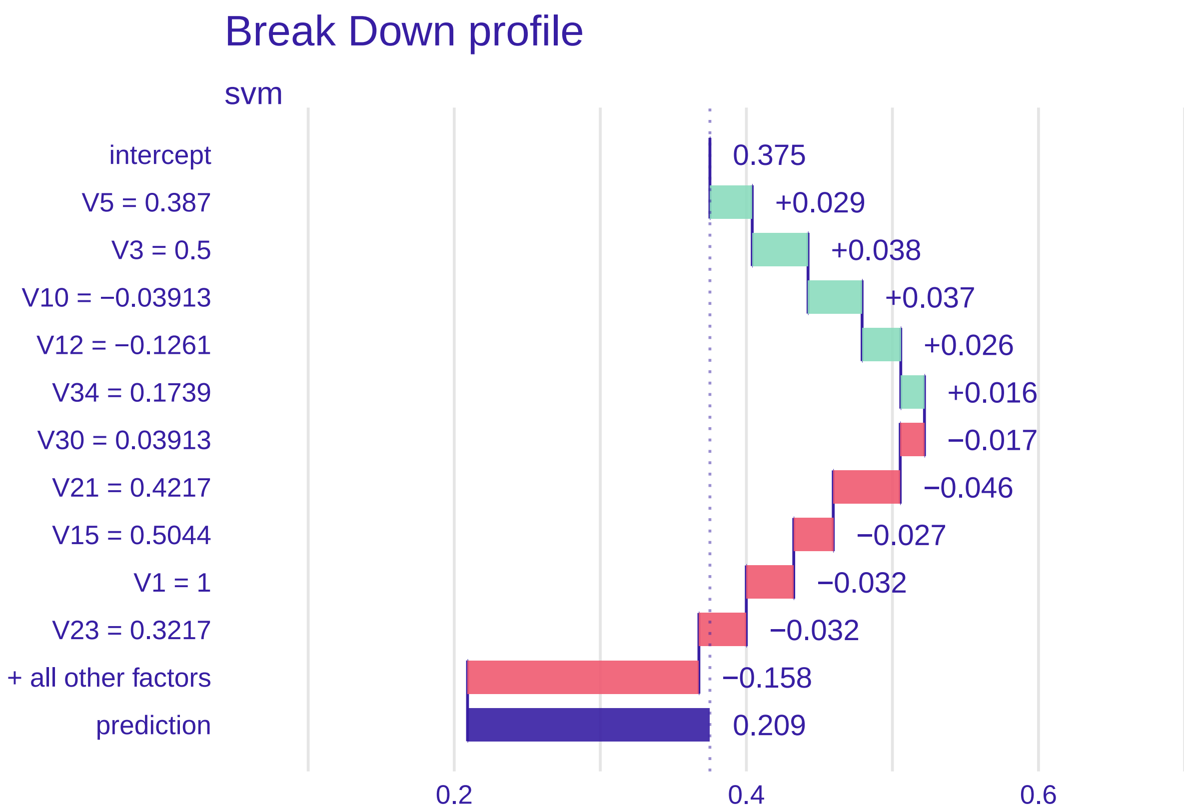

Break Down profile
svm
intercept
V27 = 1
V5 = 1
V3 = 1
V7 = 1
V22 = 0.3529
V25 = 1
V33 = 1
V8 = 0.4118
V21 = 1
V23 = 1
+ all other factors
prediction
0.375
+0.045
−0.041
−0.053
−0.055
+0.025
−0.037
+0.023
−0.024
−0.032
−0.023
−0.168
0.034
0.0
0.2
0.4
0.6

```
plot(bd_svm_2)
```

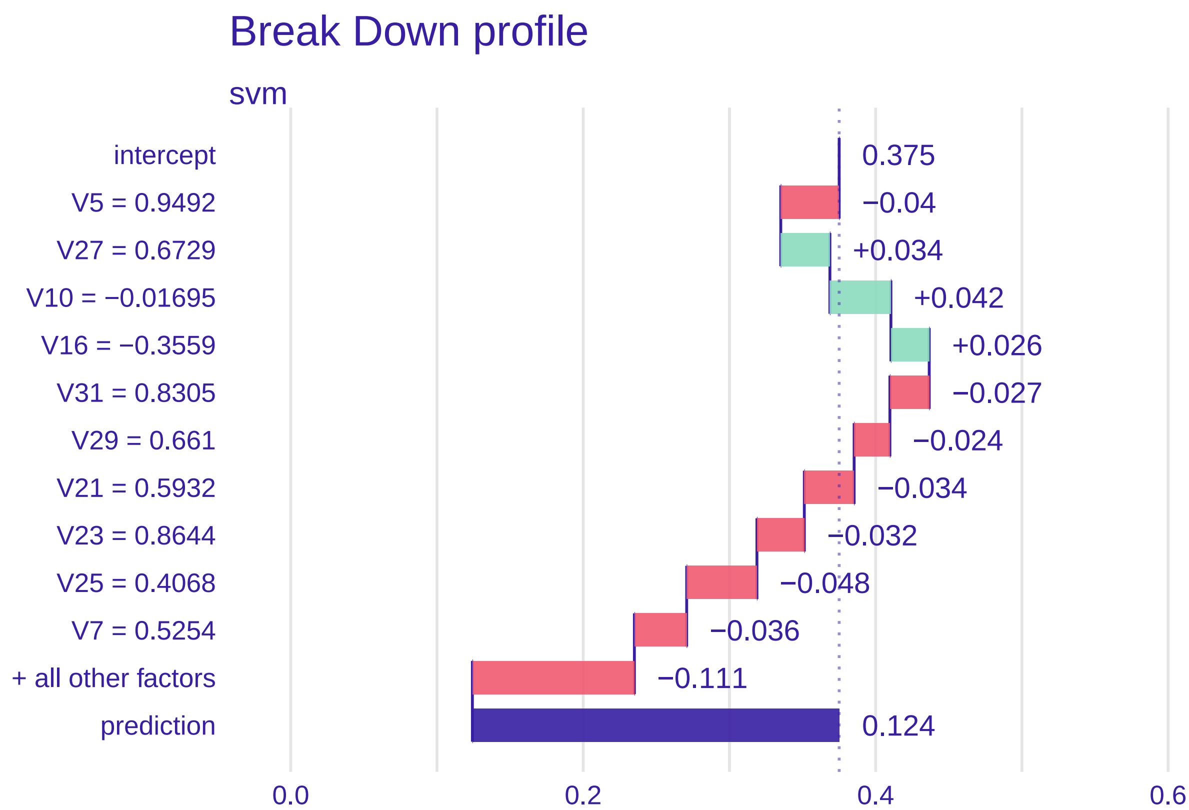


Break Down profile
svm
intercept
V6 = −0.7609
V5 = 0.9197
V27 = 0.8926
V25 = 1
V24 = 0.7742
V21 = 0.6488
V23 = 0.602
V13 = 0.7007
V28 = 0.2222
V1 = 1
+ all other factors
prediction
0.375
+0.069
−0.036
+0.061
−0.047
+0.045
−0.035
−0.034
−0.033
−0.036
−0.033
−0.02
0.274
0.2
0.3
0.4
0.5
0.6

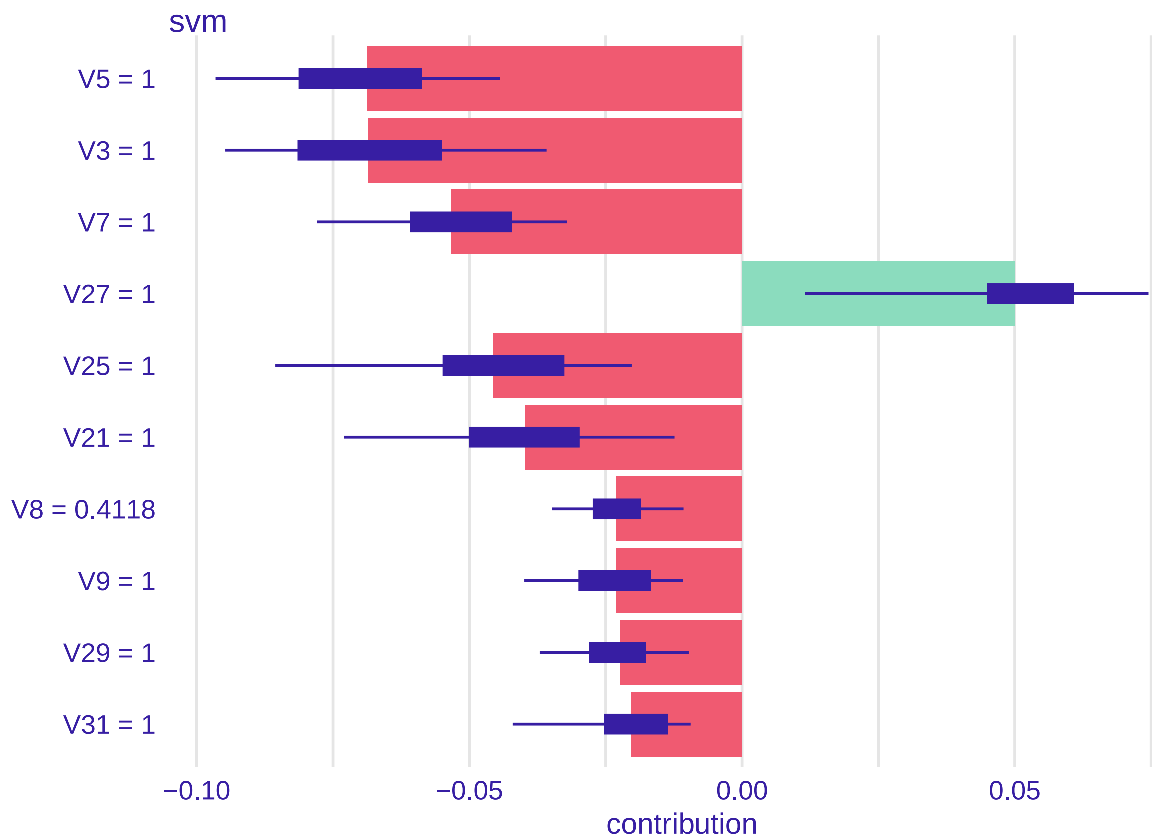

```
plot(bd_svm_3)
```


Break Down profile
svm
intercept
V5 = 0.387
V3 = 0.5
V10 = −0.03913
V12 = −0.1261
V34 = 0.1739
V30 = 0.03913
V21 = 0.4217
V15 = 0.5044
V1 = 1
V23 = 0.3217
+ all other factors
prediction
0.375
+0.029
+0.038
+0.037
+0.026
+0.016
−0.017
−0.046
−0.027
−0.032
−0.032
−0.158
0.209
0.2
0.4
0.6

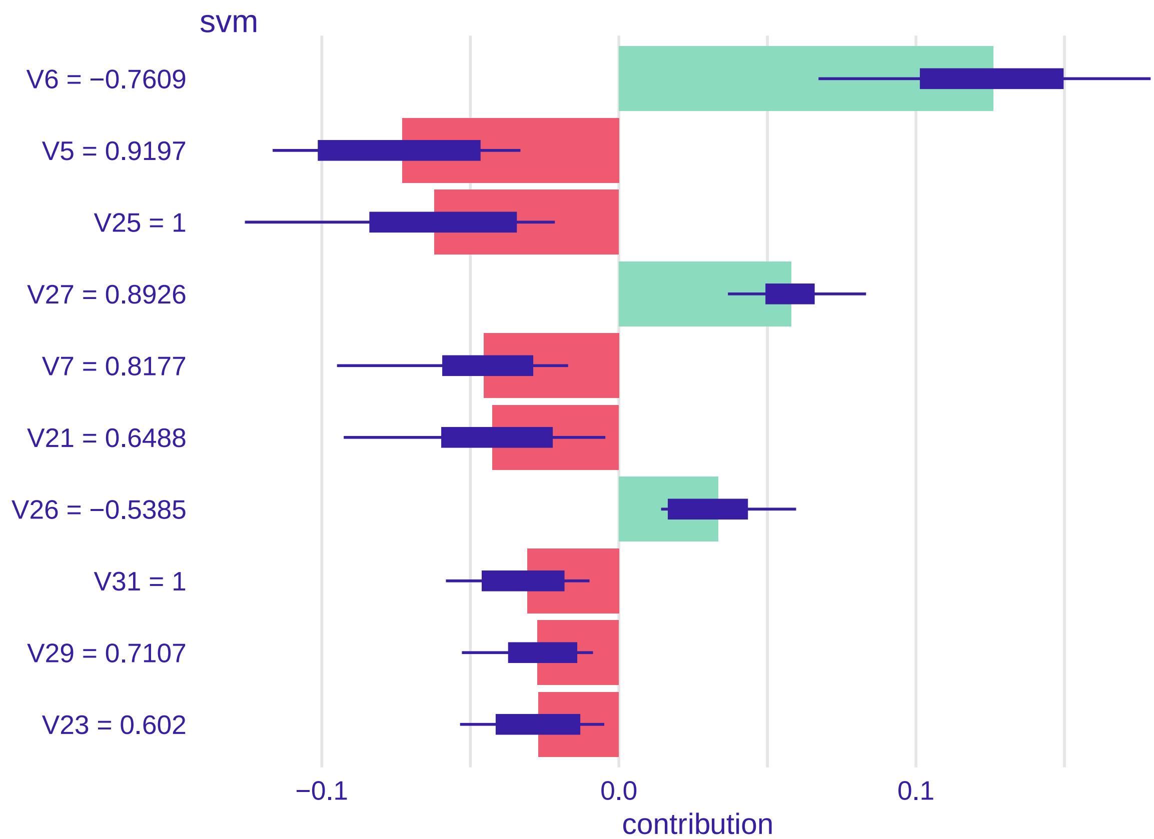
```
plot(bd_svm_4)
```



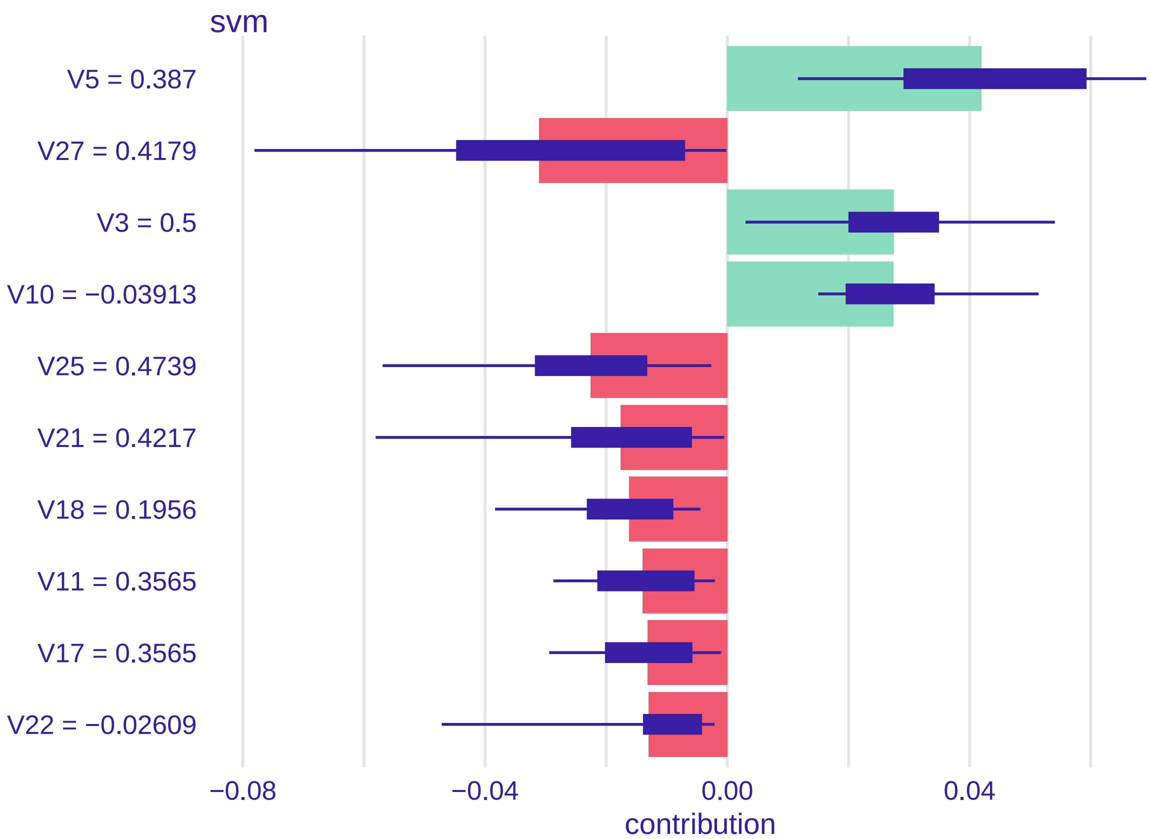

```
plot(shap_svm_1)
```



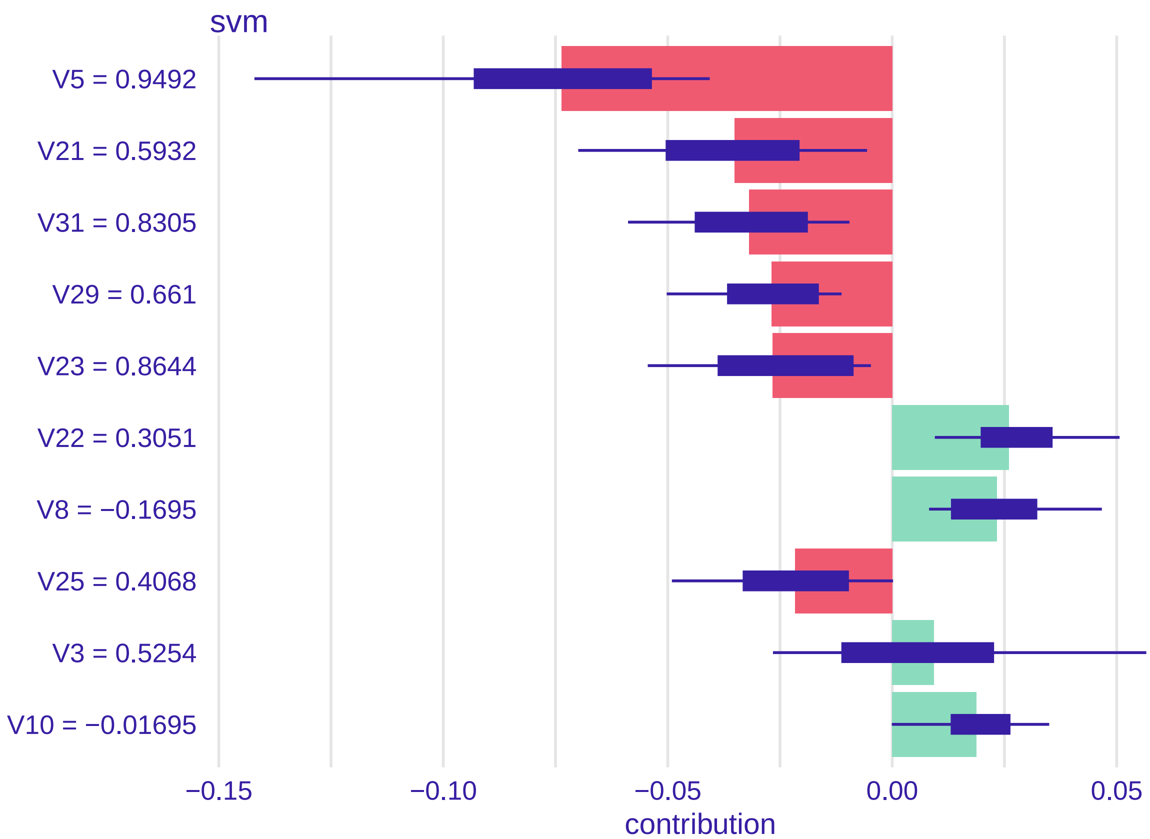

```
plot(shap_svm_2)
```




```
plot(shap_svm_3)
```

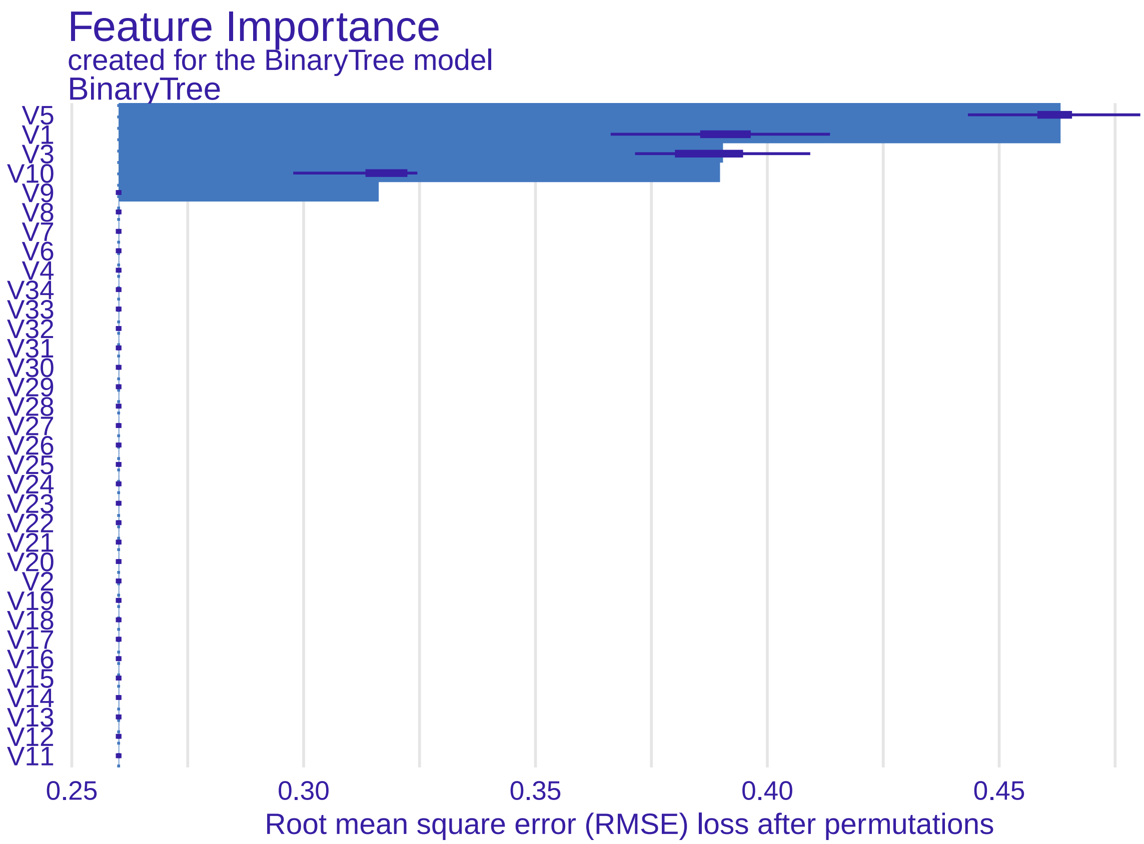



```
plot(shap_svm_4)
```

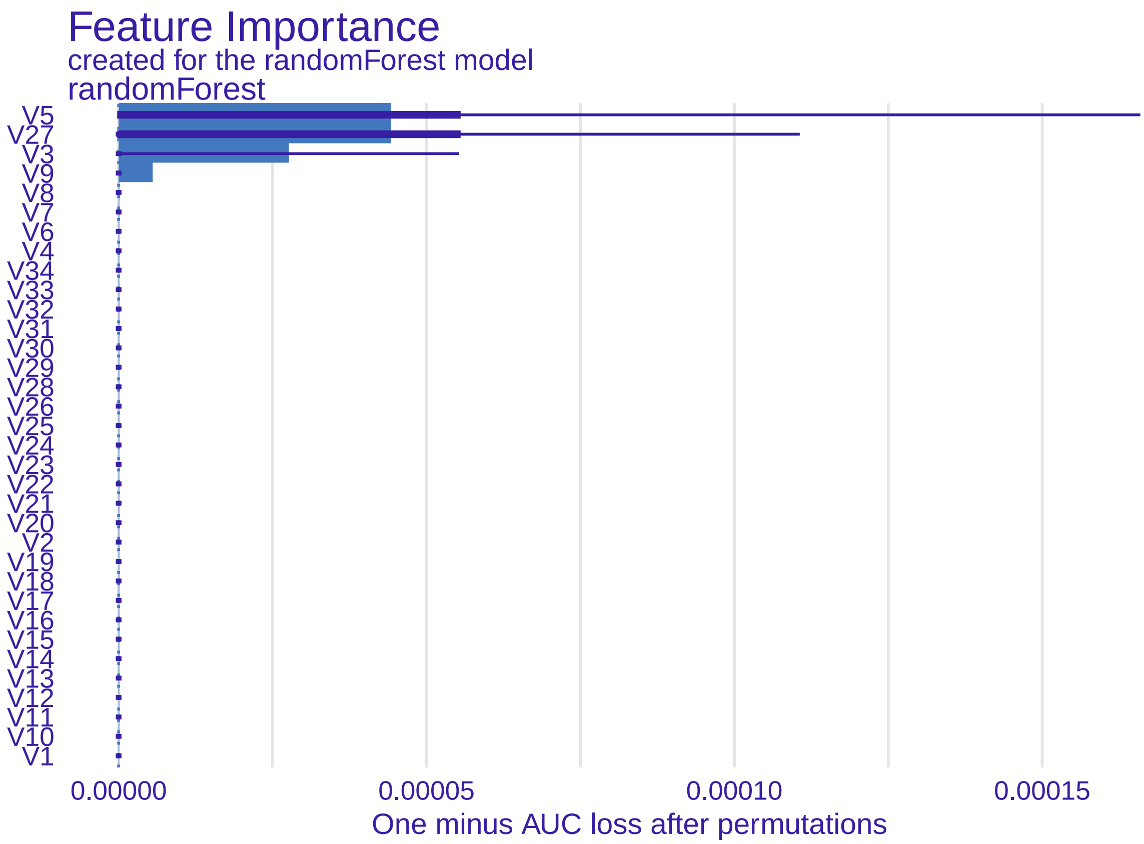



**Variable Importance**

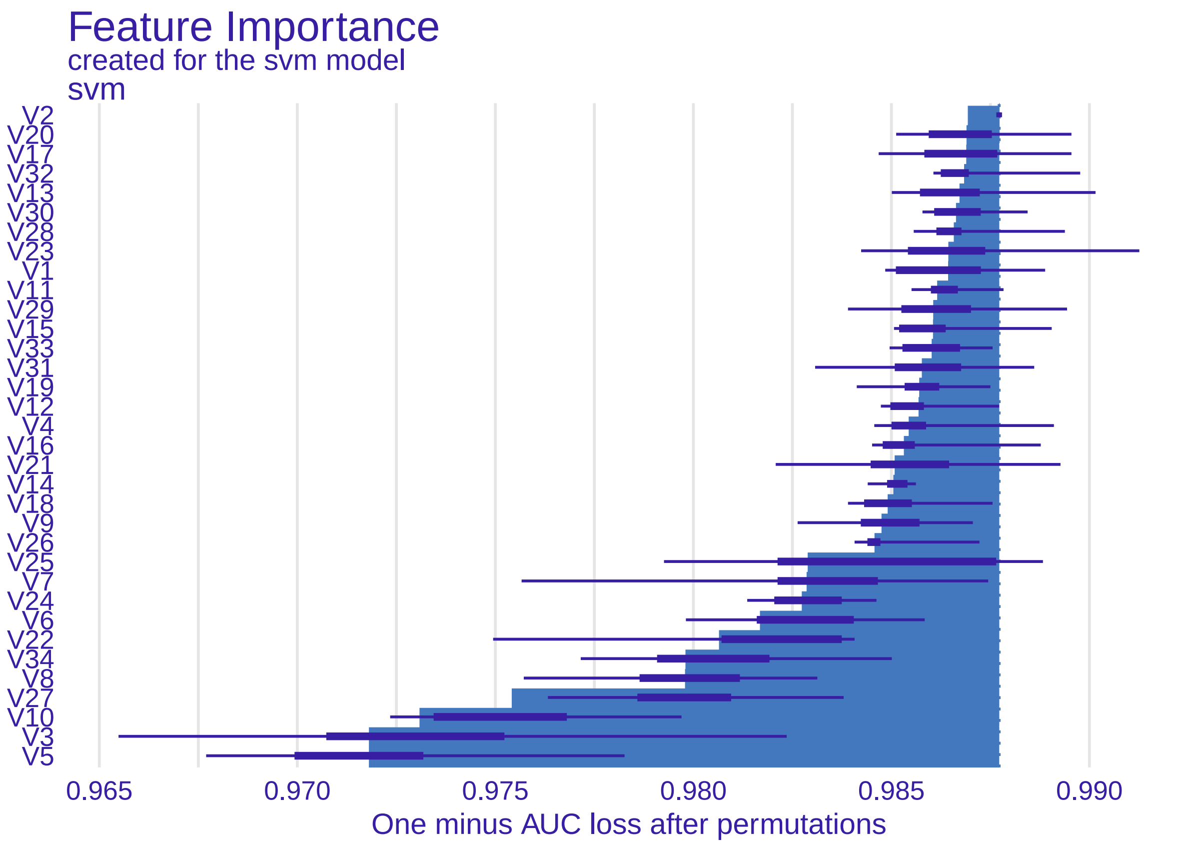
```
plot(dt_var_imp)
```




```
plot(rf_var_imp)
```




```
plot(svm_var_imp)
```




**DT Modeli:**

- 'V5', 'V3' ve 'V1' özellikleri, modelin RMSE'sini en çok etkileyen özellikler olarak sıralanmıştır. Bu, bu özelliklerin değiştirilmesi veya karıştırılması durumunda modelin performansının önemli ölçüde kötüleşeceği anlamına gelir.
- Orta seviyede öneme sahip diğer özellikler ('V9', 'V8', 'V7', vb.) de var ve bunlar da modelin performansını etkiler, ancak en üstteki özellikler kadar belirgin değil.
- En alttaki özellikler ('V10', 'V11', 'V12', vb.), modelin RMSE'sini etkileme konusunda en az öneme sahip özelliklerdir.

**RandomForest Modeli:**

- 'V5', 'V3' ve 'V7' özellikleri, AUC kaybını en çok etkileyen özellikler olarak sıralanmıştır. Bu özellikler, modelin performansı üzerinde önemli bir etkiye sahip olduğunu gösterir.
- Özelliklerin çoğu, modele göre oldukça düşük bir öneme sahip görünüyor, bu da RandomForest modelinin özellik seçiminde daha yaygın bir etki dağılımına sahip olabileceğine işaret ediyor.

**SVM Modeli:**

- SVM modeli için 'V2', 'V18', ve 'V28' özellikleri en önemli özellikler olarak sıralanmıştır.
- Bu model, diğer iki modelden farklı olarak, özelliklerin çok geniş bir önem spektrumu gösterdiği için özellik seçiminde daha belirgin bir farklılığa sahip.
- Özellik öneminin daha geniş bir dağılımı olduğunu ve pek çok özelliğin modelin AUC performansını etkilediğini görebiliyoruz.

**Genel Yorum:**

- Üç model arasında, en önemli özelliklerin sıralamasında bazı tutarlılıklar (örneğin, ‘V5’ ve ‘V3’ özelliklerinin yüksek önemi) ve bazı farklılıklar (örneğin, RandomForest modelinde ‘V7’ özelliğinin yüksek önemi ve SVM modelinde ‘V18’ özelliğinin önemi) bulunmaktadır.