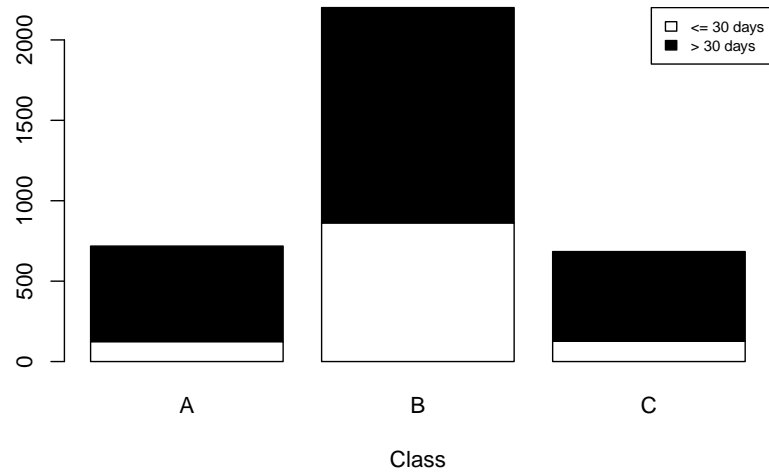


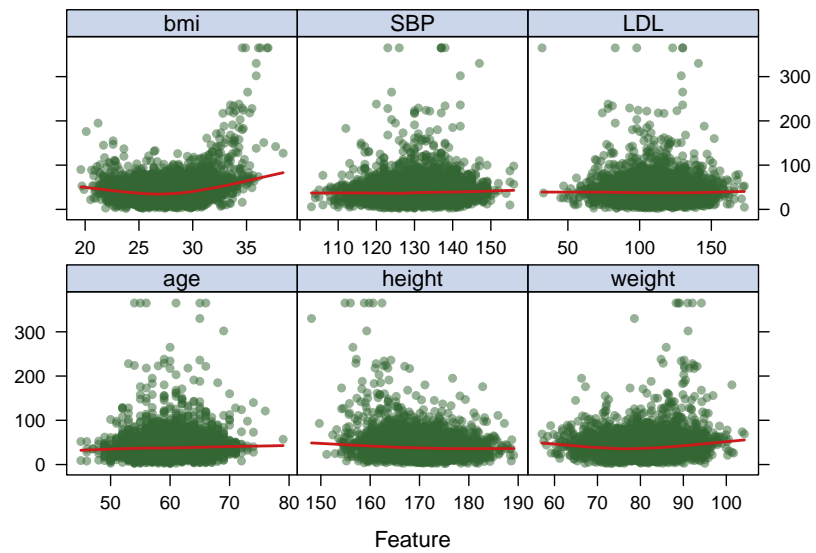
Appendix – Primary Analysis of Covid Recovery Time

Exploratory analysis and data visualization

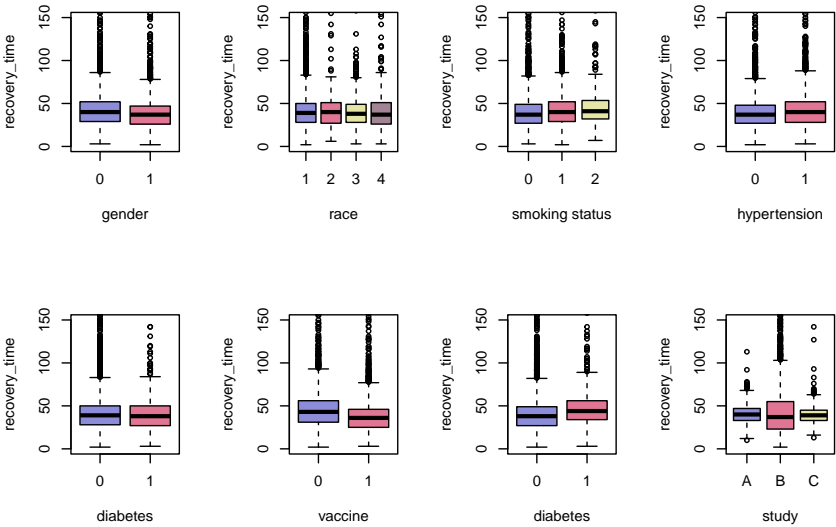
Number of cases seperated by 30 days in recovery by Study Group



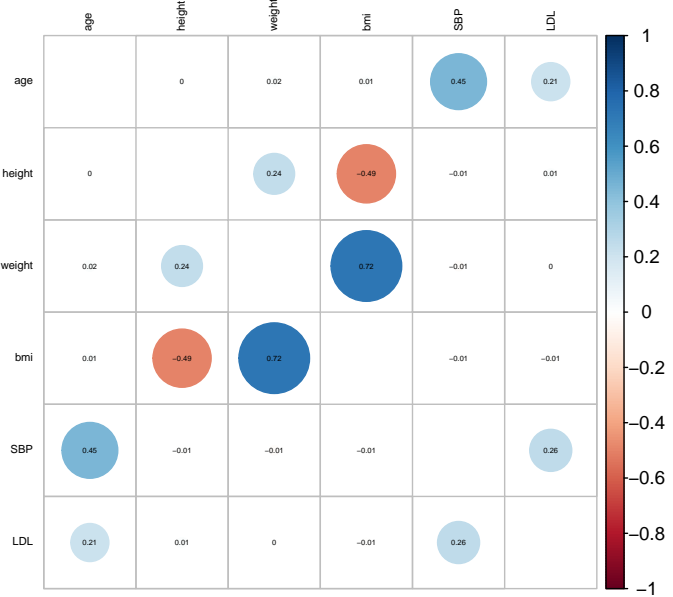
Visualize potential relationship between reponse variable and numeric predictors



Visualize potential relationship between reponse variable and categorical predictors

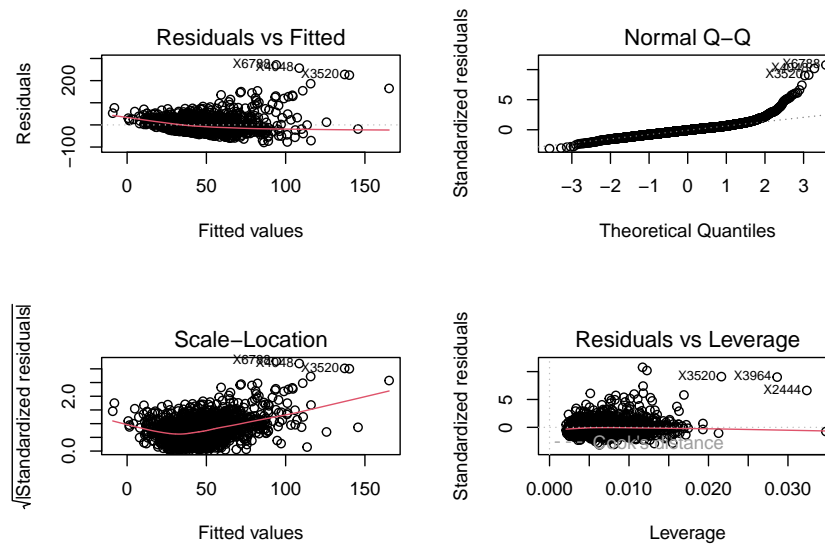


Correlation plot to check collinearity between covariates (based on training data)

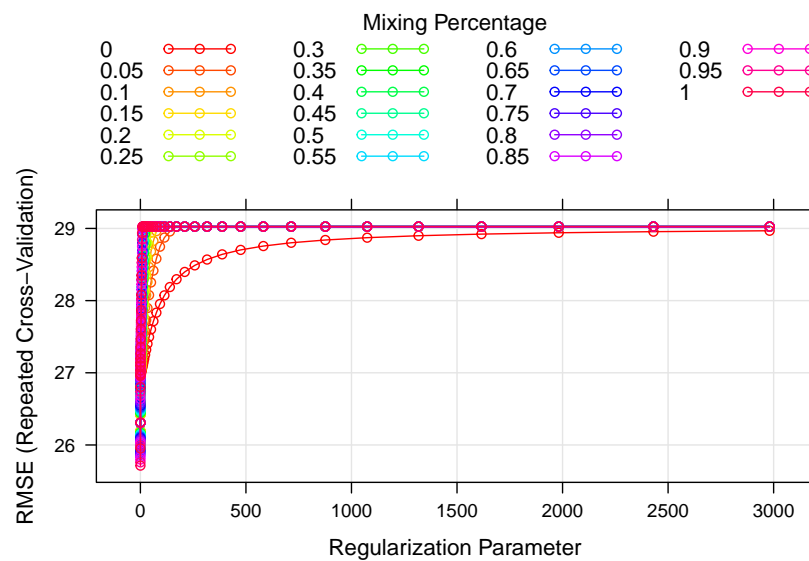


Model Training

Ordinary Least square

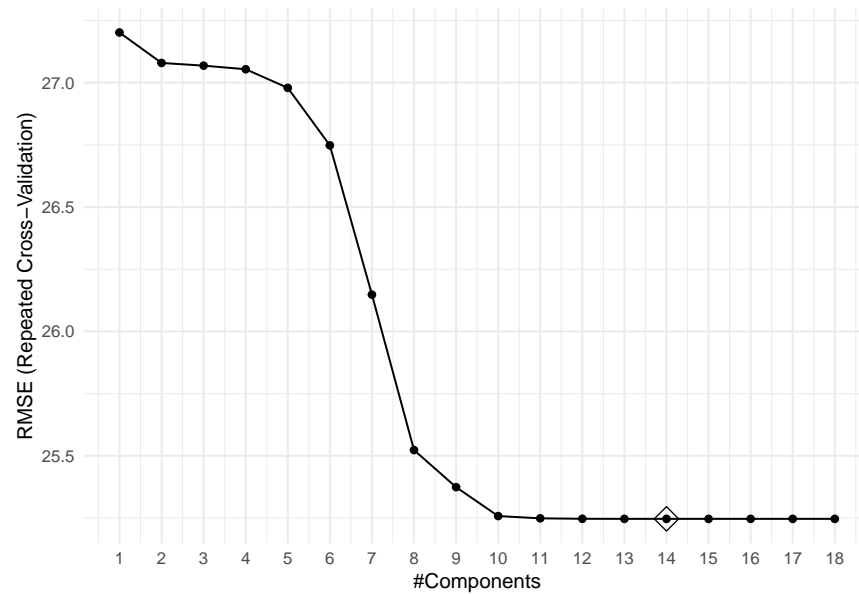


Elastic net regression

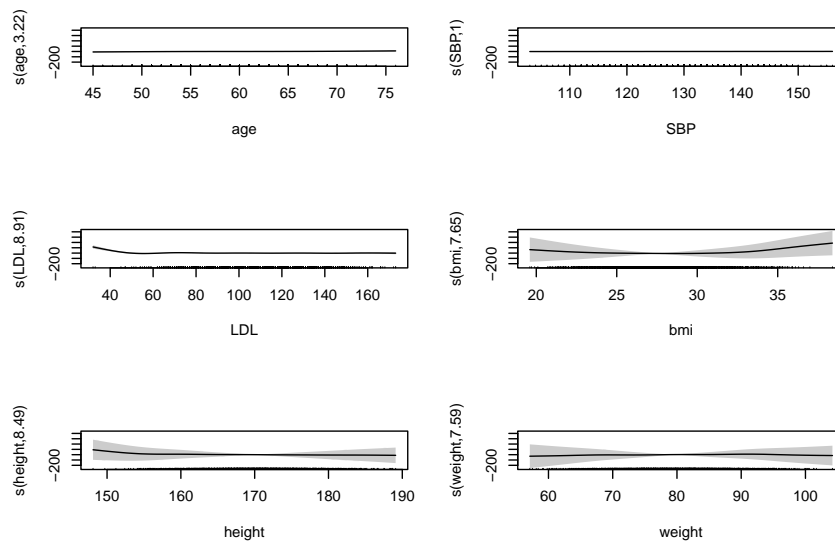


```
##      alpha      lambda
## 1001      1 0.1353353
```

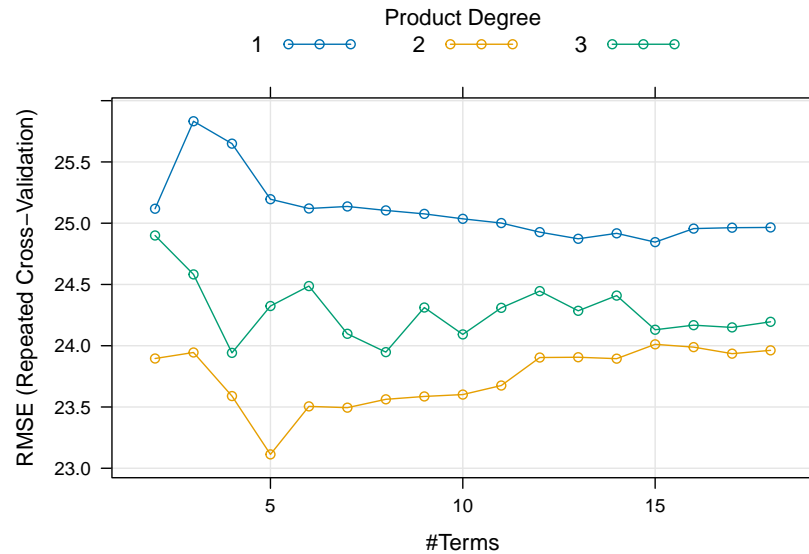
Partial least squares (PLS)



Generalized Additive Models (GAM)



Multivariate adaptive regression spline model (MARS)

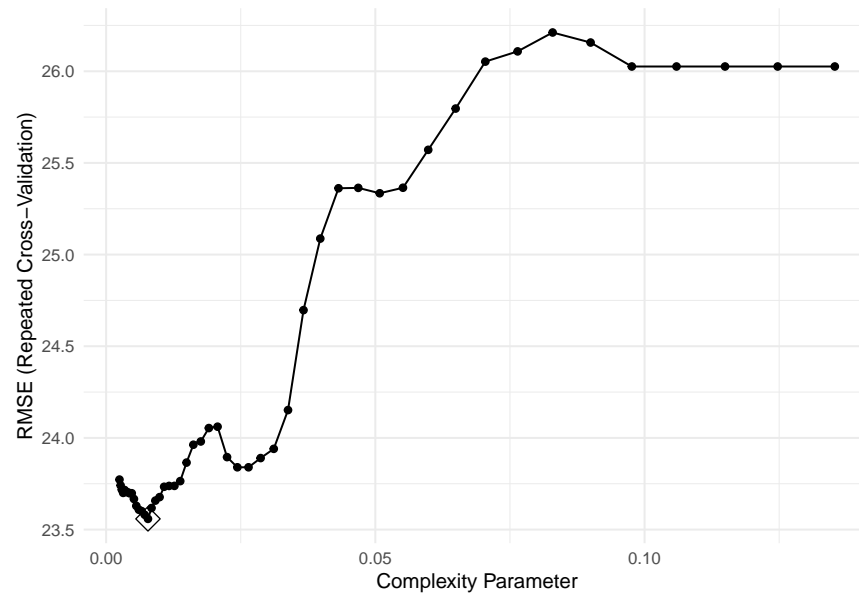


```
## Call: earth(x=matrix[2884,18], y=c(17,33,92,20,4...), keepxy=TRUE, degree=2,
##           nprune=5)
##
##               coefficients
## (Intercept)      -2.941972
## vaccine1         -9.506054
## h(bmi-24.5)       7.647737
## h(30.6-bmi)       7.320881
## h(bmi-30.6) * studyB 20.811559
##
## Selected 5 of 24 terms, and 3 of 18 predictors (nprune=5)
## Termination condition: Reached nk 37
## Importance: bmi, studyB, vaccine1, age-unused, gender1-unused, ...
## Number of terms at each degree of interaction: 1 3 1
## GCV 516.0731    RSS 1477023    GRSq 0.4041199    RSq 0.4082465
```

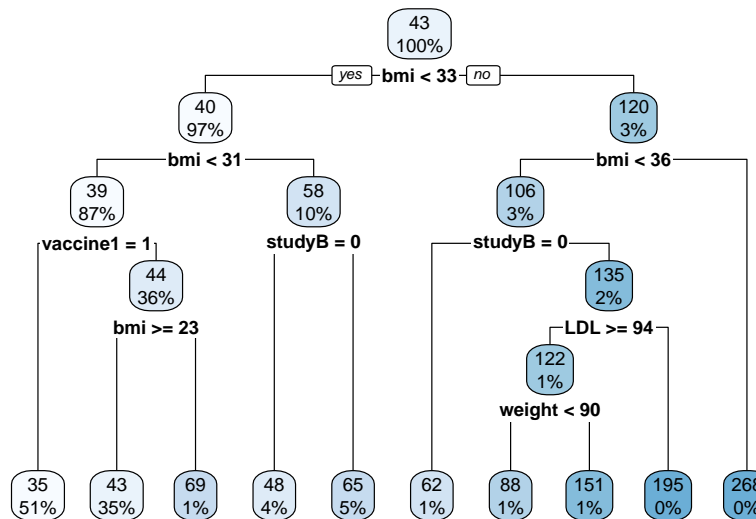
K-Nearest Neighbors (KNN)

```
##      k
## 9 13
```

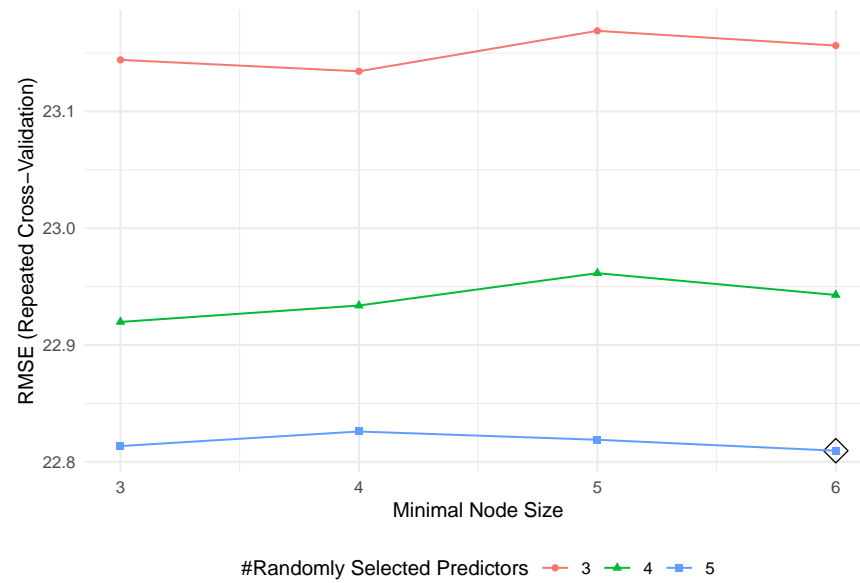
Regression Tree (CART)



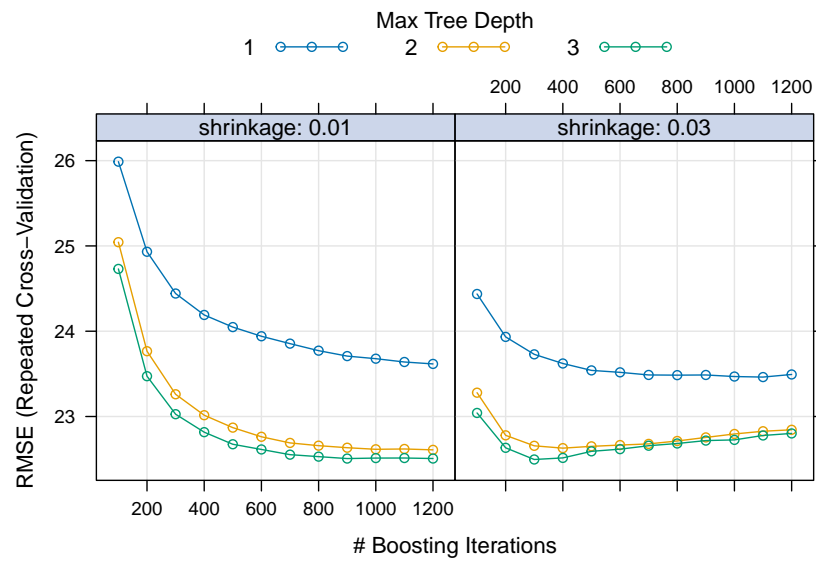
```
##          cp
## 15 0.00777266
```



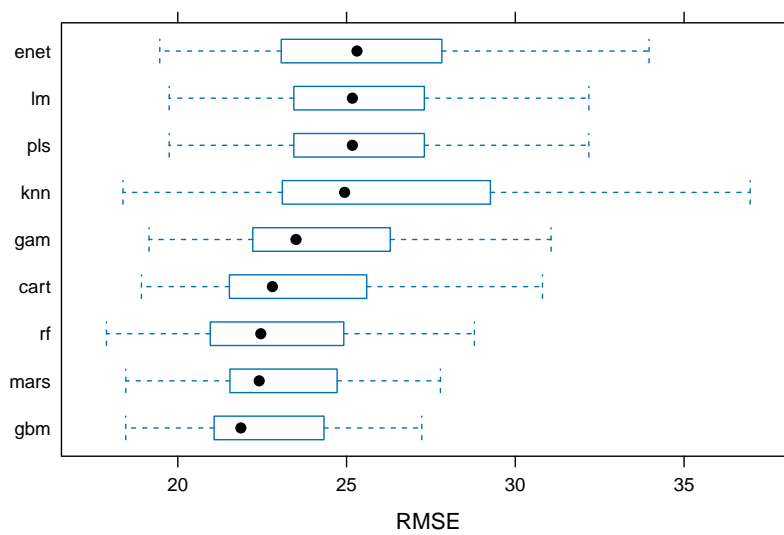
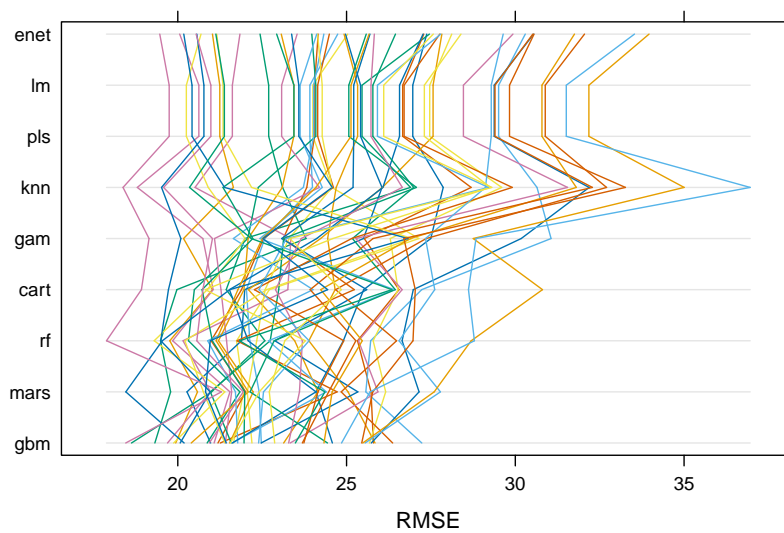
Random Forest



Generalized Boosted Regression



Models comparsion based on cross validation error

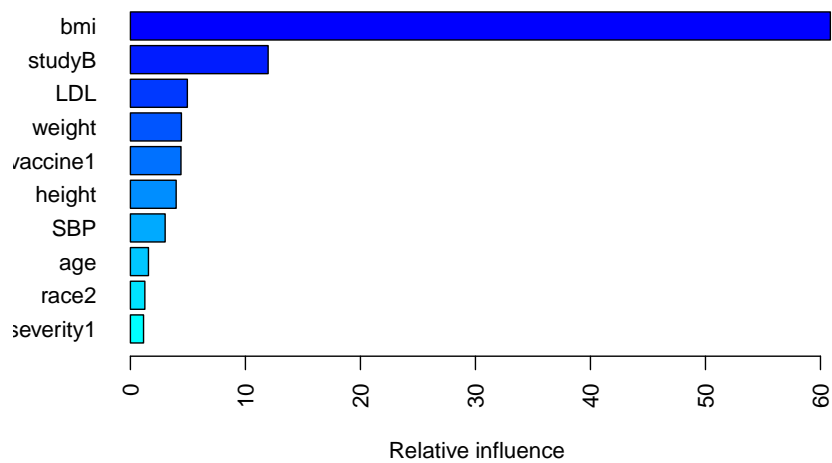


Results

Test Mean Squared Error

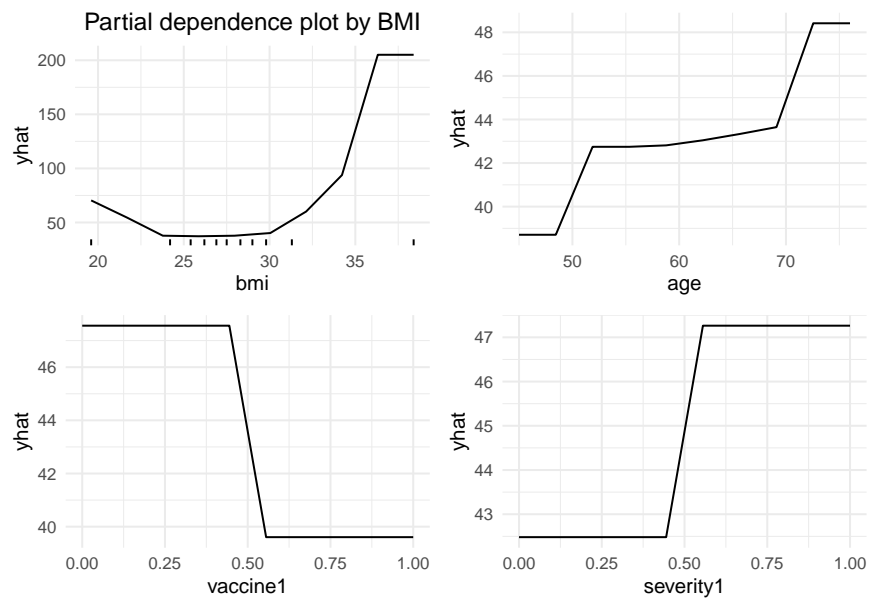
```
## [1] 492.5701
```


Variable importance plots



	rel.inf
bmi	60.8581546
studyB	11.9739399
LDL	4.9517050
weight	4.4347751
vaccine1	4.3964715
height	3.9777372
SBP	3.0197676
age	1.5690459
race2	1.2559401
severity1	1.1426109
gender1	1.0789580
smoking2	0.5346932
smoking1	0.3760449
hypertension1	0.1605474
diabetes1	0.1503396
race4	0.1009740
race3	0.0182951
studyC	0.0000000

Partial dependence plots



Individual Conditional Expectation (ICE) plot

