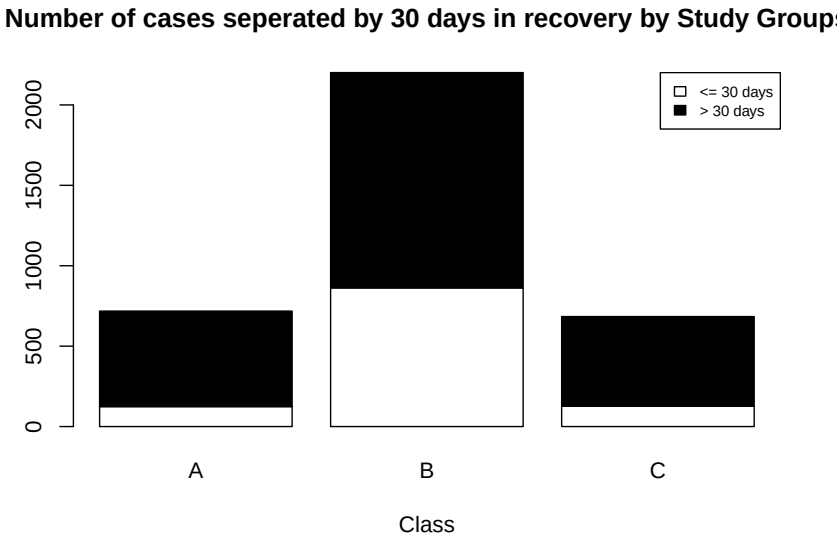
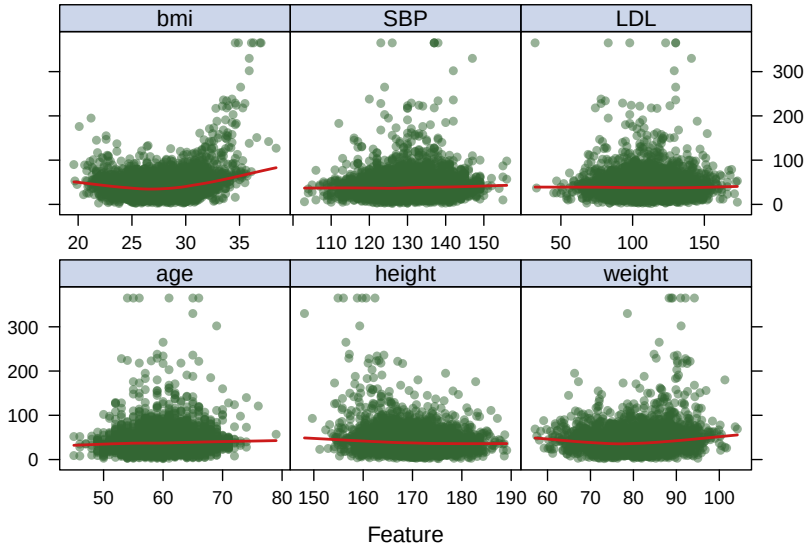
Appendix – Primary Analysis of Covid Recovery Time

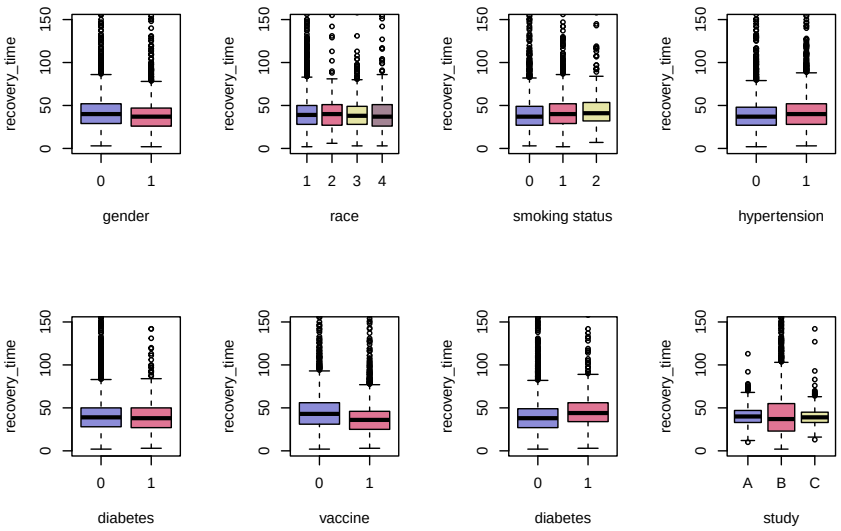
Exploratory analysis and data visualization



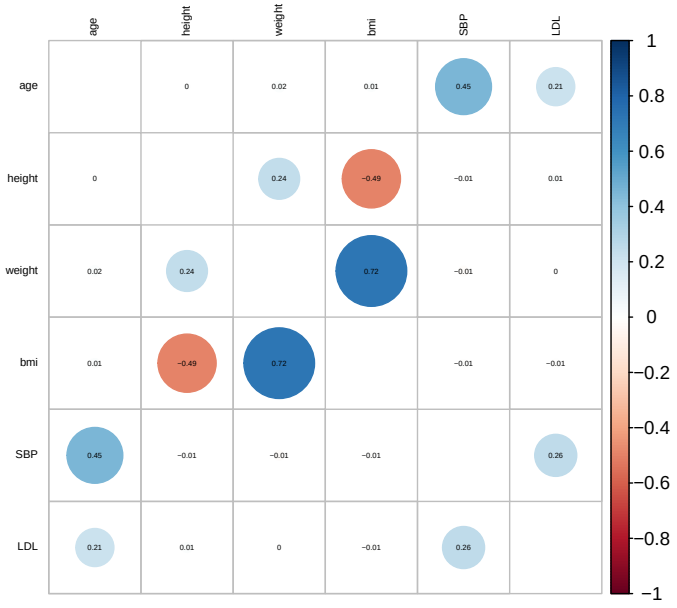
Visualize potential relationship between reponse variable and numeric predictors



Visualize potential relationship between reponse variable and categorical predictors

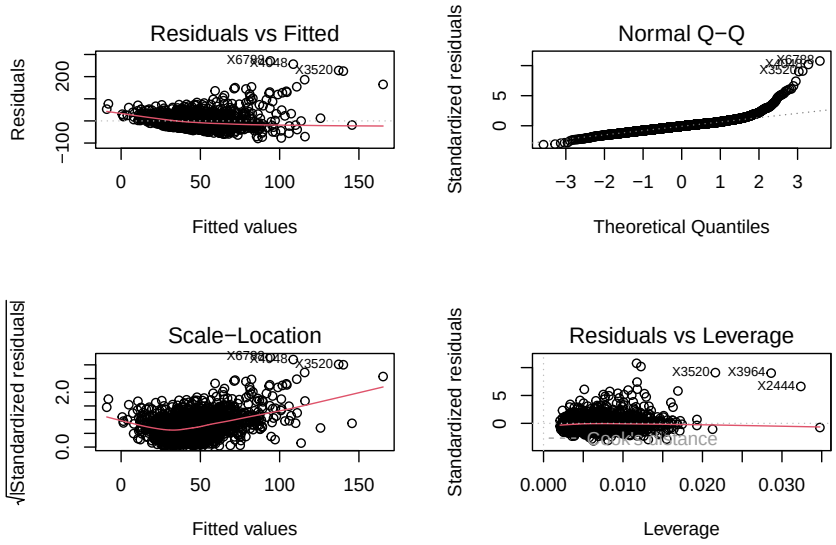


Correlation plot to check collinearity between covariates (based on training data)

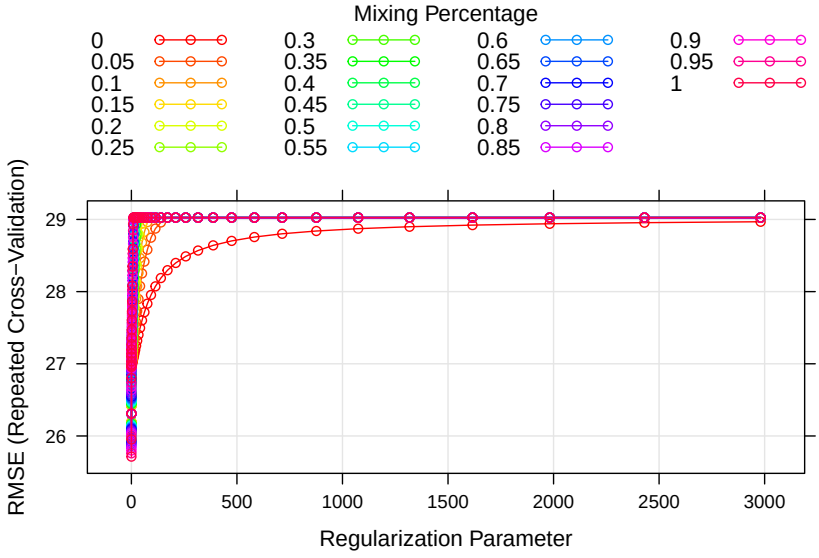


Model Training

Ordinary Least square

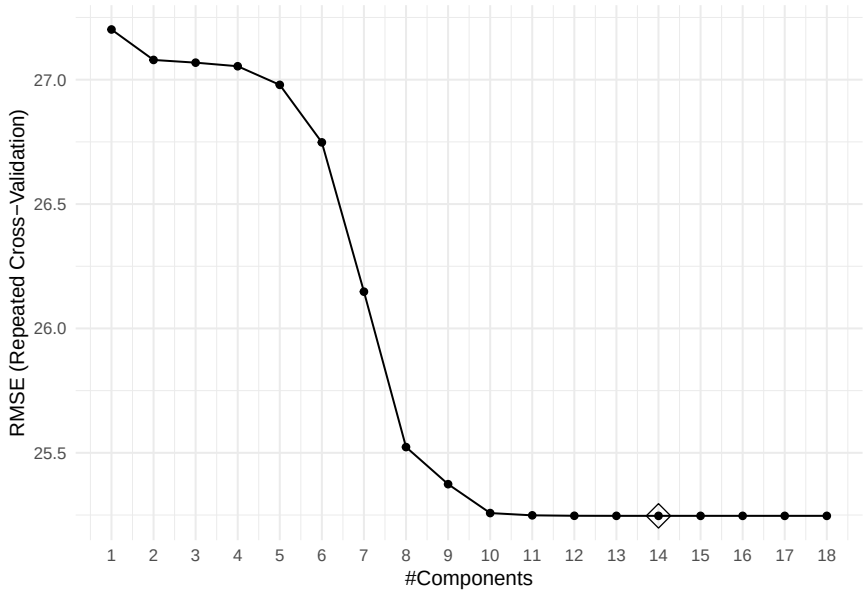


Elastic net regression

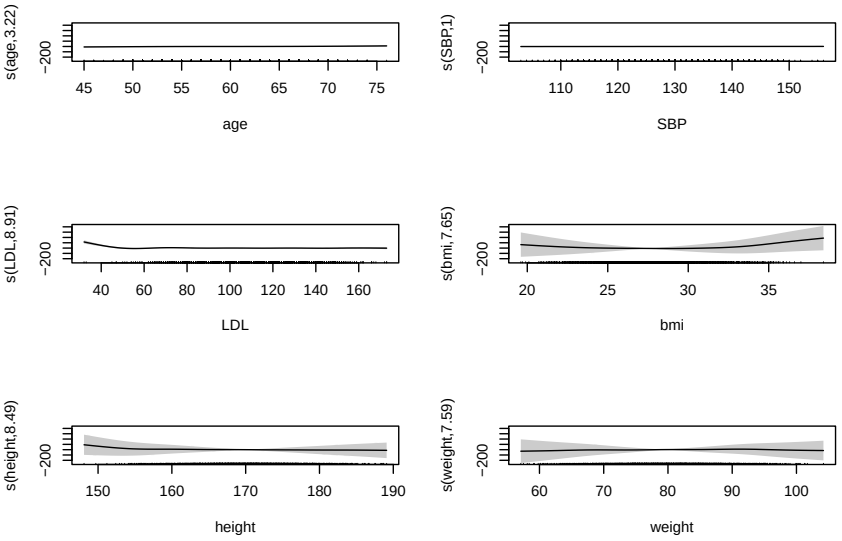



```
##      alpha    lambda
## 1001     1 0.1353353
```

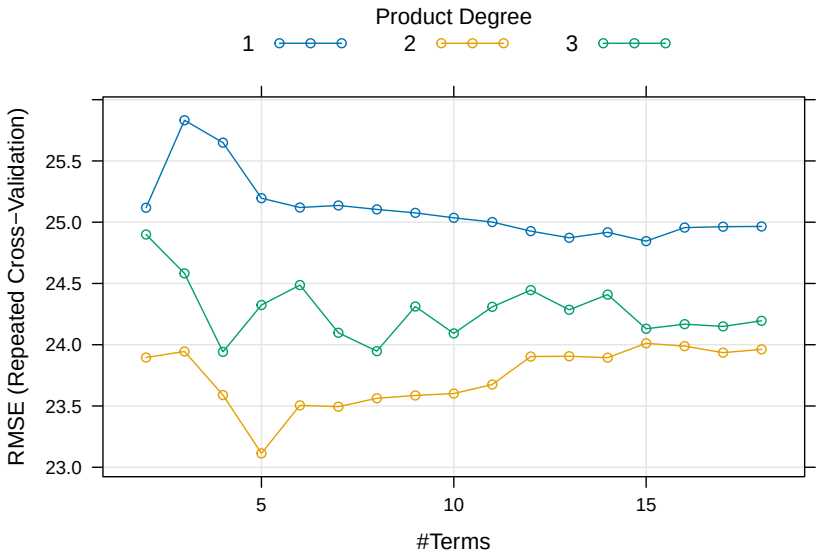
Partial least squares (PLS)



Generalized Additive Models (GAM)



Multivariate adaptive regression spline model (MARS)

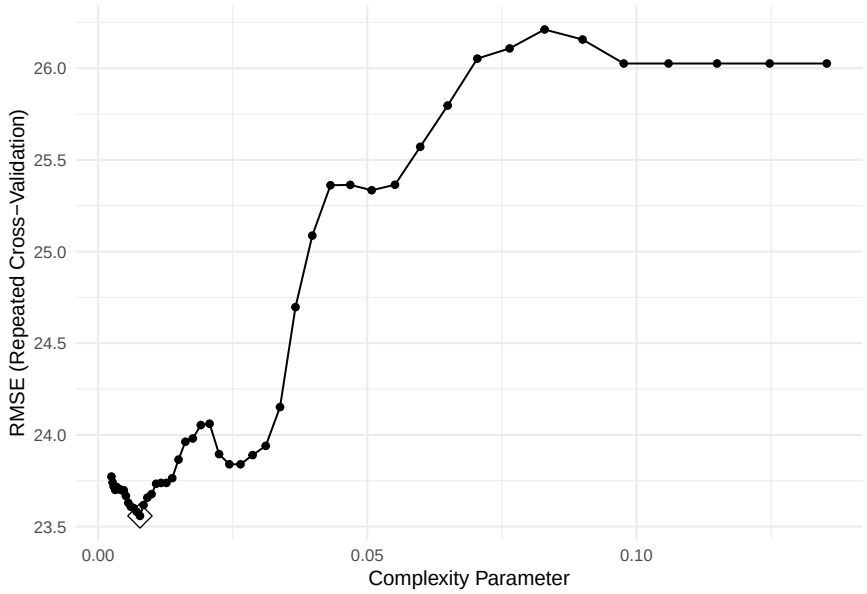



```
## Call: earth(x=matrix[2884,18], y=c(17,33,92,20,4...), keepxy=TRUE, degree=2,
##             nprune=5)
##
##                       coefficients
## (Intercept)              -2.941972
## vaccine1                 -9.506054
## h(bmi-24.5)               7.647737
## h(30.6-bmi)               7.320881
## 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
```

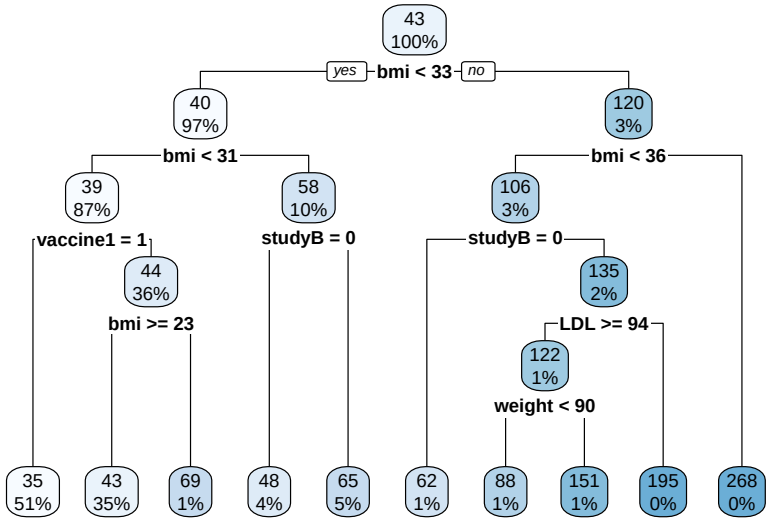
K-Nearest Neighbors (KNN)

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

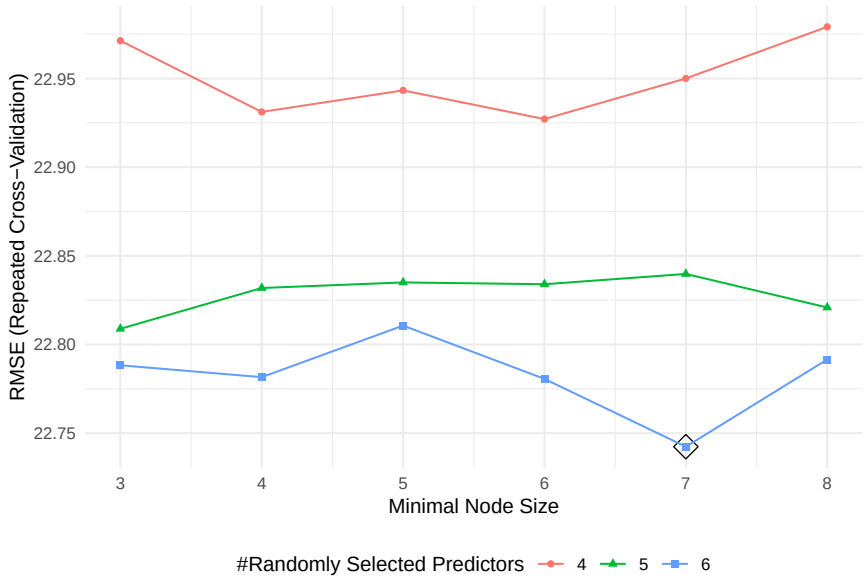
Regression Tree (CART)



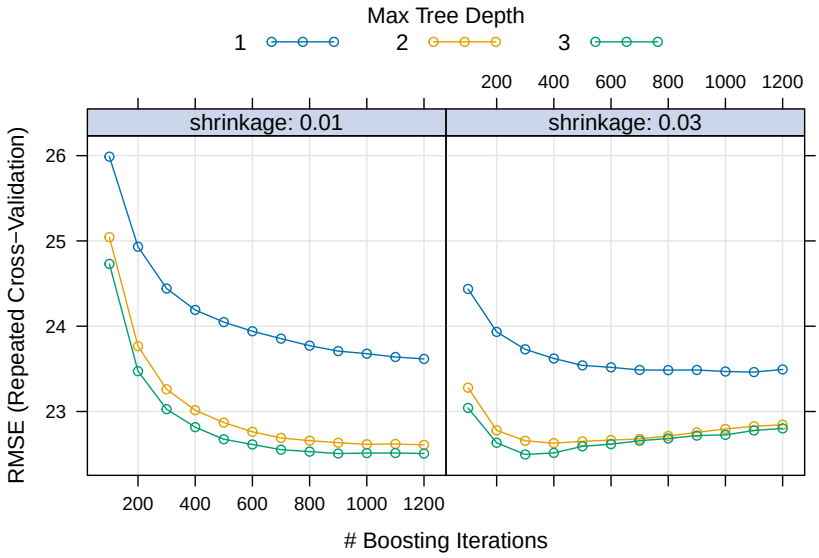

```
##           cp
## 15 0.00777266
```



Random Forest



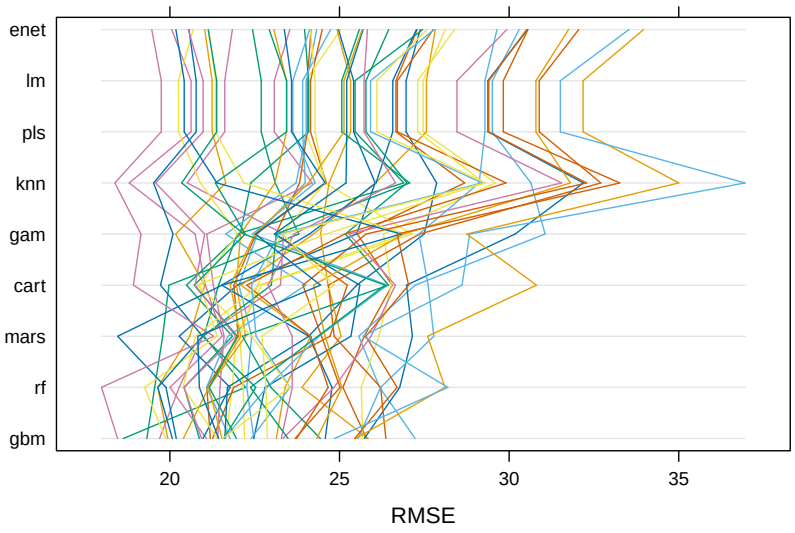
Generalized Boosted Regression

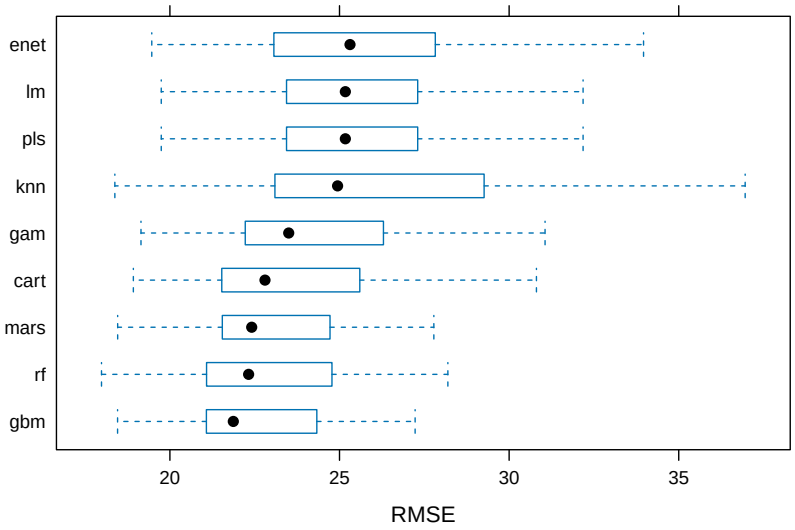


Models comparsion based on cross validation error

```
##
## Call:
## summary.resamples(object = resamp)
##
## Models: lm, enet, pls, gam, mars, knn, gbm, cart, rf
## Number of resamples: 50
```

```
##
## MAE
##          Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## lm   15.16547 16.10302 16.78177 16.83446 17.42209 19.04602    0
## enet 14.56452 15.92803 16.52216 16.46752 17.01105 18.64227    0
## pls  15.16548 16.10302 16.78177 16.83446 17.42208 19.04604    0
## gam  14.35229 15.33438 15.83680 15.95120 16.56241 17.79521    0
## mars 13.93931 14.88663 15.26803 15.37548 15.95721 17.55273    0
## knn  13.77991 14.91692 15.66110 15.73718 16.40355 18.31290    0
## gbm  13.16340 14.34295 14.85952 14.96472 15.45562 16.75934    0
## cart 14.24679 14.87635 15.29857 15.47210 16.11928 17.75836    0
## rf   13.66061 14.41903 14.83788 15.08850 15.72662 17.32203    0
##
## RMSE
##          Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## lm   19.74488 23.43892 25.17216 25.24668 27.21931 32.18024    0
## enet 19.46546 23.14204 25.30888 25.71390 27.81579 33.96273    0
## pls  19.74490 23.43892 25.17216 25.24668 27.21931 32.18025    0
## gam  19.14744 22.27447 23.50243 24.21183 26.16733 31.05932    0
## mars 18.46040 21.55759 22.41118 23.11308 24.68399 27.78102    0
## knn  18.37738 23.09791 24.94296 26.01258 29.24863 36.95880    0
## gbm  18.45737 21.10448 21.86910 22.49591 24.18224 27.22898    0
## cart 18.92266 21.53195 22.80303 23.55843 25.57517 30.80530    0
## rf   17.98350 21.08297 22.32057 22.74234 24.74996 28.19421    0
##
## Rsquared
##             Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## lm   0.08620897 0.1900059 0.2572871 0.2531044 0.3099187 0.4654627    0
## enet 0.08078819 0.1811701 0.2277041 0.2290413 0.2731629 0.3775725    0
## pls  0.08620900 0.1900061 0.2572879 0.2531045 0.3099185 0.4654625    0
## gam  0.12269099 0.2416508 0.3294108 0.3350657 0.4183874 0.6223217    0
## mars 0.07601256 0.2807337 0.3859944 0.3750641 0.4765129 0.6877643    0
## knn  0.06268675 0.1685928 0.2095025 0.2280360 0.2800124 0.4727332    0
## gbm  0.07050574 0.3111809 0.4021608 0.4026223 0.4940622 0.7041863    0
## cart 0.05216633 0.2359045 0.3635301 0.3517752 0.4471816 0.6223333    0
## rf   0.09295909 0.3246594 0.3947281 0.3929464 0.4732256 0.6063936    0
```



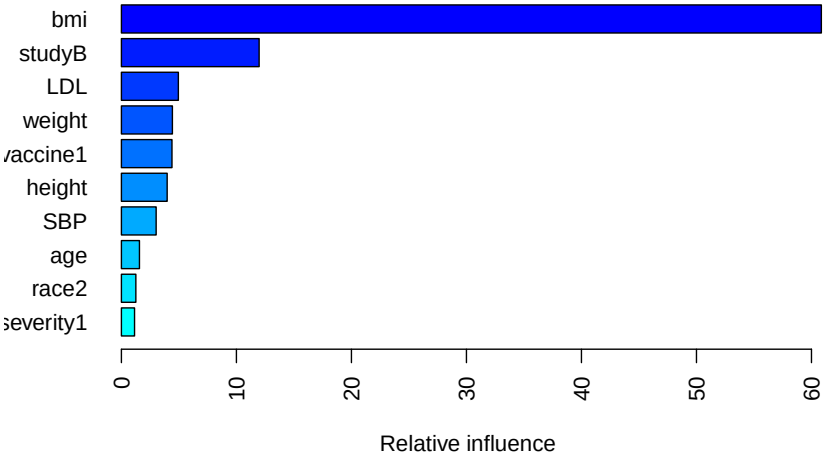


Results

Test Mean Squared Error

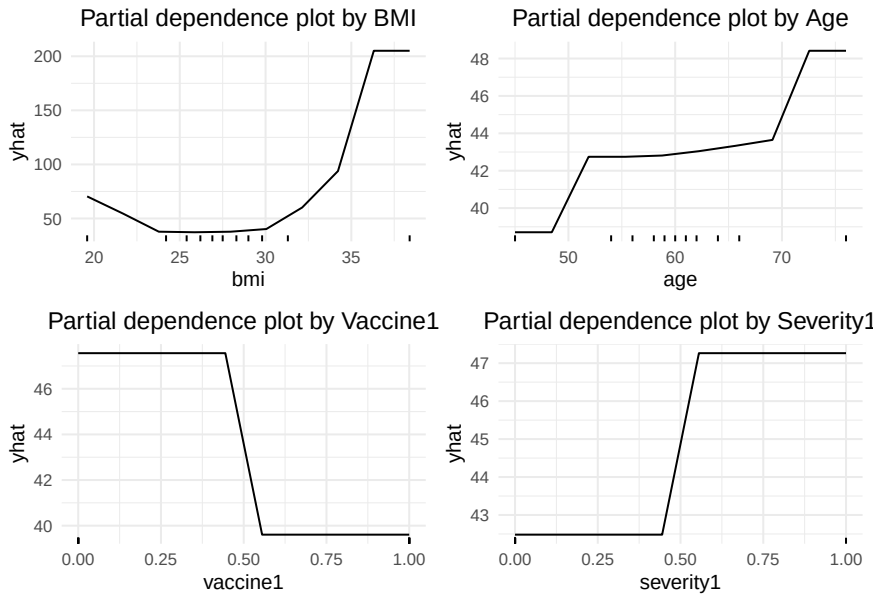
```
## [1] 492.5701
```

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 dependance plots



Individual Conditional Expectation (ICE) plot