

Figures and outputs

Exploratory analysis and data visualization

Table 1: Data summary

Name	Piped data
Number of rows	2000
Number of columns	16
Column type frequency:	
factor	8
numeric	8
Group variables	None

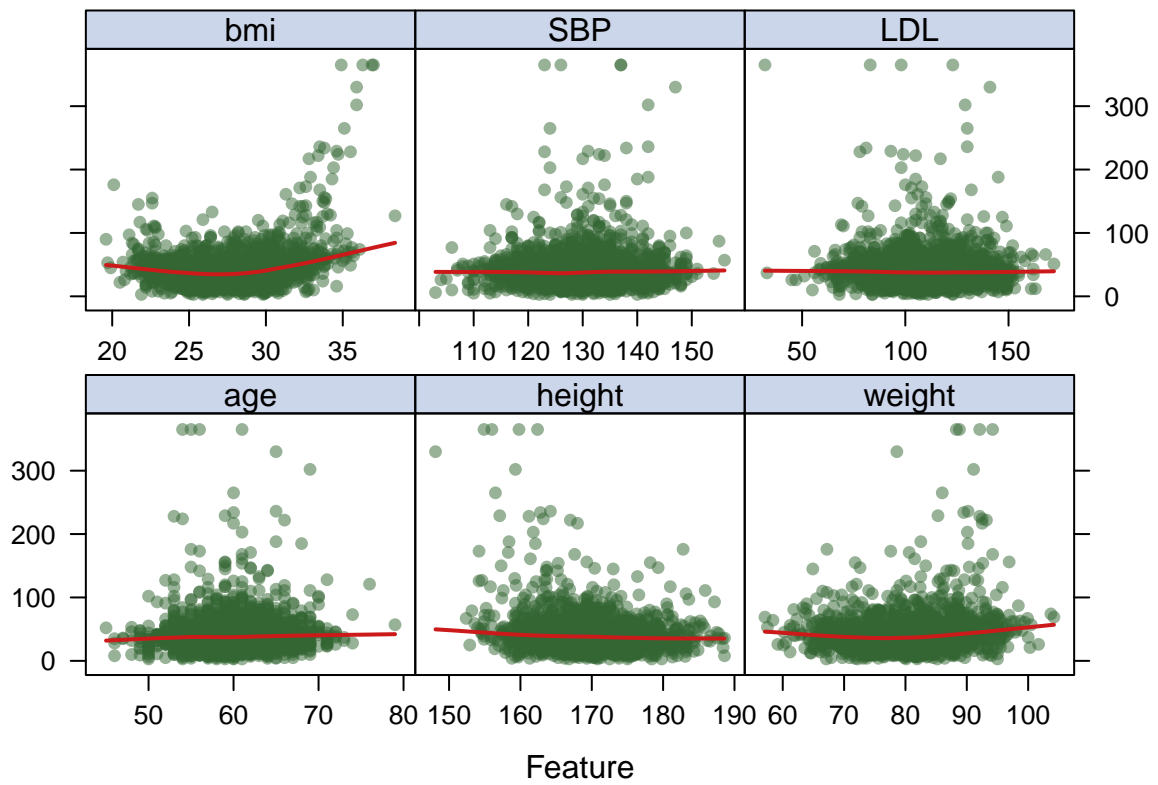
Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
gender	0	1	FALSE	2	0: 1049, 1: 951
race	0	1	FALSE	4	1: 1304, 3: 406, 4: 197, 2: 93
smoking	0	1	FALSE	3	0: 1221, 1: 588, 2: 191
hypertension	0	1	FALSE	2	0: 1053, 1: 947
diabetes	0	1	FALSE	2	0: 1699, 1: 301
vaccine	0	1	FALSE	2	1: 1168, 0: 832
severity	0	1	FALSE	2	0: 1807, 1: 193
study	0	1	FALSE	3	B: 1229, A: 390, C: 381

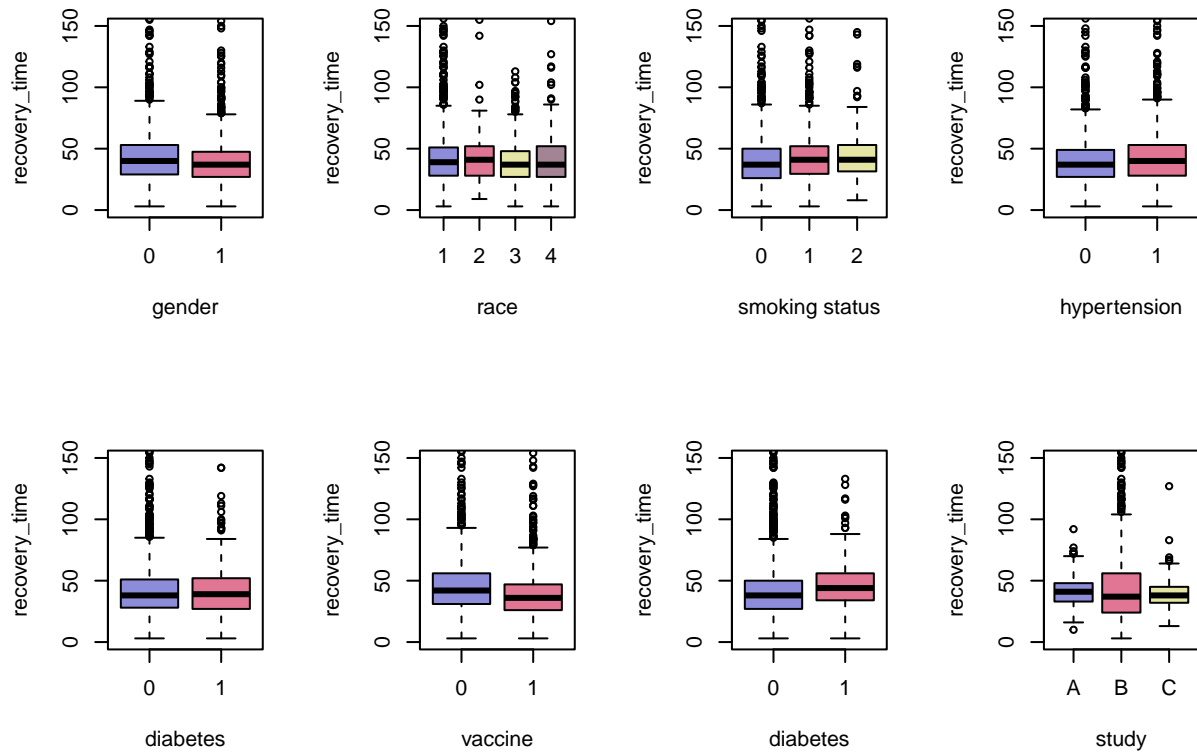
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
id	0	1	4995.42	2832.70	2.0	2603.25	4966.5	7424.5	9999.0	
age	0	1	60.16	4.47	45.0	57.00	60.0	63.0	79.0	
height	0	1	170.00	6.00	148.1	166.00	170.0	173.9	188.6	
weight	0	1	79.96	7.20	57.1	75.10	79.9	84.8	104.2	
bmi	0	1	27.73	2.82	19.6	25.80	27.6	29.5	38.4	
SBP	0	1	129.92	7.95	103.0	125.00	130.0	135.0	156.0	
LDL	0	1	110.19	19.80	32.0	97.00	110.0	124.0	172.0	
recovery_time	0	1	43.50	31.09	3.0	28.00	39.0	51.0	365.0	

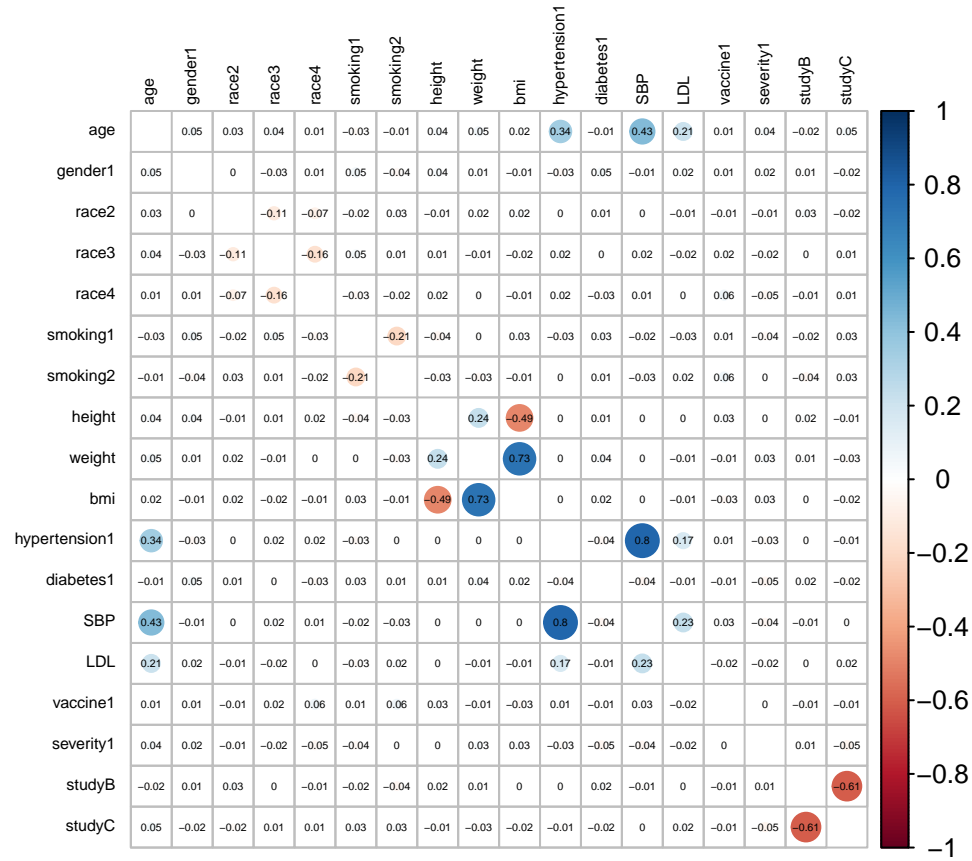
Visualize potential relationship between reponse variable and numeric predictors



Visualize potential relationship between reponse variable and categorical predictors

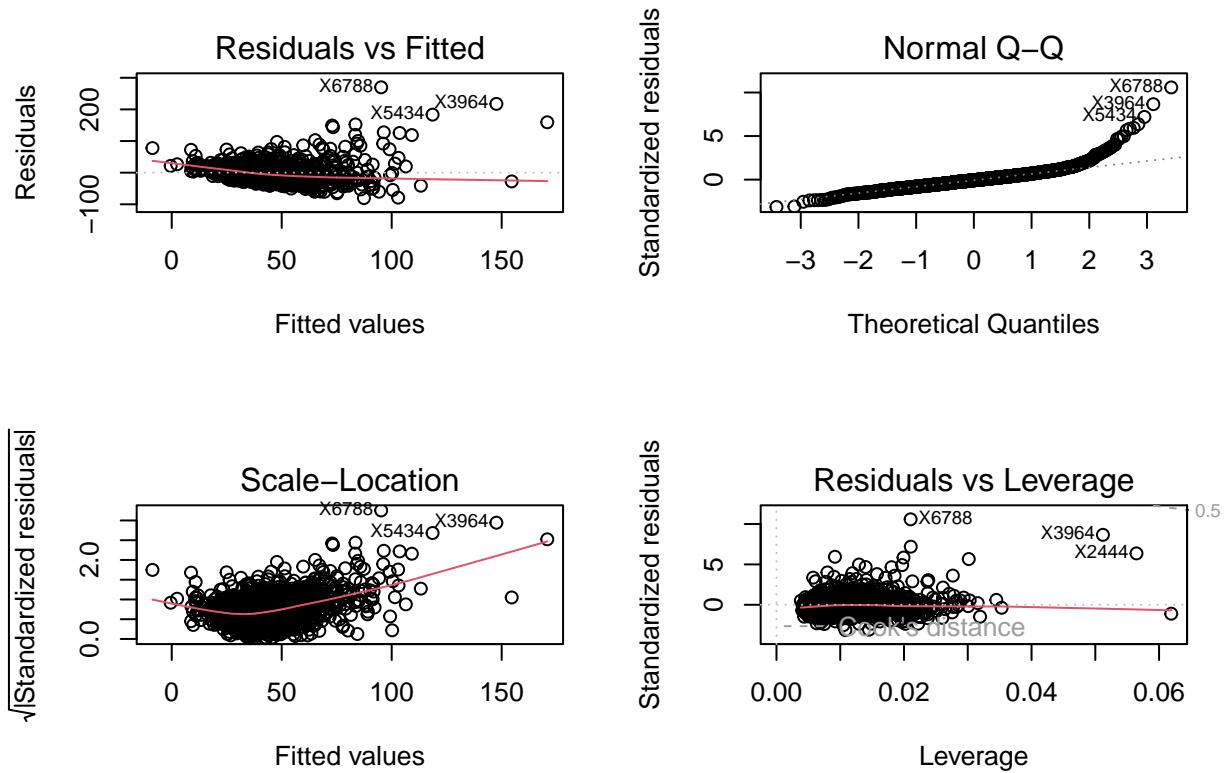


Correlation plot to check collinearity between covariates (based on training data)

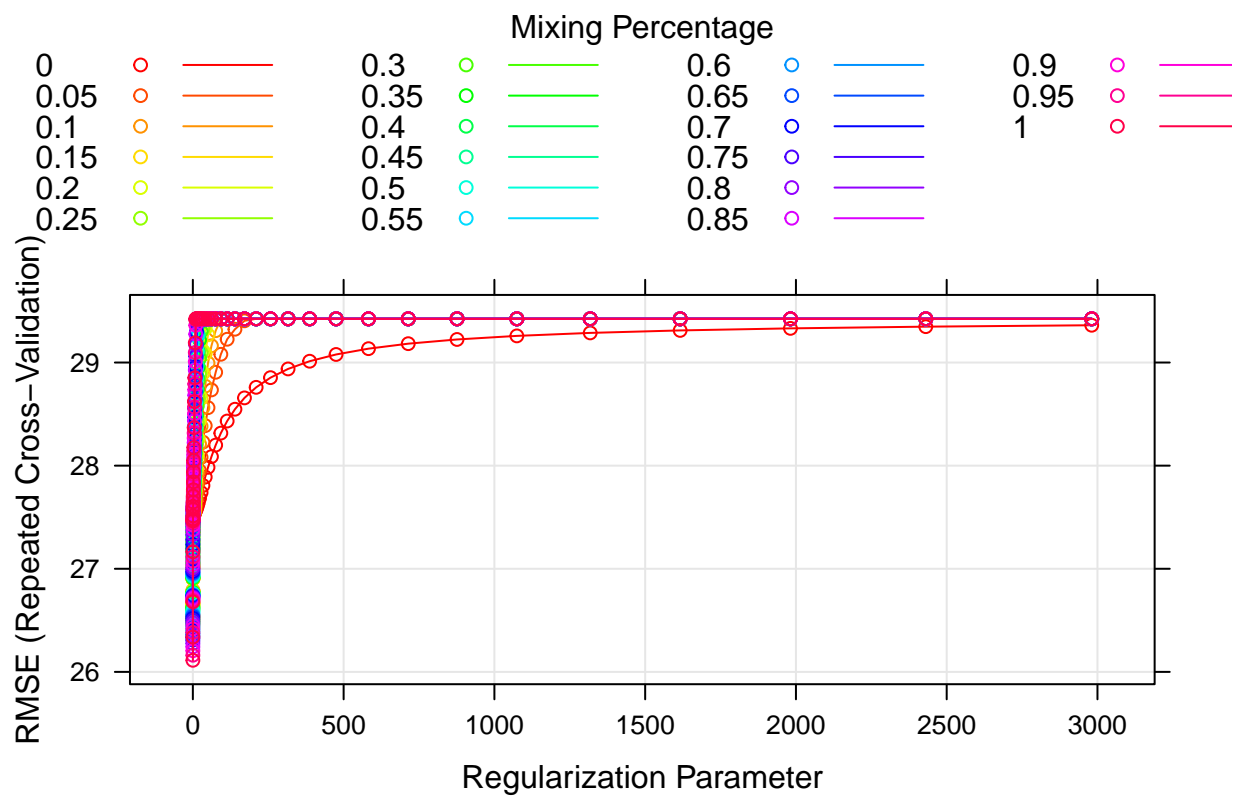


Model Training

Ordinary Least square

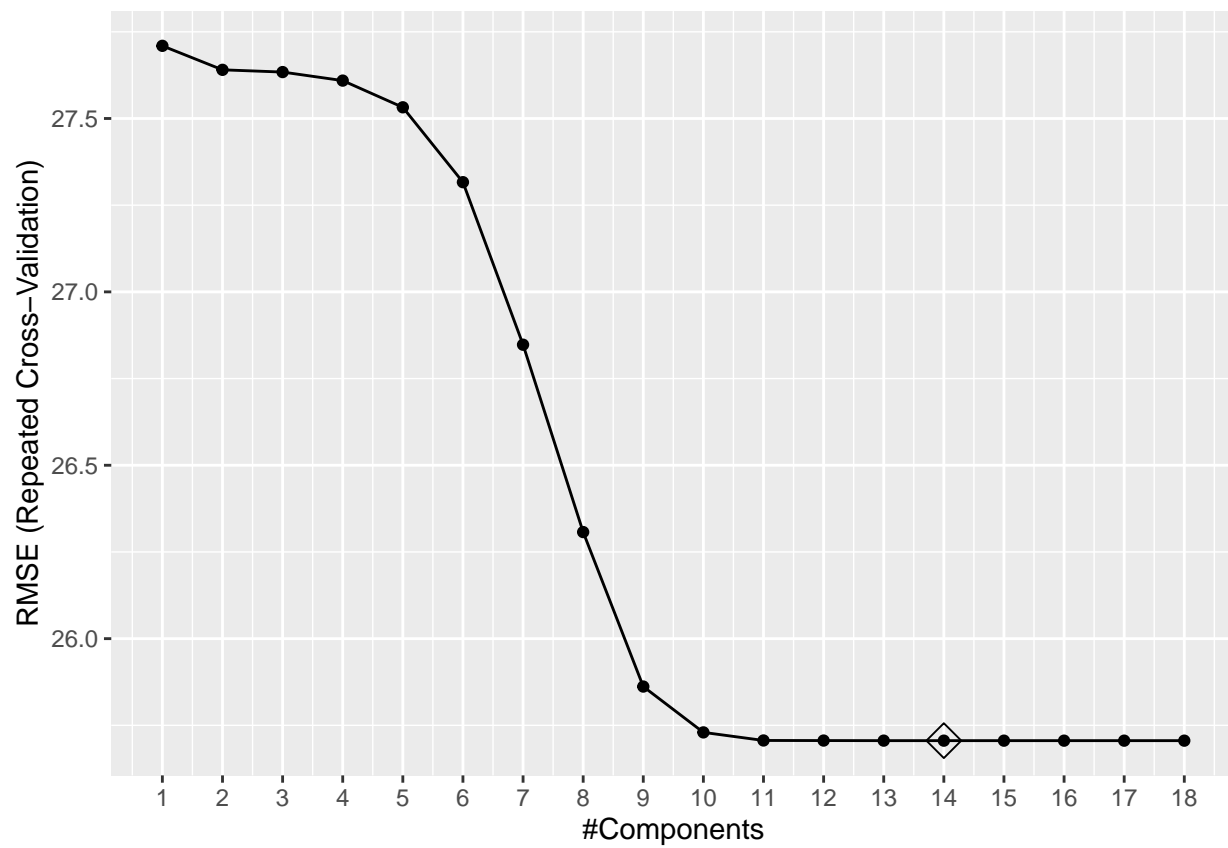


Elastic net regression

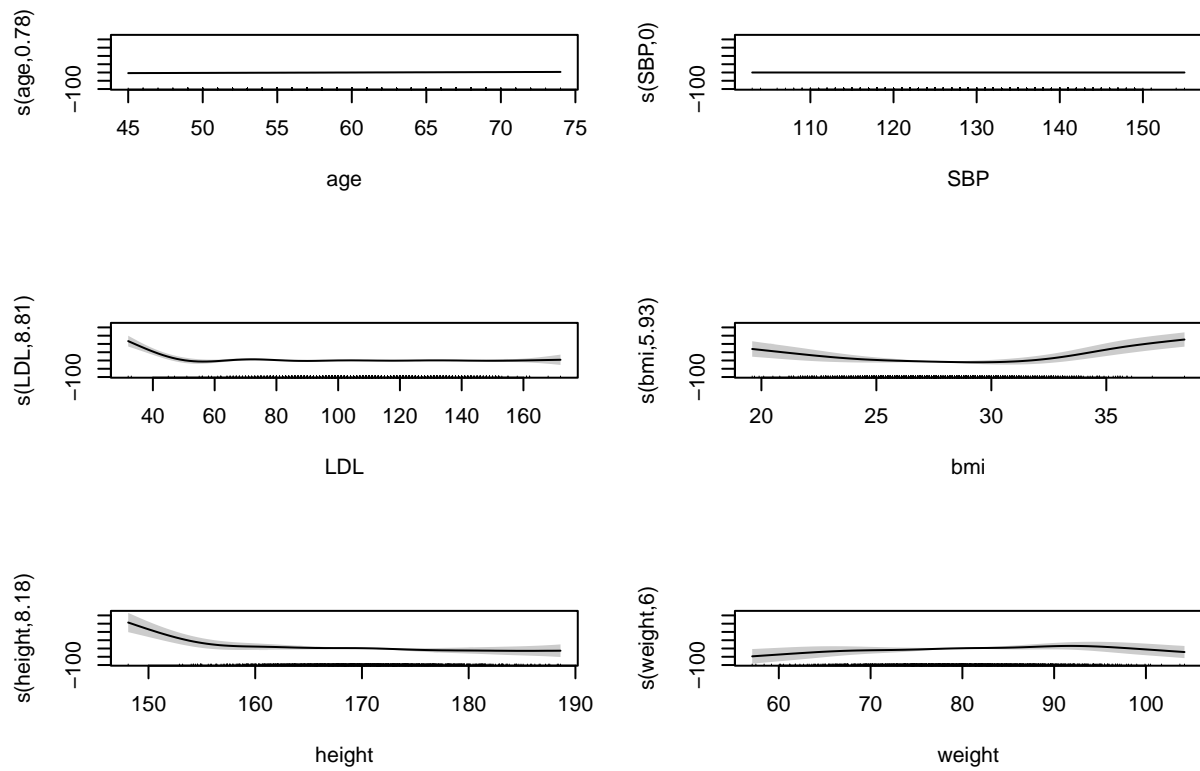


```
##      alpha      lambda
## 1001      1 0.1353353
```

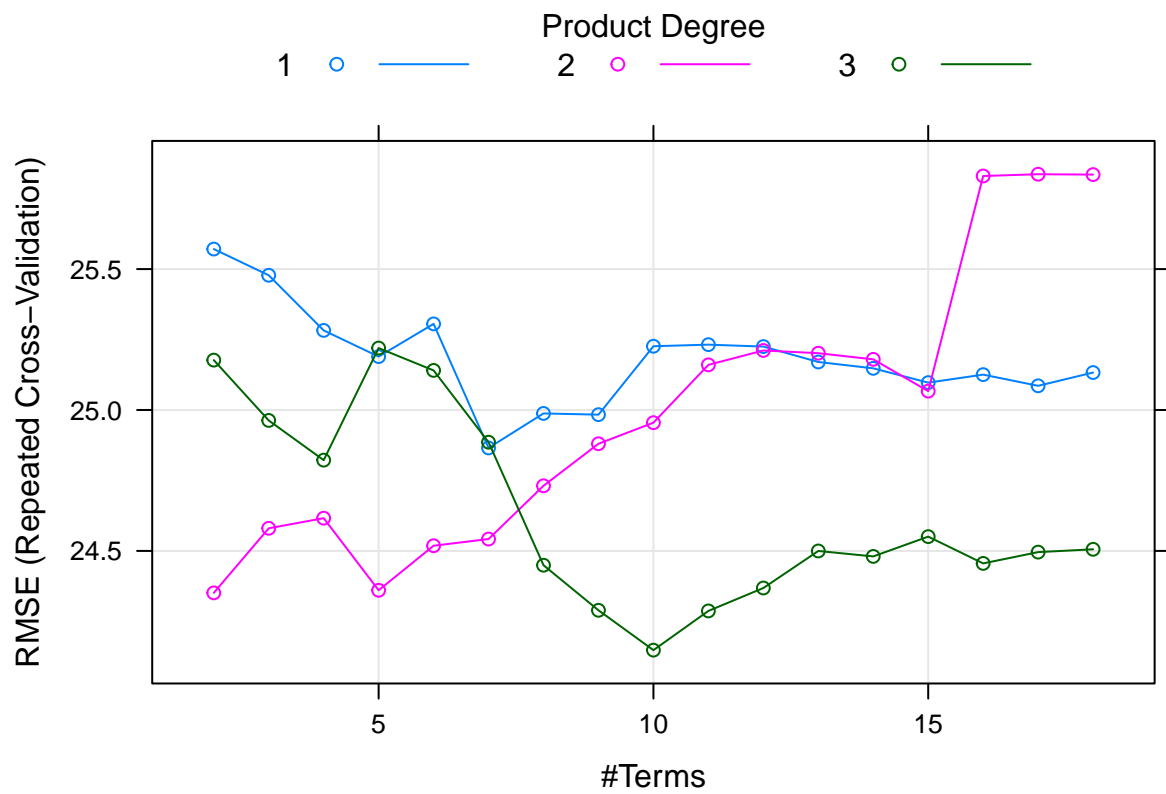
Partial least squares



Generalized Additive Models (GAM)



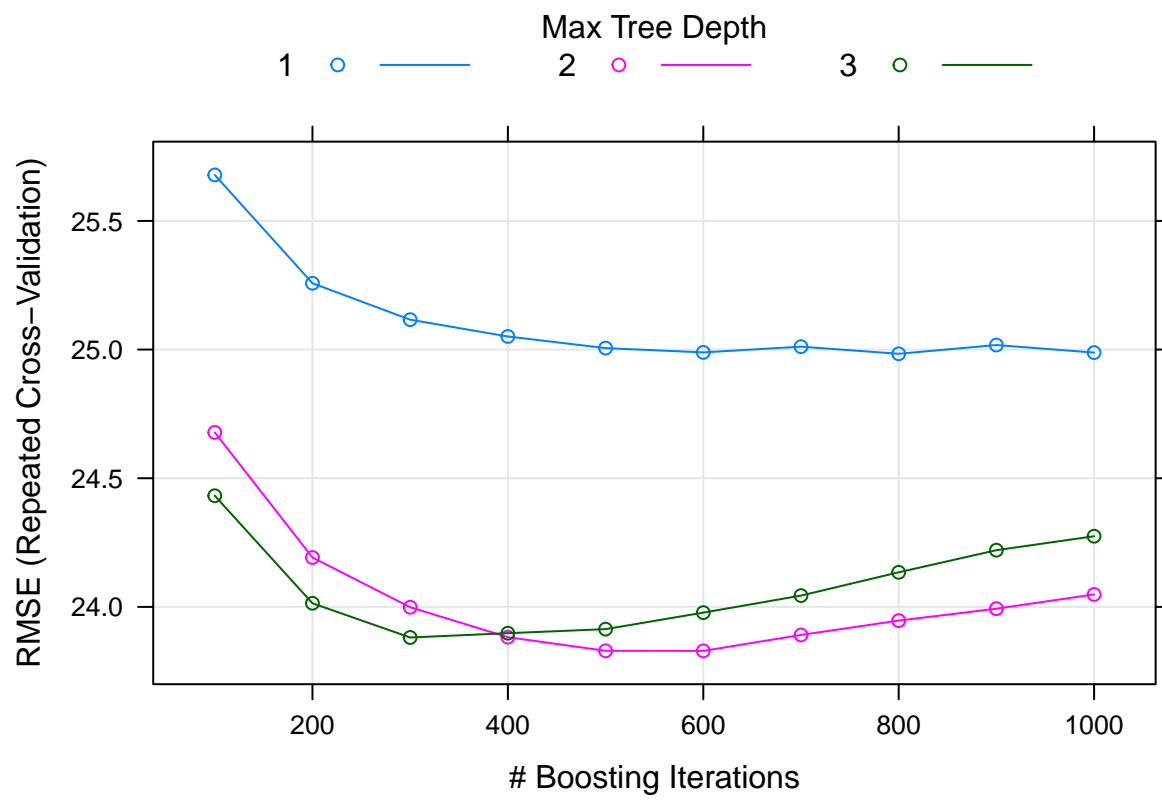
Multivariate adaptive regression spline model (MARS)



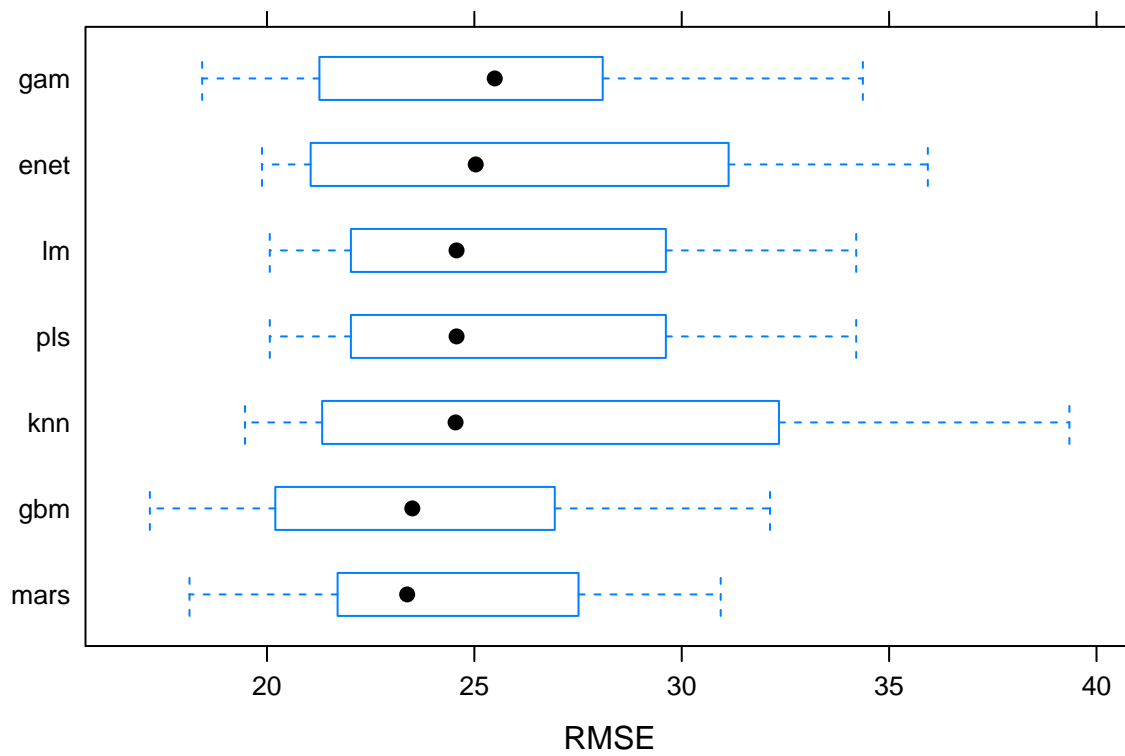
K-Nearest Neighbors (KNN)

```
##      k
## 8 12
```

Generalized Boosted Regression



Models comparsion based on cross validation error

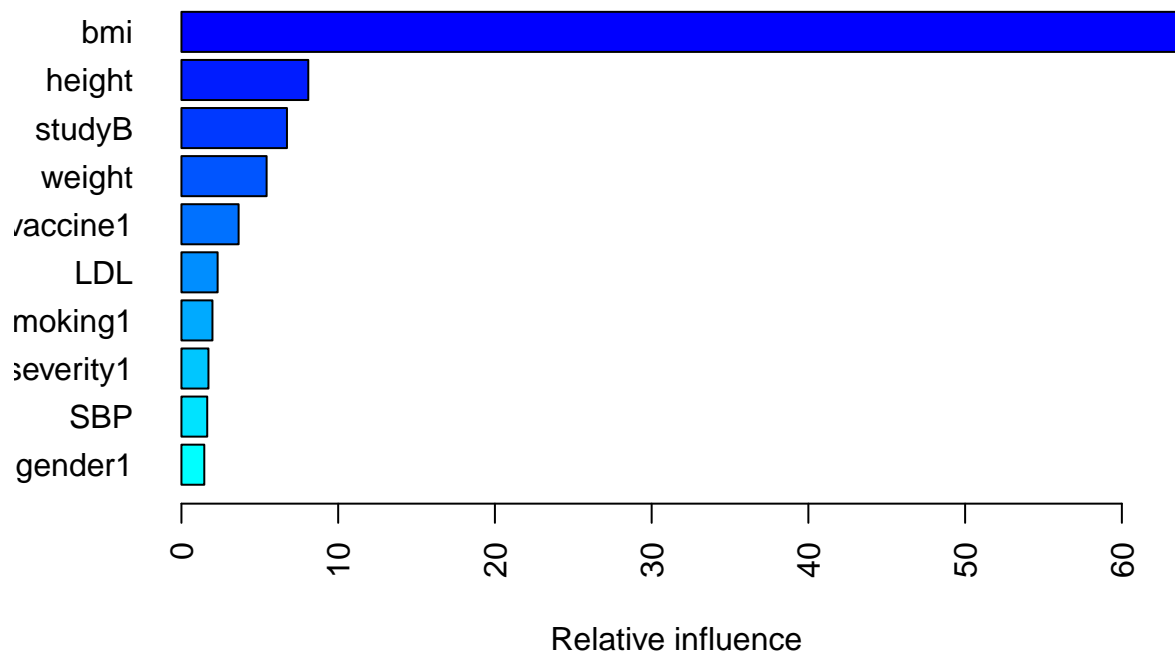


Results

Test Mean Squared Error

```
## [1] 693.644
```

Variable importance plots



	rel.inf
bmi	63.7963911
height	8.0927211
studyB	6.7327380
weight	5.4349899
vaccine1	3.6472361
LDL	2.3082765
smoking1	1.9786793
severity1	1.7236598
SBP	1.6449548
gender1	1.4572454
age	1.2712470
smoking2	0.6782509
race2	0.6052246
hypertension1	0.5740454
race4	0.0543400
race3	0.0000000
diabetes1	0.0000000
studyC	0.0000000

Partial dependance plots

