

# 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(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 regression ( default )
## -> predicted values : numerical, min = 0.3426764 , mean = 0.5714852 , max = 0.7331799
## -> residual function : difference between y and yhat ( default )
## -> residuals       : numerical, min = -0.7279689 , mean = 0.0002420015 , max = 0.6396872
## 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",
```

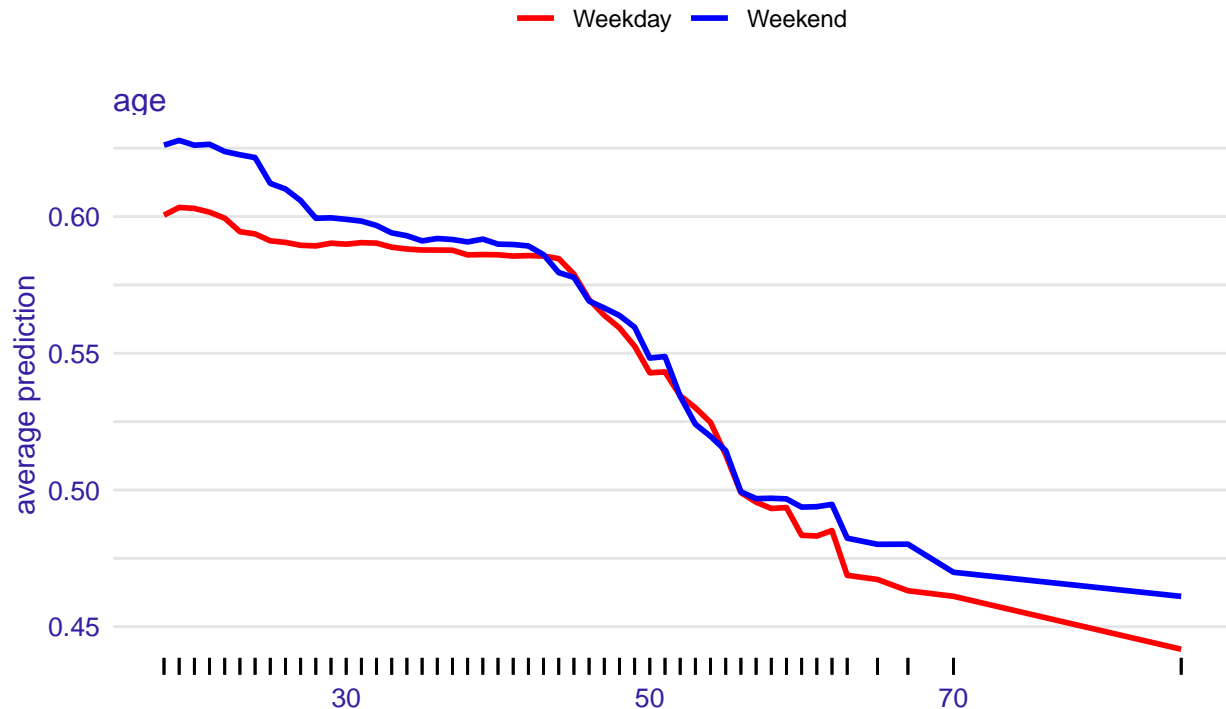
```

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

```

## PHQ2

Created for the randomForest\_Weekday, randomForest\_Weekend model



### 2.1.2. The age vs. GAD2

```

model <- randomForest(GAD2_score_cat ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modellex <- 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 regression ( default )
## -> predicted values : numerical, min = 0.4313288 , mean = 0.6607001 , max = 0.7730129
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7700419 , mean = -1.435927e-05 , max = 0.5546089
## A new explainer has been created!

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

```

```

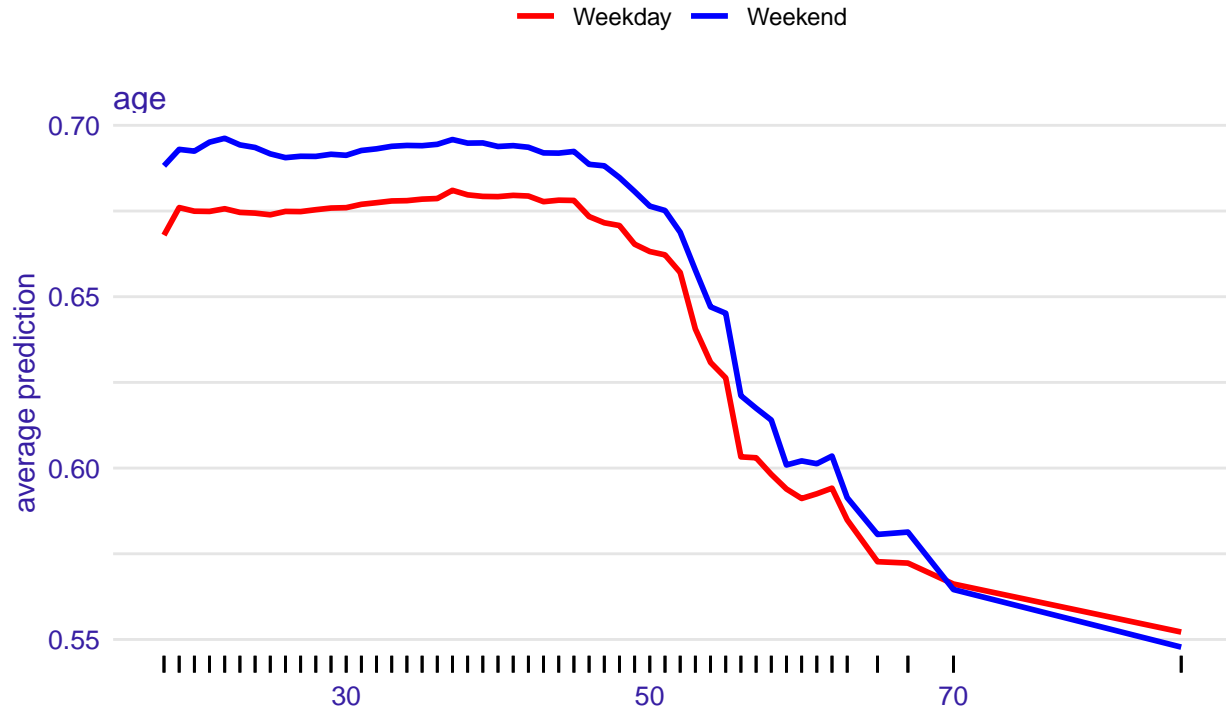
values = c("randomForest_Weekday" = "red",
           "randomForest_Weekend" = "blue"),
name    = "")

```

p2

## GAD2

Created for the randomForest\_Weekday, randomForest\_Weekend model



### 2.1.3. The hour vs. PHQ2

```

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

model <- randomForest(PHQ2_score_cat ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modellex <- 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 regression ( default )
## -> predicted values : numerical, min = 0.33895 , mean = 0.5716432 , max = 0.7325983
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7238825 , mean = 8.400072e-05 , max = 0.6429967
## A new explainer has been created!

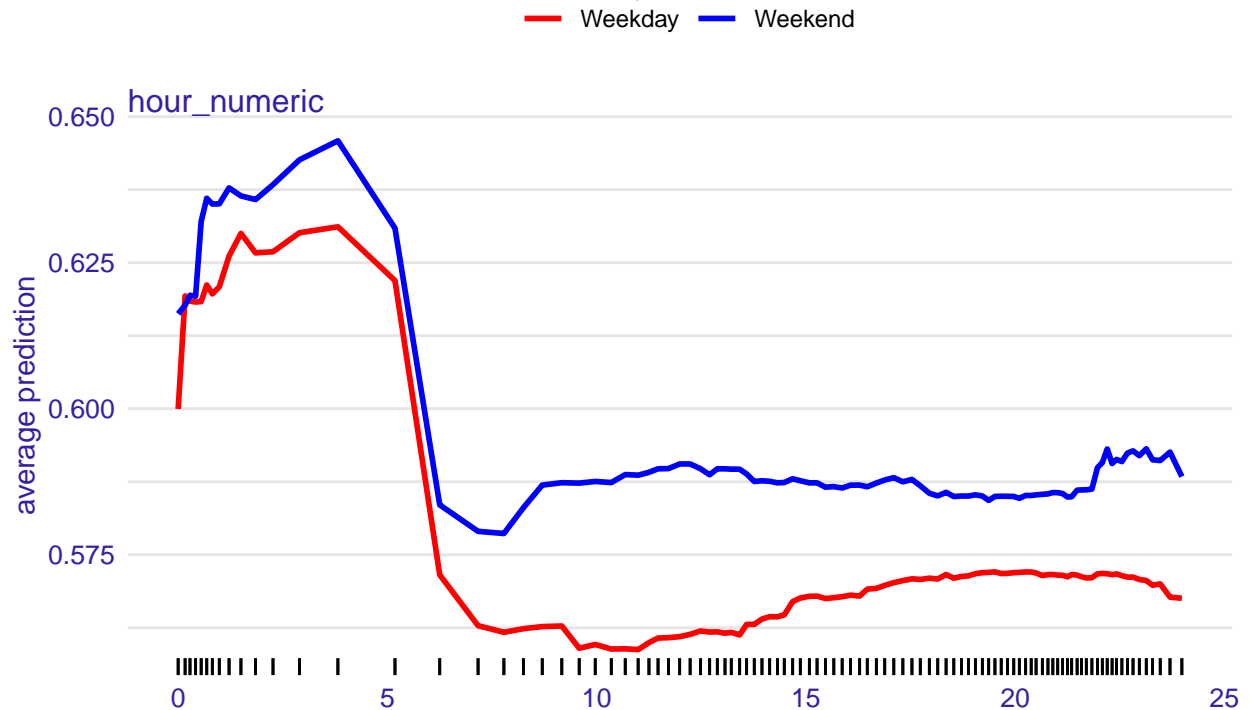
```

```
p4 <- plot(model_profile(model, variables = "hour_numeric", 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 = "")
```

p4

## PHQ2

Created for the randomForest\_Weekday, randomForest\_Weekend model



### 2.1.4. The hour vs. GAD2

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

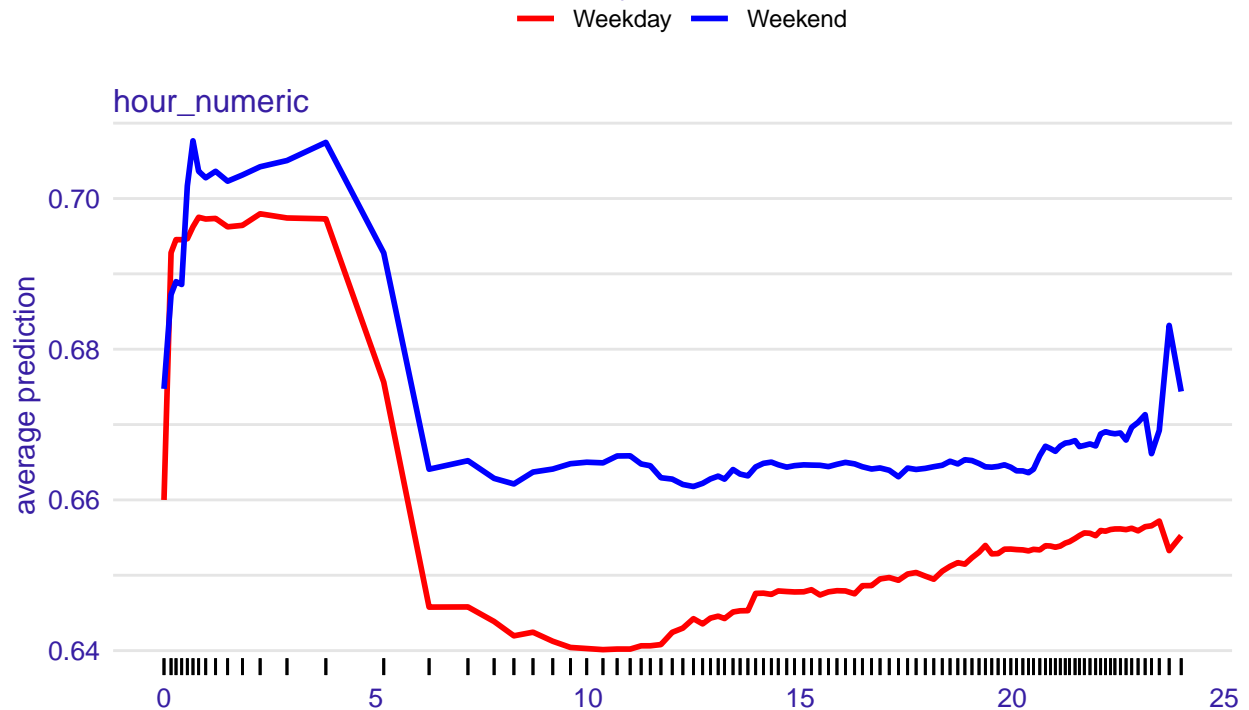
## Preparation of a new explainer is initiated
## -> model label      : randomForest ( default )
## -> data             : 34443 rows 6 cols
## -> data             : tibble converted into a data.frame
## -> target variable  : 34443 values
## -> predict function : yhat.randomForest will be used ( default )
## -> predicted values : No value for predict function target column. ( default )
## -> model_info       : package randomForest , ver. 4.7.1.2 , task regression ( default )
## -> predicted values : numerical, min = 0.4206829 , mean = 0.6608802 , max = 0.7753274
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7677867 , mean = -0.0001944491 , max = 0.5621384
## A new explainer has been created!
```

```
p5 <- plot(model_profile(modelex, variables = "hour_numeric", 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 = "")
```

p5

## GAD2

Created for the randomForest\_Weekday, randomForest\_Weekend model



## 2.2. Relationships by education levels

### 2.2.1. The age vs. PHQ2

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

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

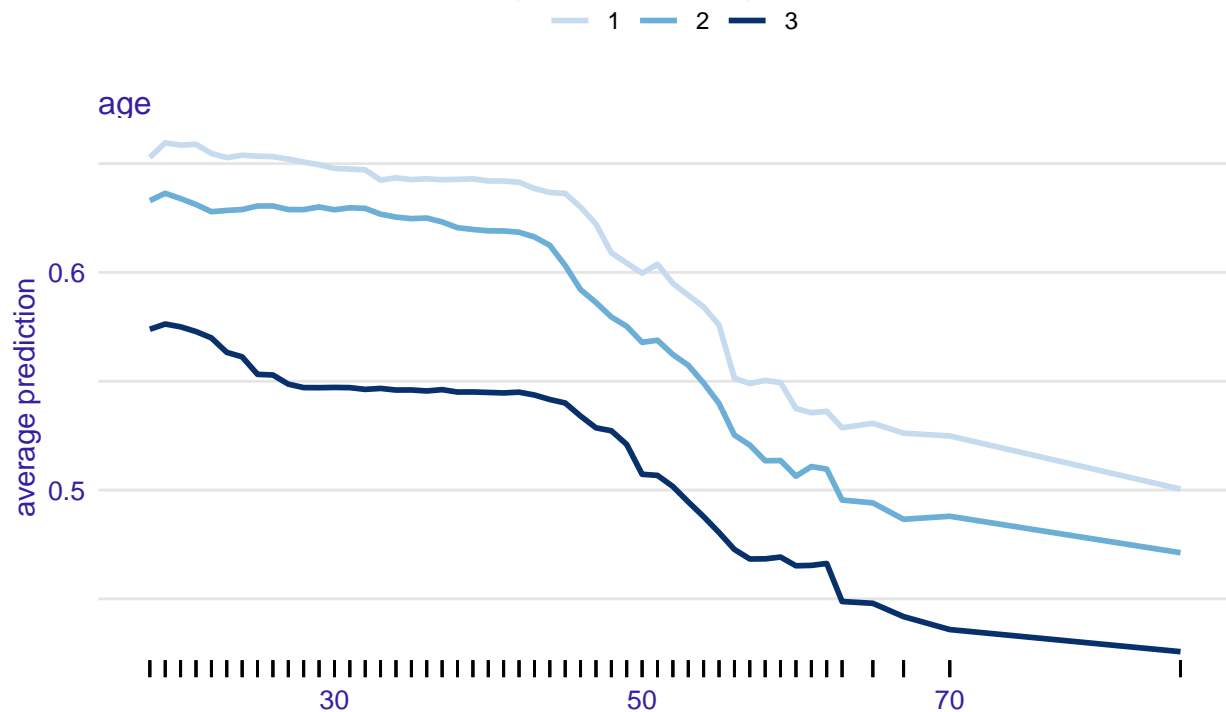
```
## -> predicted values : No value for predict function target column. ( default )
## -> model_info       : package randomForest , ver. 4.7.1.2 , task regression ( default )
## -> predicted values : numerical, min = 0.3517176 , mean = 0.5719798 , max = 0.7341967
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7287008 , mean = -0.0002526178 , max = 0.6430072
## 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

## PHQ2

Created for the randomForest\_1, randomForest\_2, randomForest\_3 model



### 2.2.2. The age vs. GAD2

```
model <- randomForest(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
```

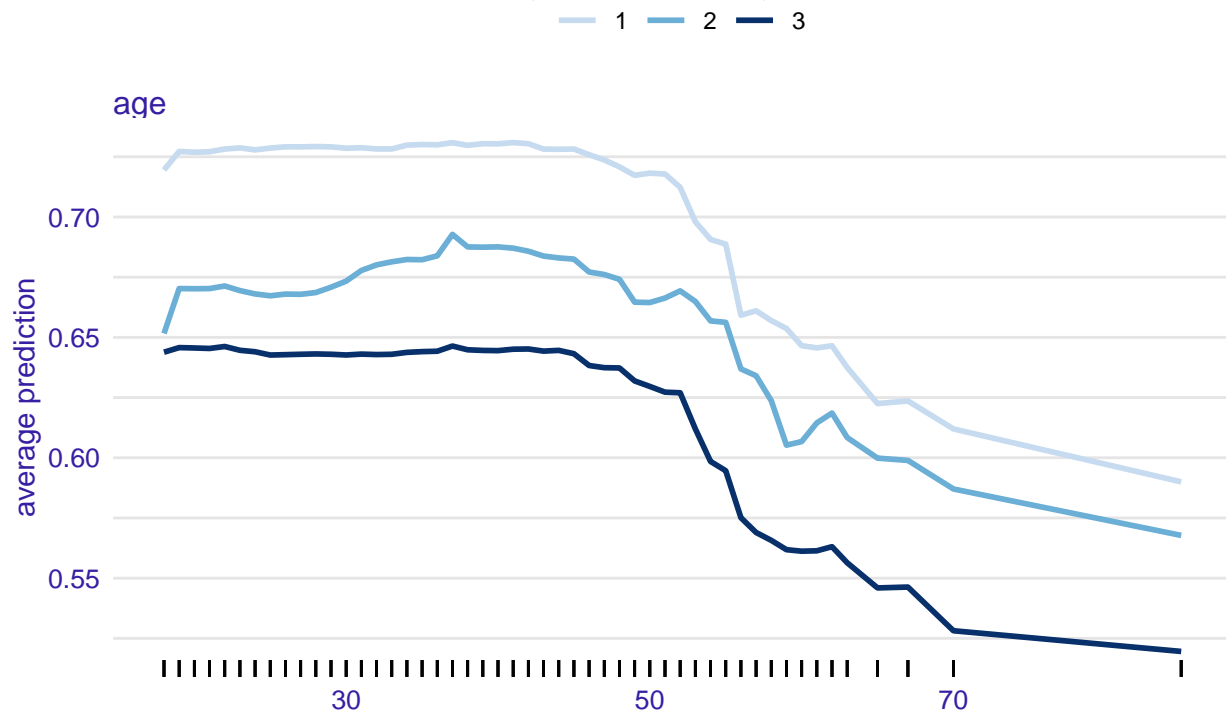
```
## -> target variable : 34443 values
## -> predict function : yhat.randomForest will be used ( default )
## -> predicted values : No value for predict function target column. ( default )
## -> model_info : package randomForest , ver. 4.7.1.2 , task regression ( default )
## -> predicted values : numerical, min = 0.4354447 , mean = 0.6607732 , max = 0.7686952
## -> residual function : difference between y and yhat ( default )
## -> residuals : numerical, min = -0.7621654 , mean = -8.740142e-05 , max = 0.553584
## 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

## GAD2

Created for the randomForest\_1, randomForest\_2, randomForest\_3 model



### 2.2.3. The hour vs. PHQ2

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

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

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

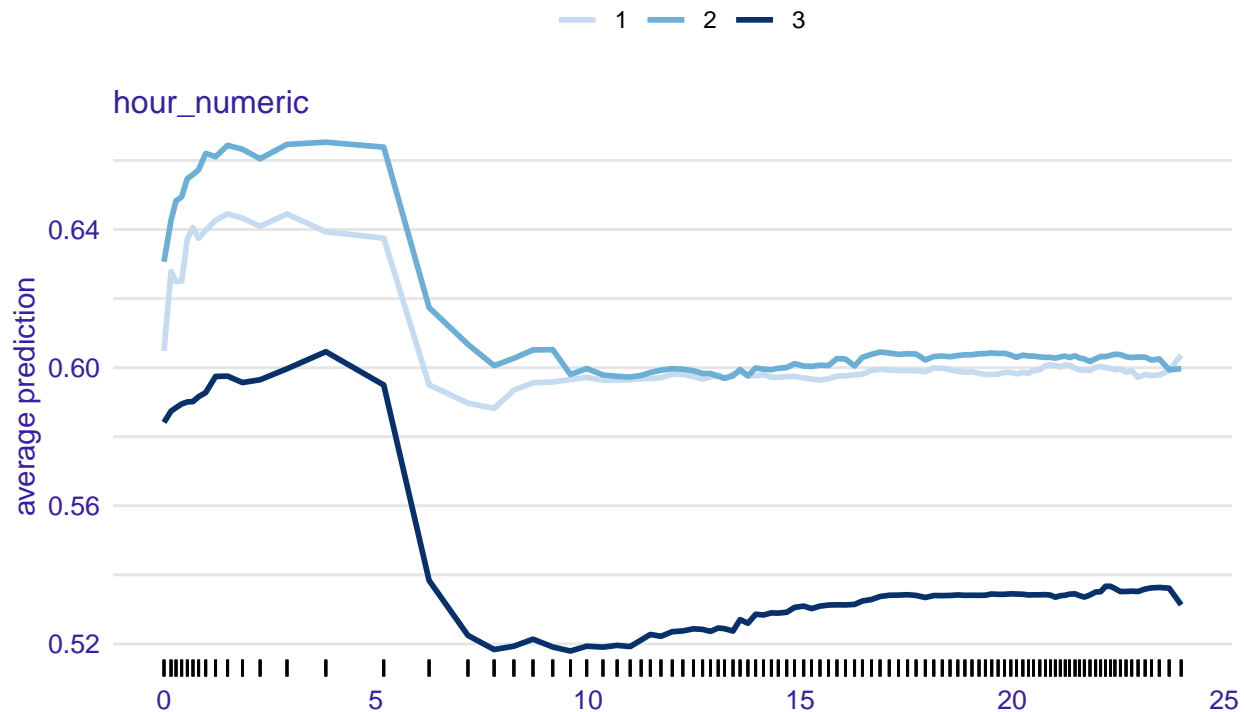
```
## Preparation of a new explainer is initiated
## -> model label      : randomForest ( default )
## -> data             : 34443 rows 6 cols
## -> data             : tibble converted into a data.frame
## -> target variable  : 34443 values
## -> predict function : yhat.randomForest will be used ( default )
## -> predicted values : No value for predict function target column. ( default )
## -> model_info       : package randomForest , ver. 4.7.1.2 , task regression ( default )
## -> predicted values : numerical, min = 0.3417992 , mean = 0.5717026 , max = 0.7321723
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7275672 , mean = 2.459055e-05 , max = 0.6455207
## A new explainer has been created!
```

```
p4 <- plot(model_profile(model_ex, 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

## PHQ2

Created for the randomForest\_1, randomForest\_2, randomForest\_3 model





## 2.2.4. The hour vs. GAD2

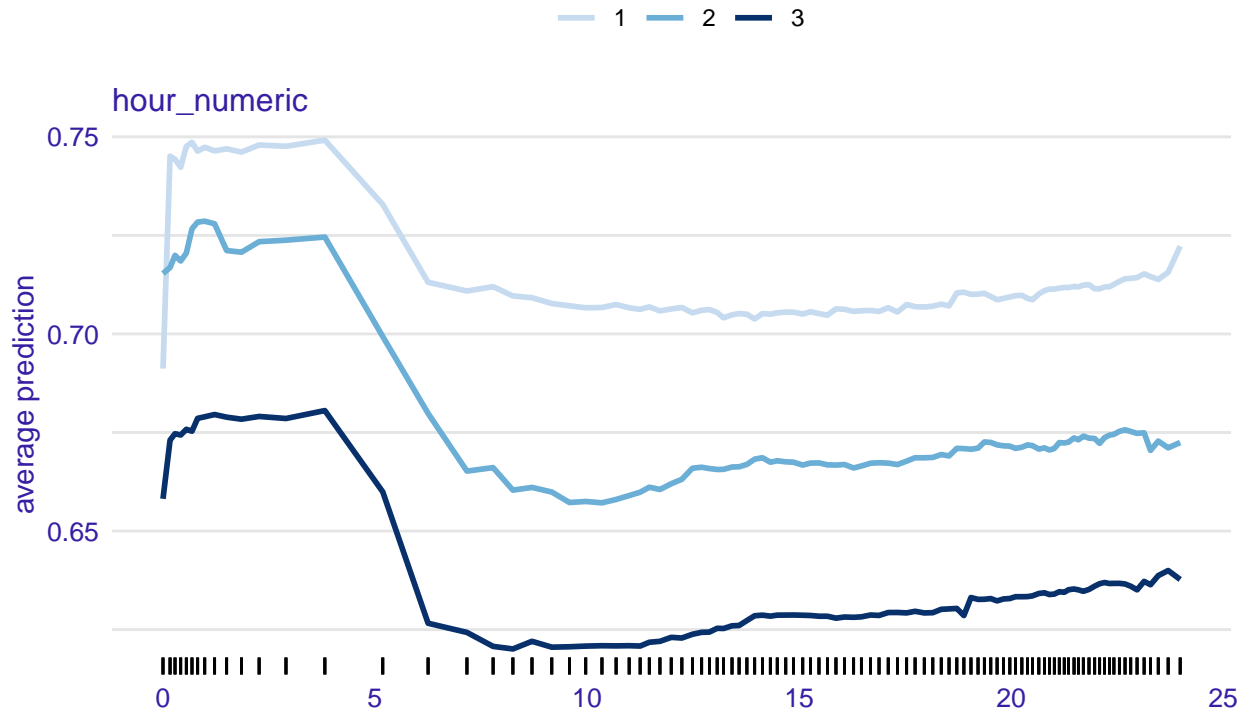
```
model <- randomForest(GAD2_score_cat ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modelex <- explain(model, data = datadf[,c(8, 9, 10, 11, 15, 16)],
                  y = datadf$GAD2_score_cat)

## 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 regression ( default )
## -> predicted values : numerical, min = 0.4414311 , mean = 0.6606569 , max = 0.7729693
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7652176 , mean = 2.886801e-05 , max = 0.5454419
## 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
```

## GAD2

Created for the randomForest\_1, randomForest\_2, randomForest\_3 model



### 2.3. Relationships by gender

#### 2.3.1. The age vs. PHQ2

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

model <- randomForest(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 regression ( default )
## -> predicted values : numerical, min = 0.3306704 , mean = 0.5716797 , max = 0.7349671
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7301232 , mean = 4.75396e-05 , max = 0.6544156
## 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",
```

```

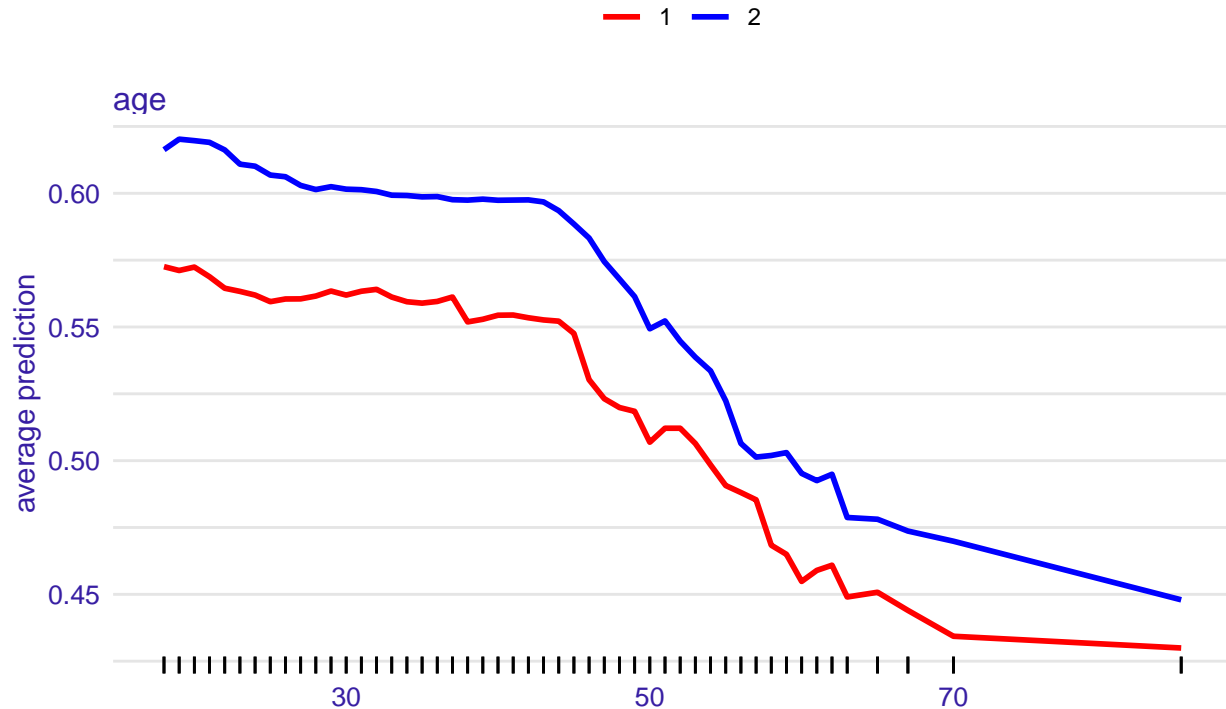
      "randomForest_2" = "2"),
values = c("randomForest_1" = "red",
      "randomForest_2" = "blue"),
name    = "")

```

p1

## PHQ2

Created for the randomForest\_1, randomForest\_2 model



### 2.3.2. The age vs. GAD2

```

model <- randomForest(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 regression ( default )
## -> predicted values : numerical, min = 0.4367064 , mean = 0.6607382 , max = 0.7692376
## -> residual function : difference between y and yhat ( default )
## -> residuals        : numerical, min = -0.7622171 , mean = -5.247157e-05 , max = 0.5557547
## A new explainer has been created!

```

```

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

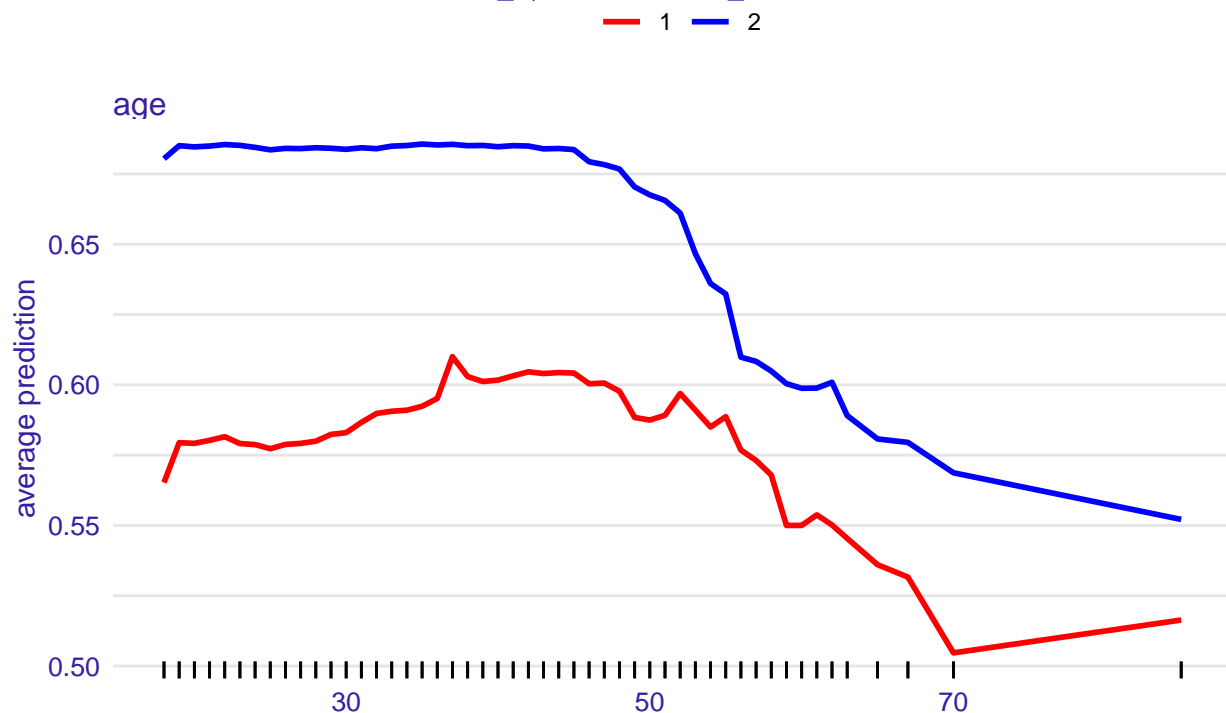
```

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

p2

## GAD2

Created for the randomForest\_1, randomForest\_2 model



### 2.3.3. The hour vs. PHQ2

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

model <- randomForest(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)

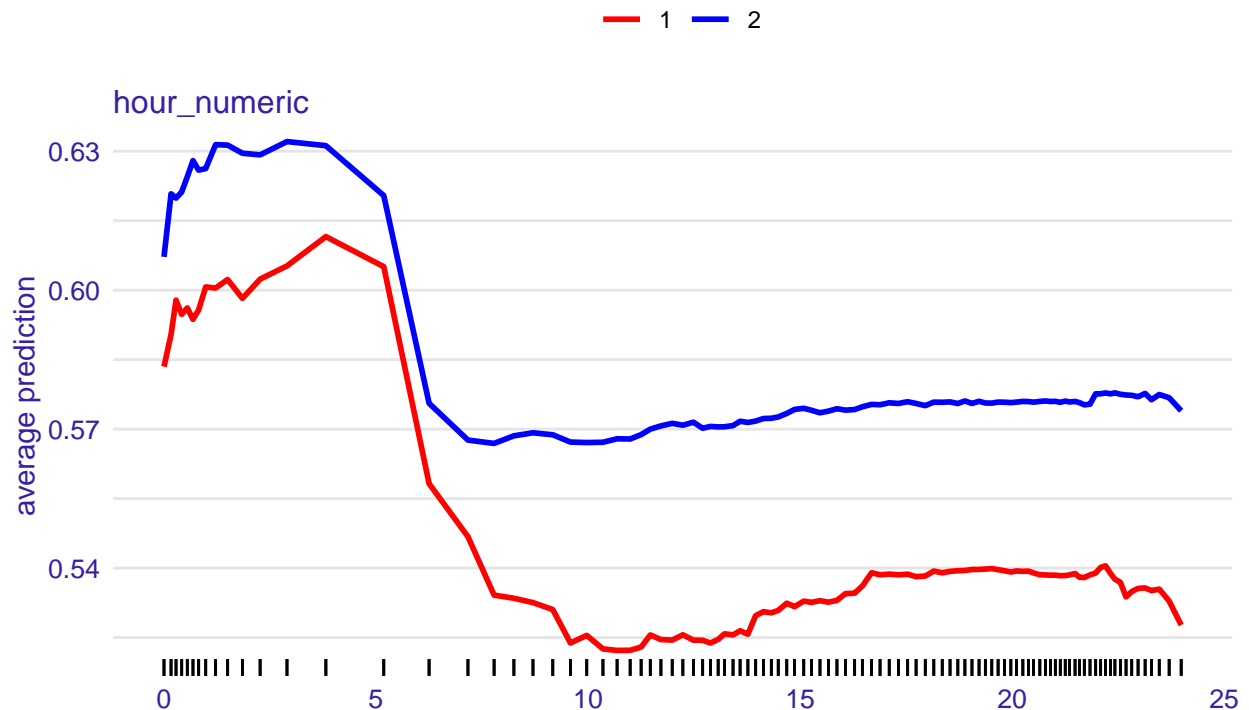
## Preparation of a new explainer is initiated
## -> model label      : randomForest ( default )
## -> data             : 34443 rows 6 cols
## -> data             : tibble converted into a data.frame
## -> target variable  : 34443 values
## -> predict function : yhat.randomForest will be used ( default )
## -> predicted values : No value for predict function target column. ( default )
## -> model_info       : package randomForest , ver. 4.7.1.2 , task regression ( default )
## -> predicted values : numerical, min = 0.342336 , mean = 0.5716836 , max = 0.7307588
## -> residual function : difference between y and yhat ( default )
```

```
## -> residuals      : numerical, min = -0.7239865 , mean = 4.355311e-05 , max = 0.6351379
## A new explainer has been created!

p4 <- plot(model_profile(model, 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
```

## PHQ2

Created for the randomForest\_1, randomForest\_2 model



### 2.3.4. The hour vs. GAD2

```
model <- randomForest(GAD2_score_cat ~ age + education + sex + weekday_weekend + hour_numeric,
                      data = datadf)
modellex <- 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 regression ( default )
## -> predicted values : numerical, min = 0.4405806 , mean = 0.6606958 , max = 0.7696138
## -> residual function : difference between y and yhat ( default )
```

```
## -> residuals      : numerical, min = -0.7662779 , mean = -9.996047e-06 , max = 0.5497871
## 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

## GAD2

Created for the randomForest\_1, randomForest\_2 model

— 1 — 2

