RashPy

2025-06-29

Data importing

Partial Dependence Profile Explanations

Partial dependence profiles (PDPs) allow for a comparative examination of how a model outputs to a specific variable across different subgroups. These profiles show how a selected variable influences the model's output using group-level average predictions. For example, when the relationship between the model's predicted `scores` and the `age` variable is plotted separately for `Weekday` and `Weekend` groups, it becomes possible to visualize how this relationship varies by group. Parallel curves suggest that the model predicts similar `scores` across groups as `age` changes, while vertical separations between the curves or differences in their slope and shape indicate that the model responds differently to `age` depending on the group, revealing distinct patterns in how `scores` vary with `age`.

In the following sections, the PDPs of the trained model on the intended response variable such as `PHQ2`, `GAD2`, and `PHQ4` are given. These profiles are compared the independent variables `education`, `sex`, `weekday_weekend`.

2.1. Relationships by part of the week

2.1.1. The age vs. PHQ2

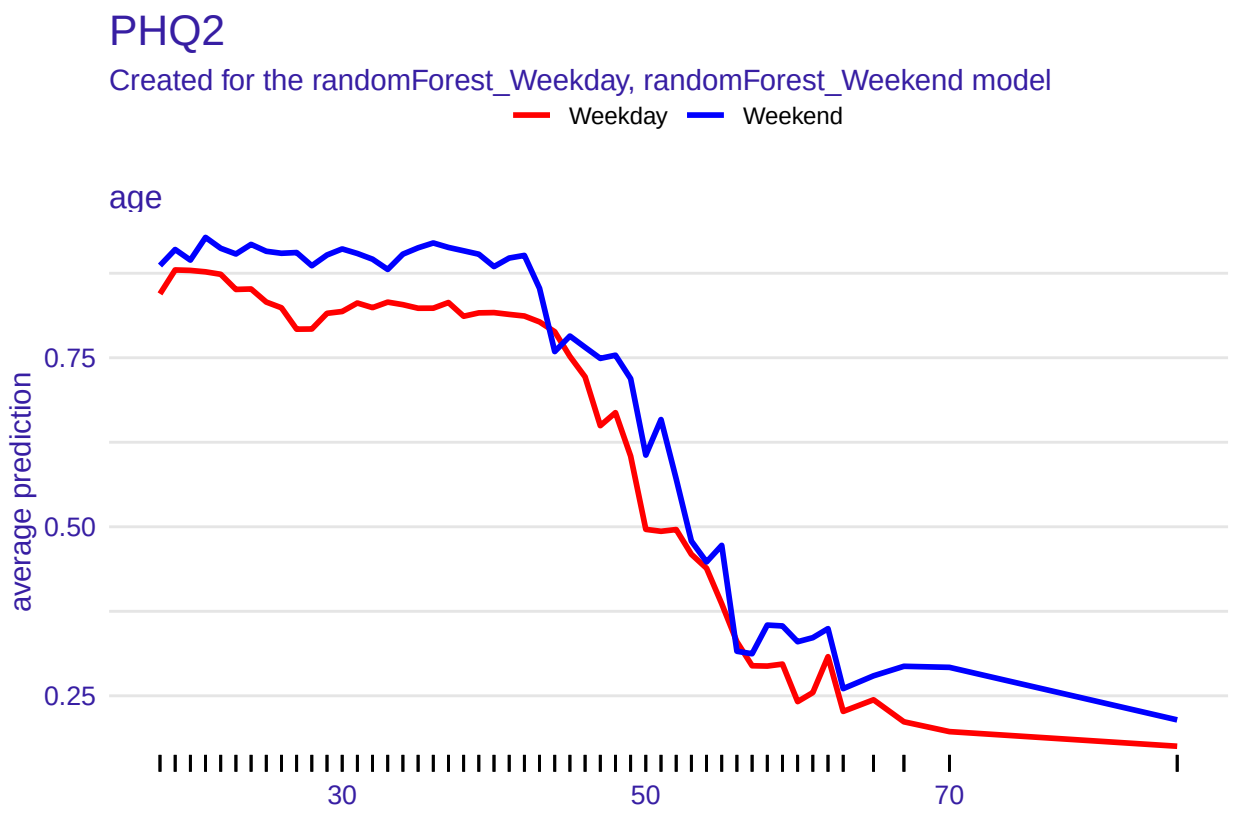
```
library(DALEX)
library(randomForest)
library(modelStudio)

model <- randomForest(factor(PHQ2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$PHQ2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.004 , mean =  0.749663 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.1779358 , max =  0.96
##   A new explainer has been created!
```

```
p1 <- plot(model_profile(modelex, variables = "age", groups = "weekday_weekend")) + ggtitle("PHQ2") +
  geom_rug(sides = "b")  +
  scale_color_manual(labels = c("randomForest_Weekday" = "Weekday",
                                "randomForest_Weekend" = "Weekend"),
                     values = c("randomForest_Weekday" = "red",
```

```
                                  "randomForest_Weekend" = "blue"),
                       name   = "")
p1
```



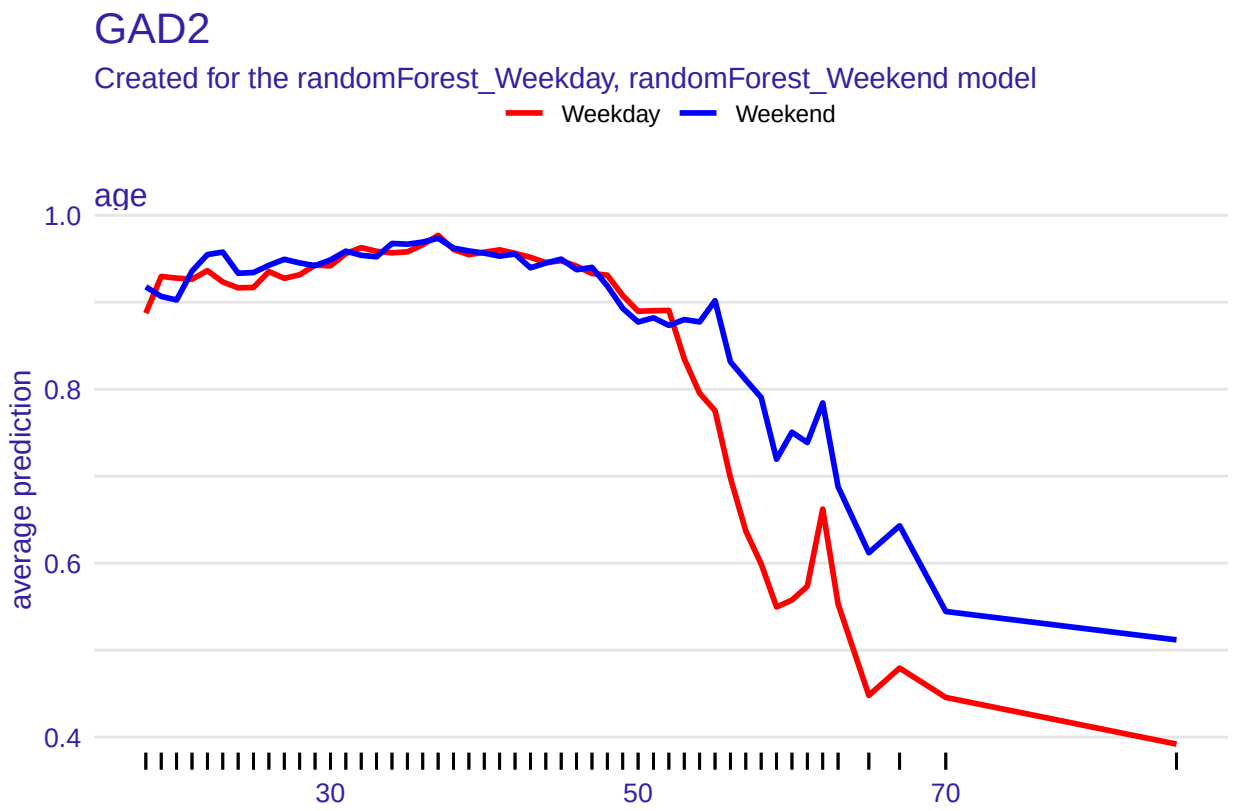
2.1.2. The age vs. GAD2

```
model <- randomForest(factor(GAD2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$GAD2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.018 , mean =  0.8790294 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.2183436 , max =  0.914
##   A new explainer has been created!
```

```
p2 <- plot(model_profile(modelex, variables = "age", groups = "weekday_weekend")) + ggtitle("GAD2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_Weekday" = "Weekday",
                                "randomForest_Weekend" = "Weekend"),
```

```
                      values = c("randomForest_Weekday" = "red",
                                 "randomForest_Weekend" = "blue"),
                      name   = "")
p2
```



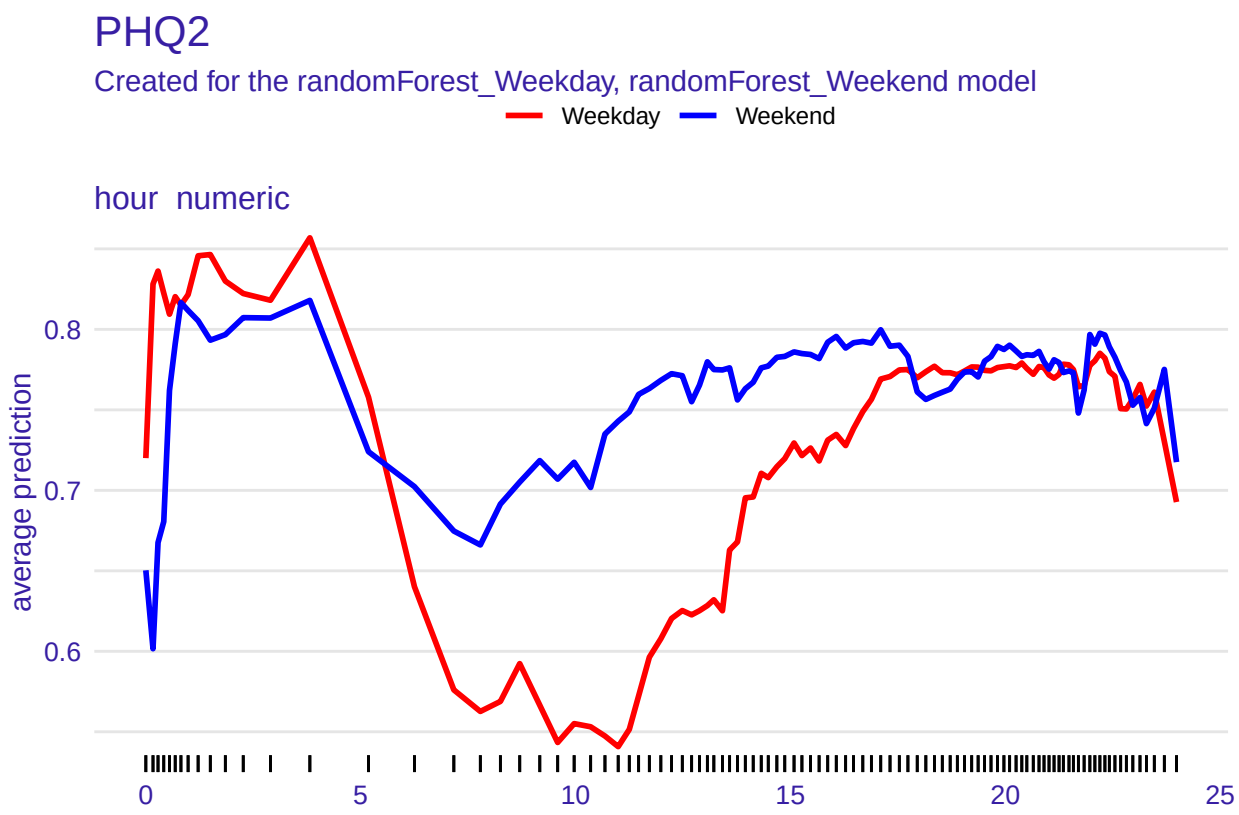
2.1.3. The hour vs. PHQ2

```
library(DALEX)
library(randomForest)
library(modelStudio)

model <- randomForest(factor(PHQ2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$PHQ2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.004 , mean =  0.7545646 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -0.998 , mean =  -0.1828374 , max =  0.94
##   A new explainer has been created!
```

```
p4 <- plot(model_profile(modelex, variables = "hour_numeric", groups = "weekday_weekend")) + ggtitle("PH
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_Weekday" = "Weekday",
                                "randomForest_Weekend" = "Weekend"),
                     values = c("randomForest_Weekday" = "red",
                                "randomForest_Weekend" = "blue"),
                     name   = "")
p4
```

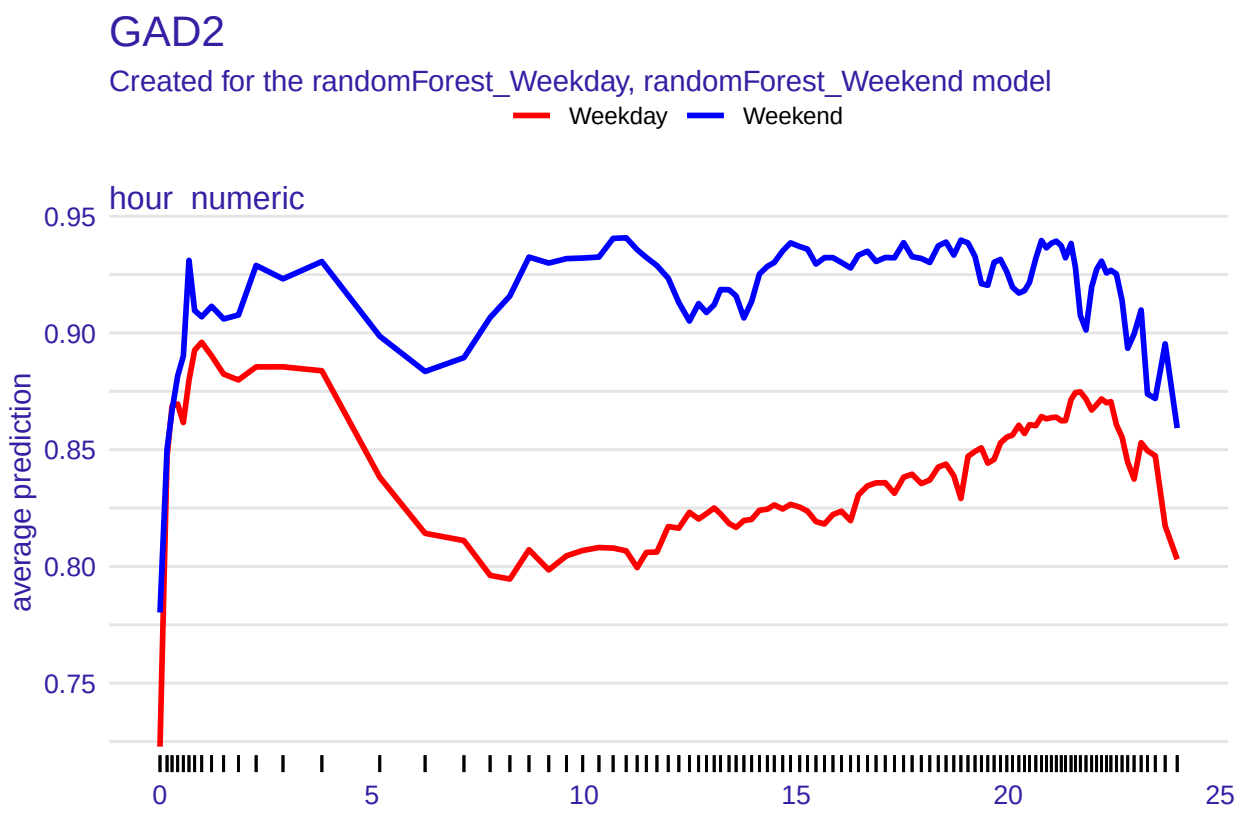


2.1.4. The hour vs. GAD2

```
model <- randomForest(factor(GAD2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$GAD2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.028 , mean =  0.8802342 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.2195485 , max =  0.926
##   A new explainer has been created!
```

```
p5 <- plot(model_profile(modelex, variables = "hour_numeric", groups = "weekday_weekend")) + ggtitle("GA
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_Weekday" = "Weekday",
                                "randomForest_Weekend" = "Weekend"),
                     values = c("randomForest_Weekday" = "red",
                                "randomForest_Weekend" = "blue"),
                     name   = "")
p5
```



2.2. Relationships by education levels

2.2.1. The age vs. PHQ2

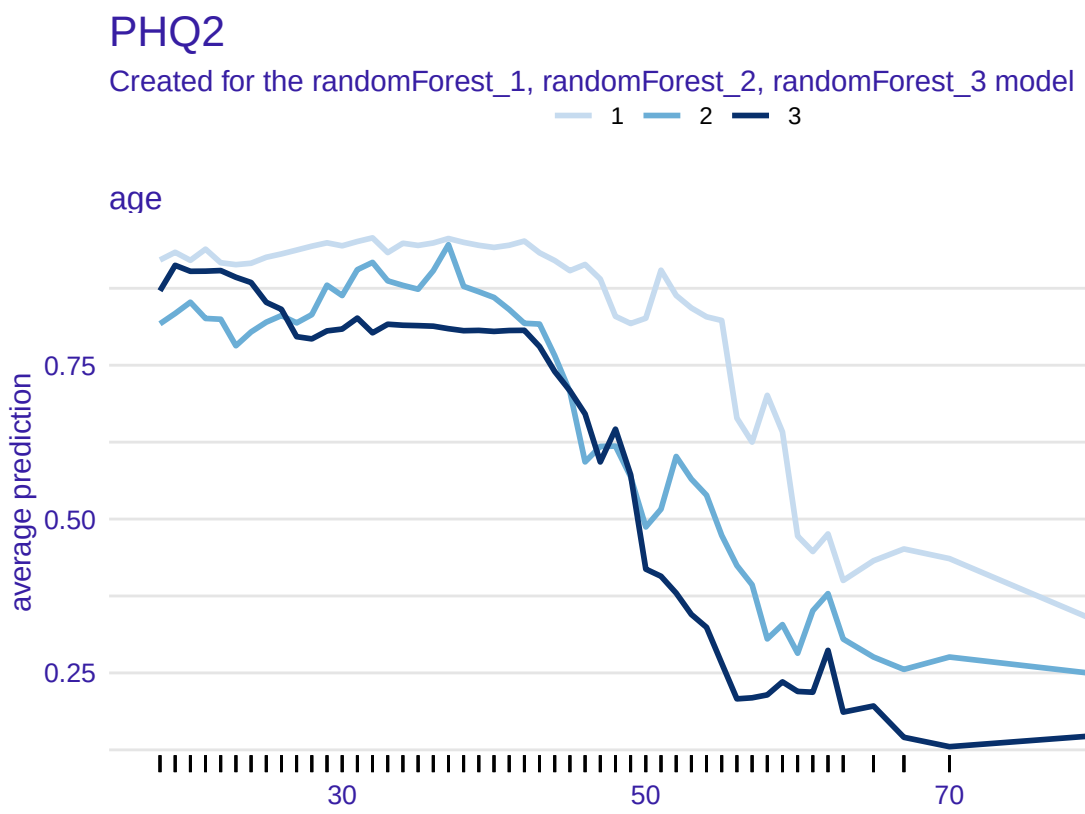
```
library(DALEX)
library(randomForest)
library(modelStudio)

model <- randomForest(factor(PHQ2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$PHQ2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
```

```
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.004 , mean =  0.7508285 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -0.998 , mean =  -0.1791013 , max =  0.944
##   A new explainer has been created!
```

```
p1 <- plot(model_profile(modelex, variables = "age", groups = "education")) + ggtitle("PHQ2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2",
                                "randomForest_3" = "3"),
                     values = c("randomForest_1" = "#c6dbef",
                                "randomForest_2" = "#6baed6",
                                "randomForest_3" = "#08306b"),
                     name   = "")
p1
```



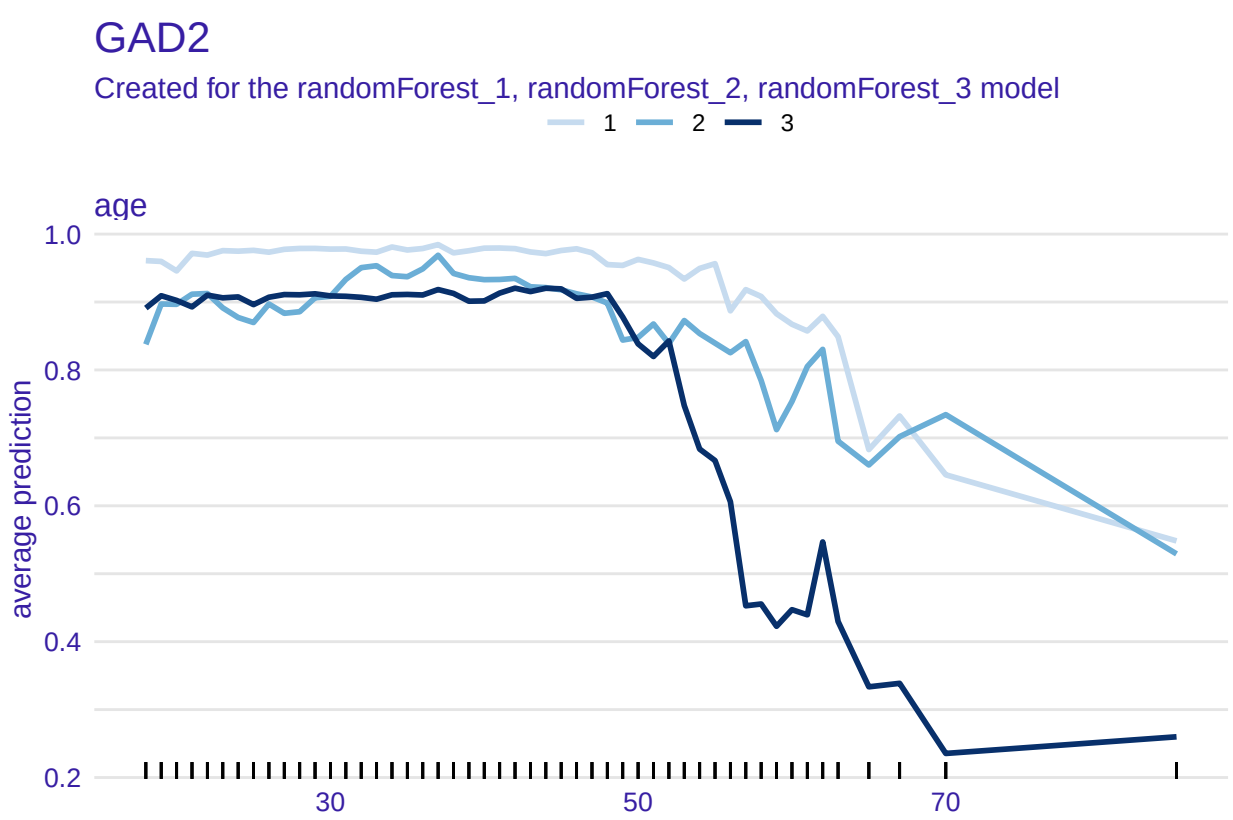
2.2.2. The age vs. GAD2

```
model <- randomForest(factor(GAD2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$GAD2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
```

```
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.024 , mean =  0.8805225 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.2198367 , max =  0.932
##   A new explainer has been created!
```

```
p2 <- plot(model_profile(modelex, variables = "age", groups = "education")) + ggtitle("GAD2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2",
                                "randomForest_3" = "3"),
                     values = c("randomForest_1" = "#c6dbef",
                                "randomForest_2" = "#6baed6",
                                "randomForest_3" = "#08306b"),
                     name   = "")
p2
```



2.2.3. The hour vs. PHQ2

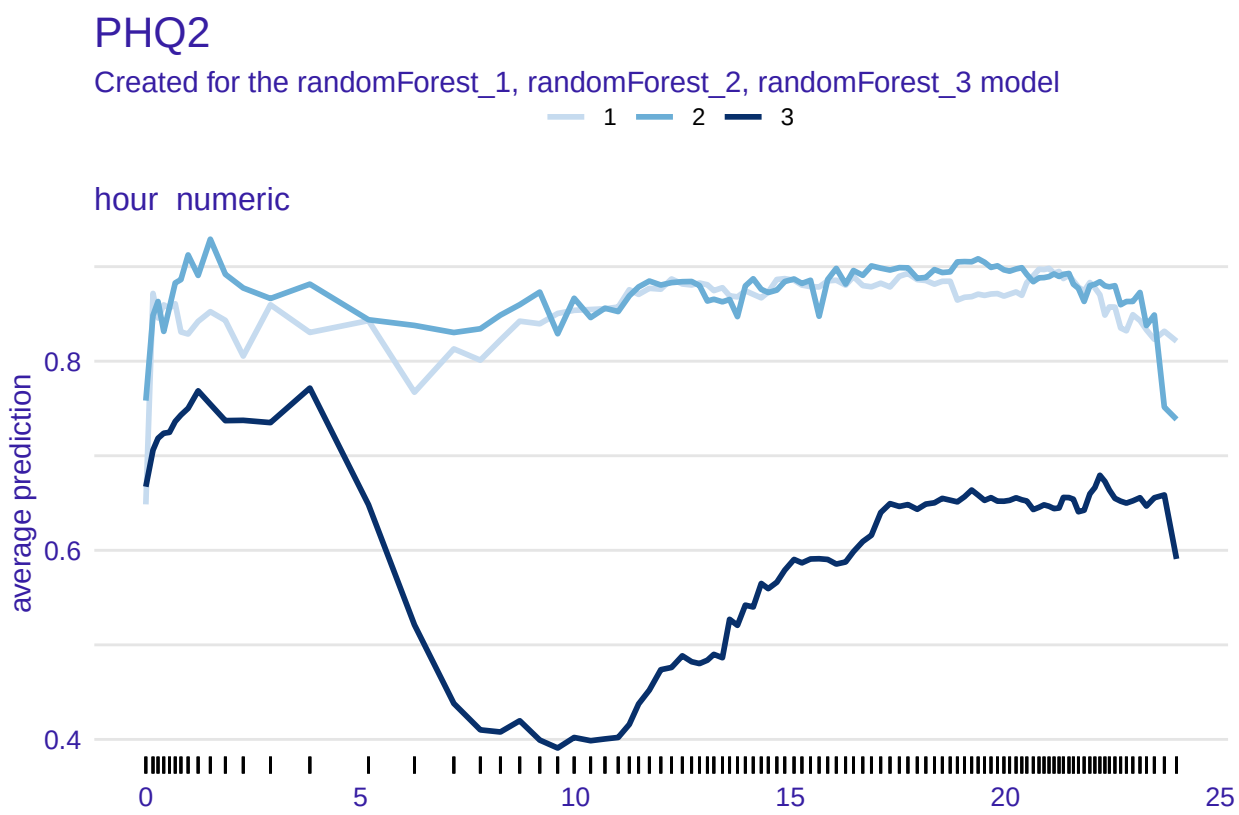
```
library(DALEX)
library(randomForest)
library(modelStudio)

model <- randomForest(factor(PHQ2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
```

```
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$PHQ2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.008 , mean =  0.74981 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -0.996 , mean =  -0.1780828 , max =  0.95
##   A new explainer has been created!
```

```
p4 <- plot(model_profile(modelex, variables = "hour_numeric", groups = "education")) + ggtitle("PHQ2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2",
                                "randomForest_3" = "3"),
                     values = c("randomForest_1" = "#c6dbef",
                                "randomForest_2" = "#6baed6",
                                "randomForest_3" = "#08306b"),
                     name   = "")
p4
```

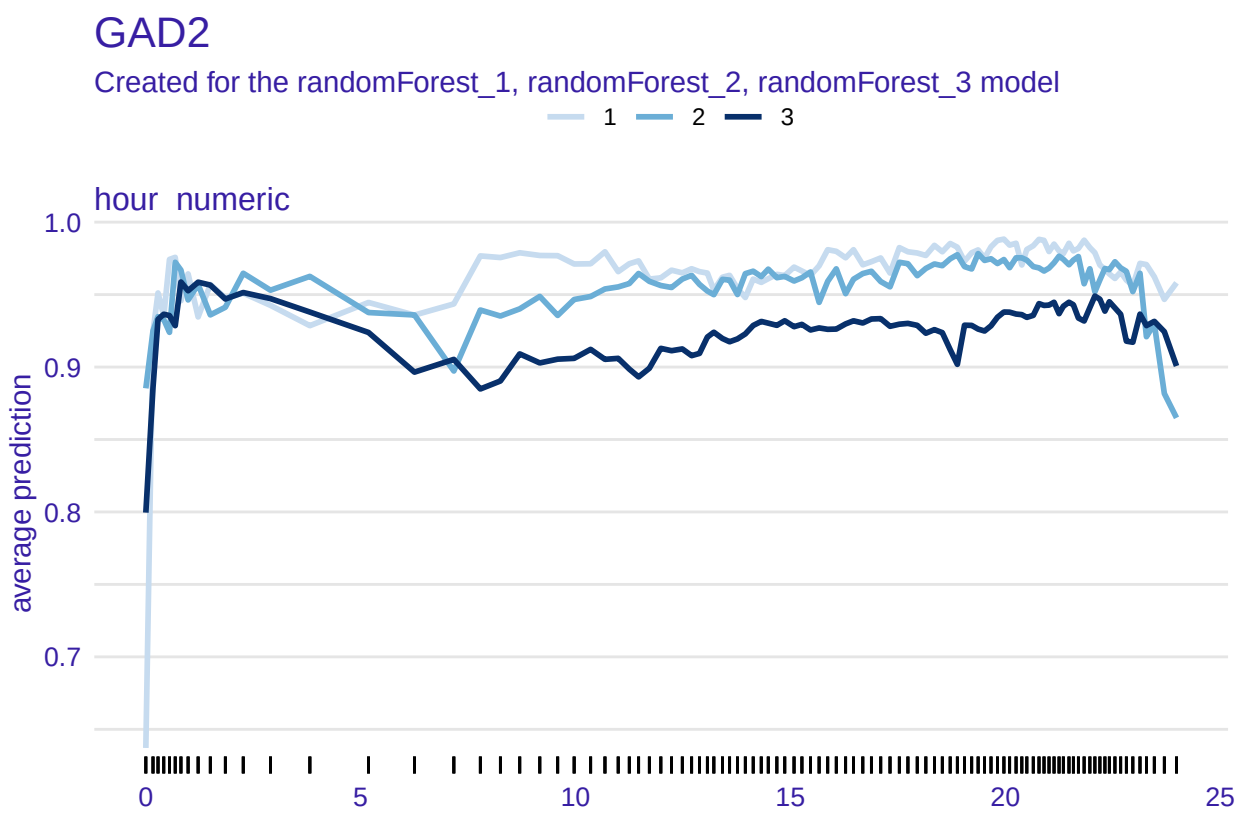


2.2.4. The hour vs. GAD2

```
model <- randomForest(factor(GAD2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$GAD2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.024 , mean =  0.8805814 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.2198957 , max =  0.922
##   A new explainer has been created!
```

```
p5 <- plot(model_profile(modelex, variables = "hour_numeric", groups = "education")) + ggtitle("GAD2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2",
                                "randomForest_3" = "3"),
                     values = c("randomForest_1" = "#c6dbef",
                                "randomForest_2" = "#6baed6",
                                "randomForest_3" = "#08306b"),
                     name   = "")
p5
```



2.3. Relationships by gender

2.3.1. The age vs. PHQ2

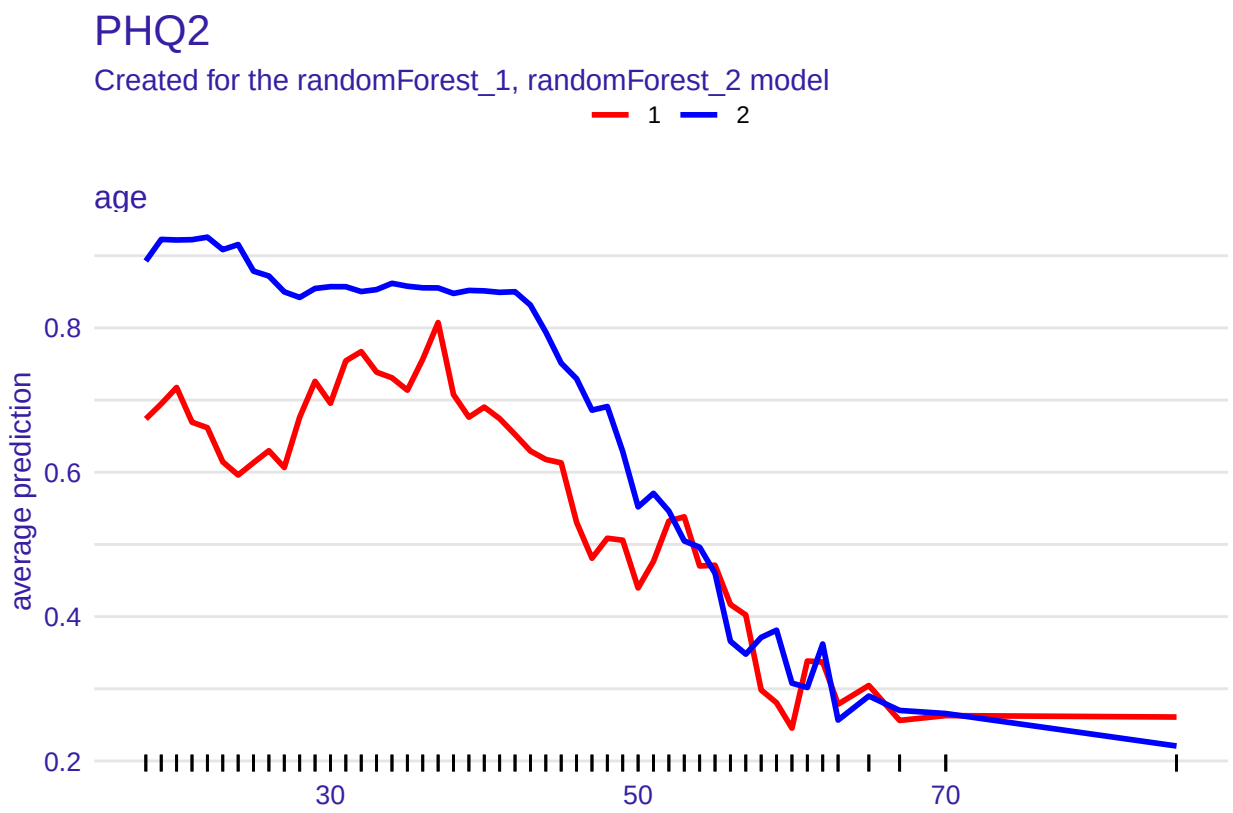
```
library(DALEX)
library(randomForest)
library(modelStudio)

model <- randomForest(factor(PHQ2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$PHQ2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.008 , mean =  0.7519774 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.1802502 , max =  0.944
##   A new explainer has been created!
```

```
p1 <- plot(model_profile(modelex, variables = "age", groups = "sex")) + ggtitle("PHQ2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
```

```
                          "randomForest_2" = "2"),
                values = c("randomForest_1" = "red",
                          "randomForest_2" = "blue"),
                name   = "")
p1
```



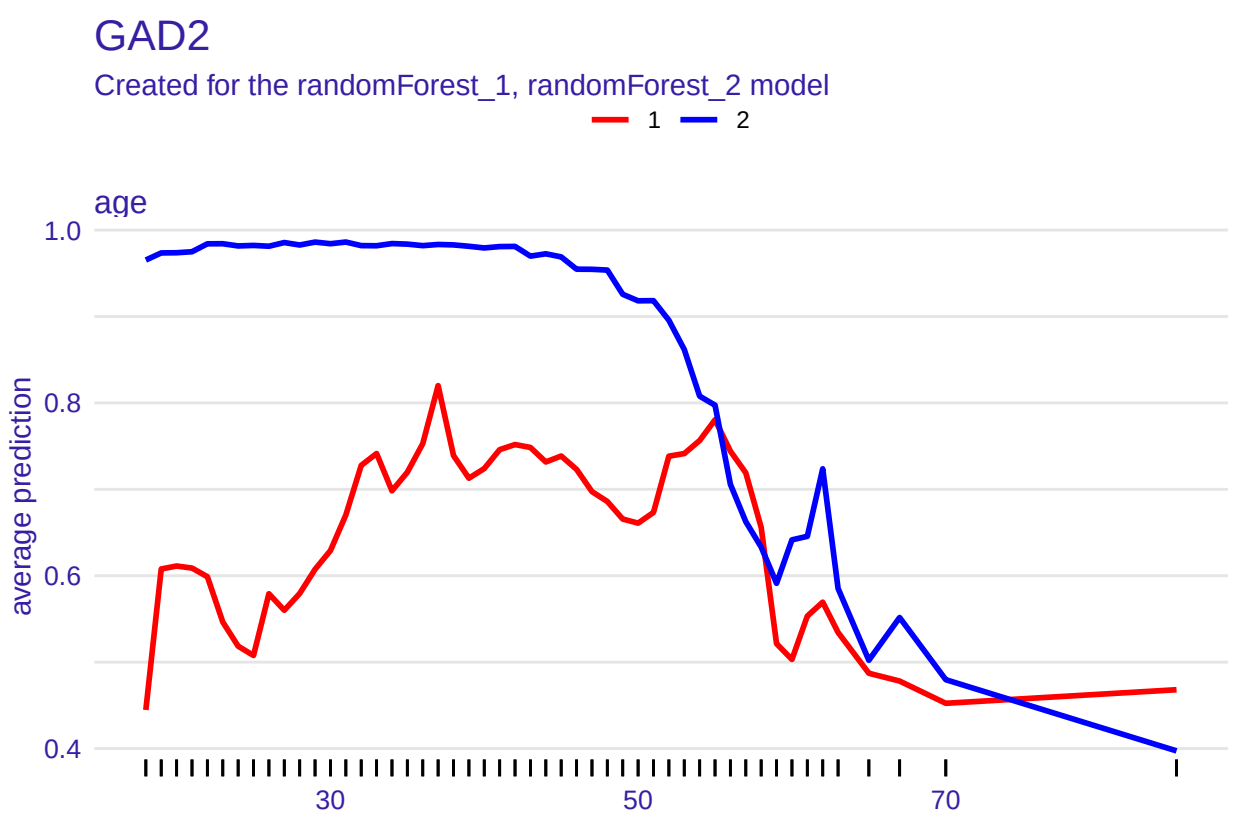
2.3.2. The age vs. GAD2

```
model <- randomForest(factor(GAD2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$GAD2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.024 , mean =  0.8793555 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
##   -> residuals         :  numerical, min =  -1 , mean =  -0.2186697 , max =  0.912
##   A new explainer has been created!
```

```
p2 <- plot(model_profile(modelex, variables = "age", groups = "sex")) + ggtitle("GAD2") +
  geom_rug(sides = "b") +
```

```
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2"),
                     values = c("randomForest_1" = "red",
                                "randomForest_2" = "blue"),
                     name   = "")
p2
```



2.3.3. The hour vs. PHQ2

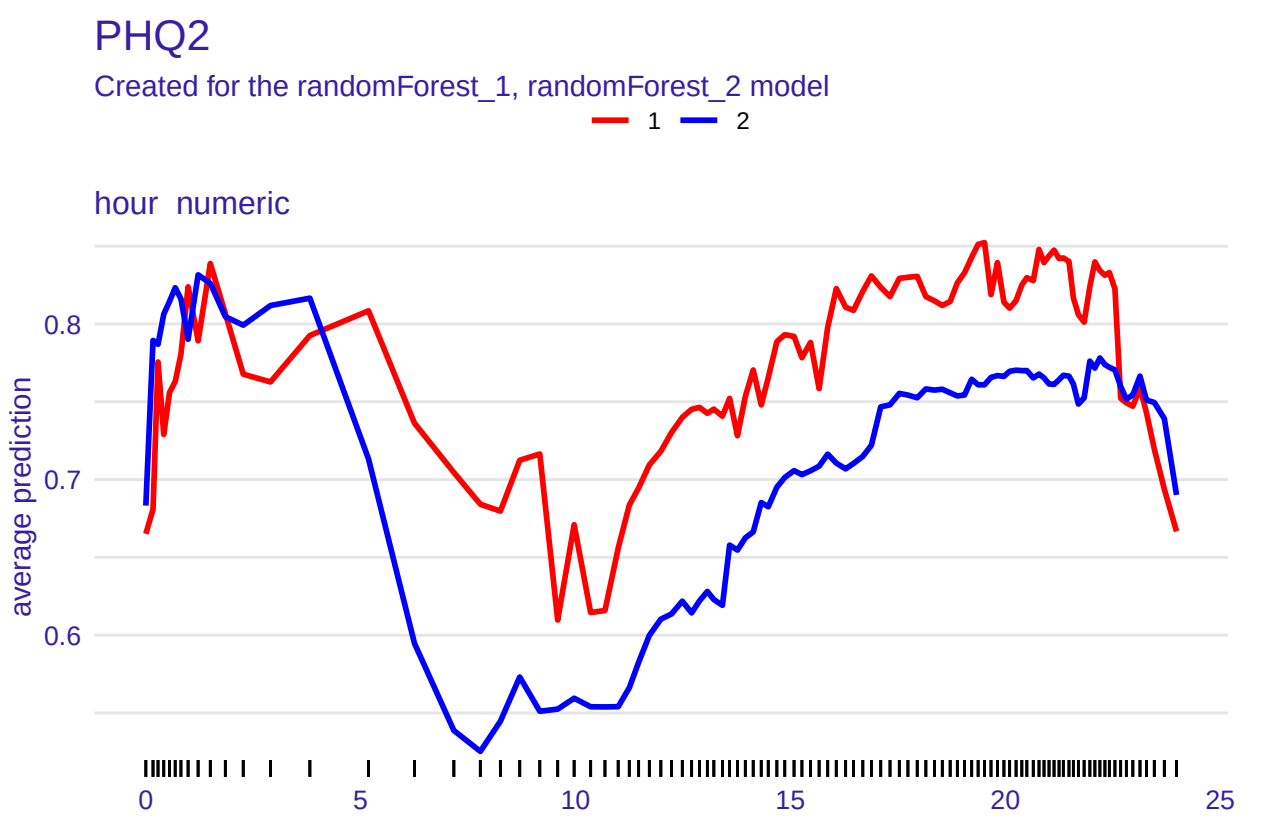
```
library(DALEX)
library(randomForest)
library(modelStudio)

model <- randomForest(factor(PHQ2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$PHQ2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.004 , mean =  0.7454884 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
```

```
##   -> residuals         :  numerical, min =  -0.998 , mean =  -0.1737612 , max =  0.956
##   A new explainer has been created!
```

```
p4 <- plot(model_profile(modelex, variables = "hour_numeric", groups = "sex")) + ggtitle("PHQ2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2"),
                     values = c("randomForest_1" = "red",
                                "randomForest_2" = "blue"),
                     name   = "")
p4
```



2.3.4. The hour vs. GAD2

```
model <- randomForest(factor(GAD2_score_cat) ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                   y = datadf$GAD2_score_cat)
```

```
## Preparation of a new explainer is initiated
##   -> model label       :  randomForest  (  default  )
##   -> data              :  34443  rows  6  cols
##   -> data              :  tibble converted into a data.frame
##   -> target variable   :  34443  values
##   -> predict function  :  yhat.randomForest  will be used (  default  )
##   -> predicted values  :  No value for predict function target column. (  default  )
##   -> model_info        :  package randomForest , ver. 4.7.1.2 , task classification (  default  )
##   -> predicted values  :  numerical, min =  0.026 , mean =  0.8796267 , max =  1
##   -> residual function :  difference between y and yhat (  default  )
```

```
##   -> residuals         :  numerical, min =  -1 , mean =  -0.218941 , max =  0.91
##   A new explainer has been created!
```

```
p5 <- plot(model_profile(modelex, variables = "hour_numeric", groups = "sex")) + ggtitle("GAD2") +
  geom_rug(sides = "b") +
  scale_color_manual(labels = c("randomForest_1" = "1",
                                "randomForest_2" = "2"),
                     values = c("randomForest_1" = "red",
                                "randomForest_2" = "blue"),
                     name   = "")
p5
```