Appendix

Joe LaRocca, Brooklynn McNeil, Malika Top

2025-03-31

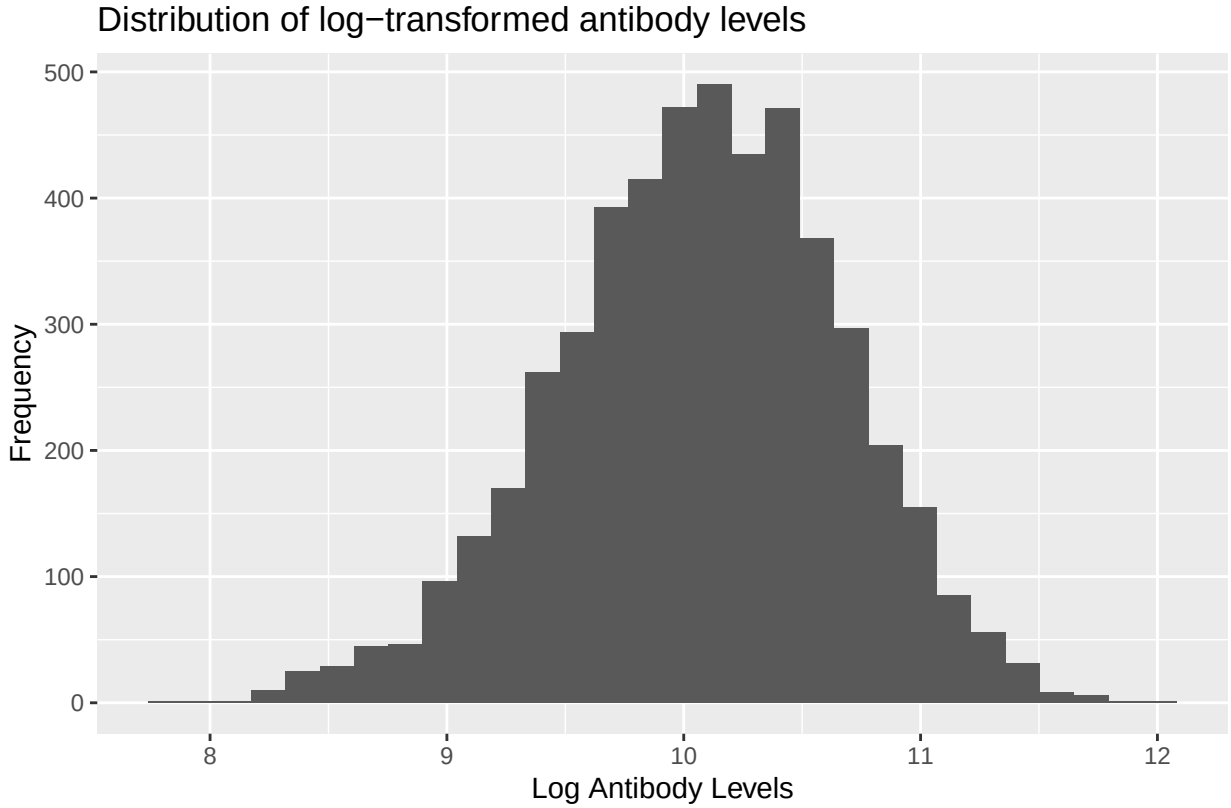
Characteristic	**N = 5,000**[1]
gender	
female	2,573 (51%)
male	2,427 (49%)
race	
white	3,221 (64%)
asian	278 (5.6%)
black	1,036 (21%)
hispanic	465 (9.3%)
smoking	
never_smoked	3,010 (60%)
former_smoker	1,504 (30%)
current_smoker	486 (9.7%)
diabetes	772 (15%)
hypertension	2,298 (46%)

[1]n (%)

Characteristic	**N = 5,000**
age	
Median (Q1, Q3)	60.0 (57.0, 63.0)
Min, Max	44.0, 75.0
height	
Median (Q1, Q3)	170.1 (166.1, 174.3)
Min, Max	150.2, 192.9
weight	
Median (Q1, Q3)	80 (75, 85)
Min, Max	57, 106
bmi	

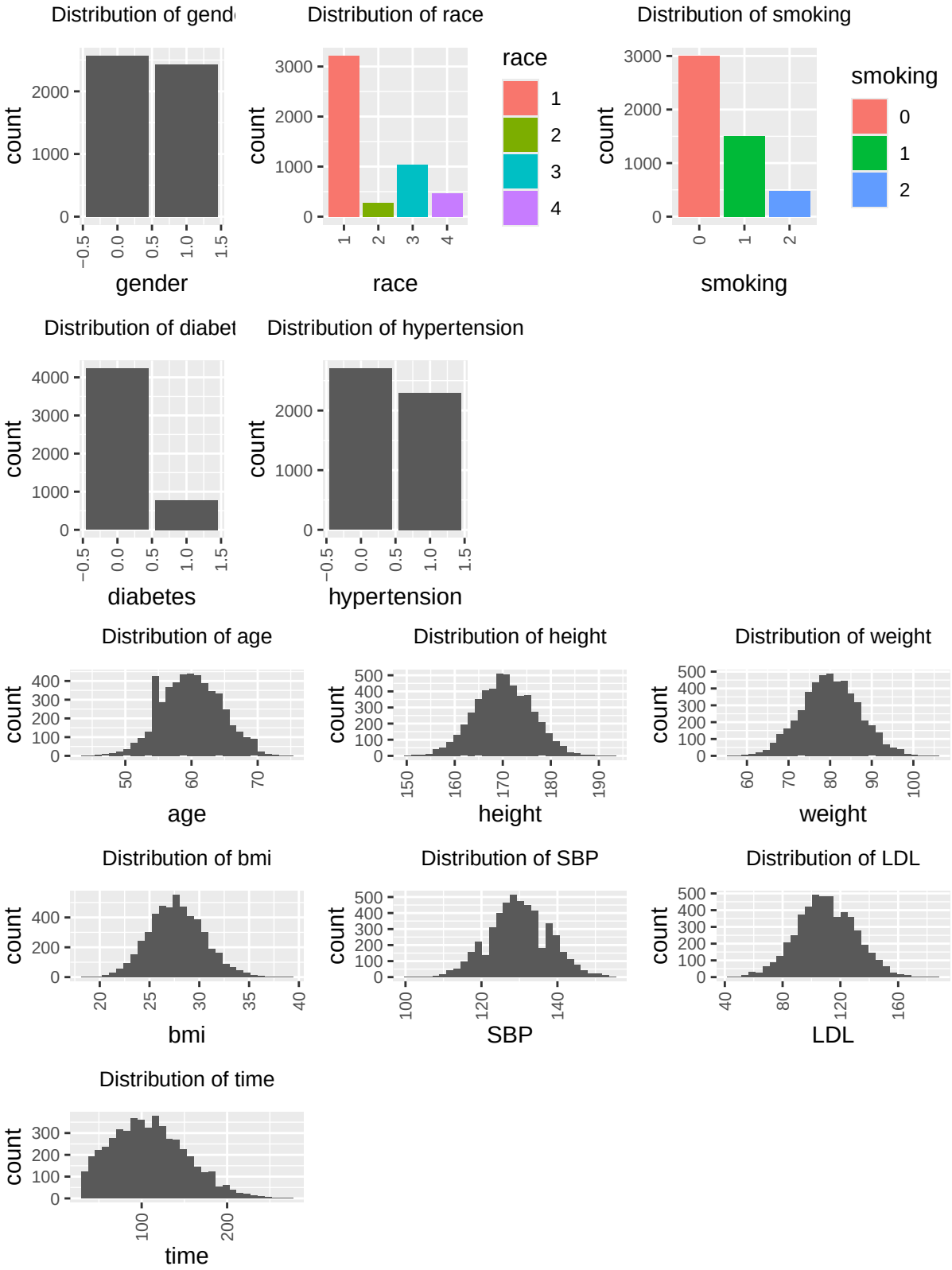
Characteristic	N = 5,000
Median (Q1, Q3)	27.60 (25.80, 29.50)
Min, Max	18.20, 38.80
SBP	
Median (Q1, Q3)	130 (124, 135)
Min, Max	101, 155
LDL	
Median (Q1, Q3)	110 (96, 124)
Min, Max	43, 185
time	
Median (Q1, Q3)	106 (76, 138)
Min, Max	30, 270

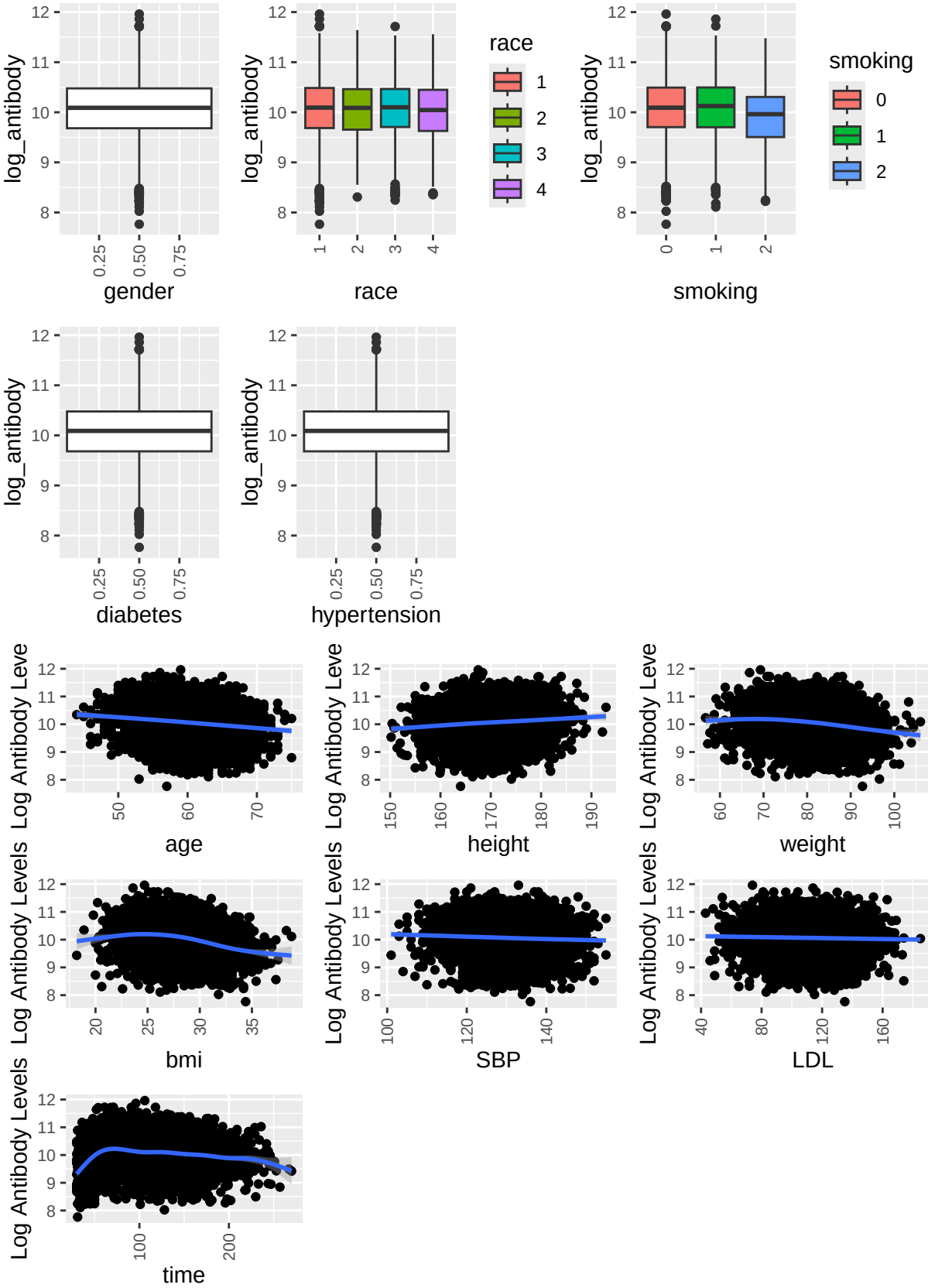
Characteristic	N = 5,000
log_antibody	
Median (Q1, Q3)	10.09 (9.68, 10.48)
Min, Max	7.77, 11.96



The distribution of the response variable is Normal so no transformations need to be applied.

Figure 2


Distribution of gend
count
gender
Distribution of race
count
race
race
1
2
3
4
Distribution of smoking
count
smoking
smoking
0
1
2
Distribution of diabet
count
diabetes
Distribution of hypertension
count
hypertension
Distribution of age
count
age
Distribution of height
count
height
Distribution of weight
count
weight
Distribution of bmi
count
bmi
Distribution of SBP
count
SBP
Distribution of LDL
count
LDL
Distribution of time
count
time


log_antibody
gender
race
smoking
diabetes
hypertension
Log Antibody Levels
age
height
weight
bmi
SBP
LDL
time

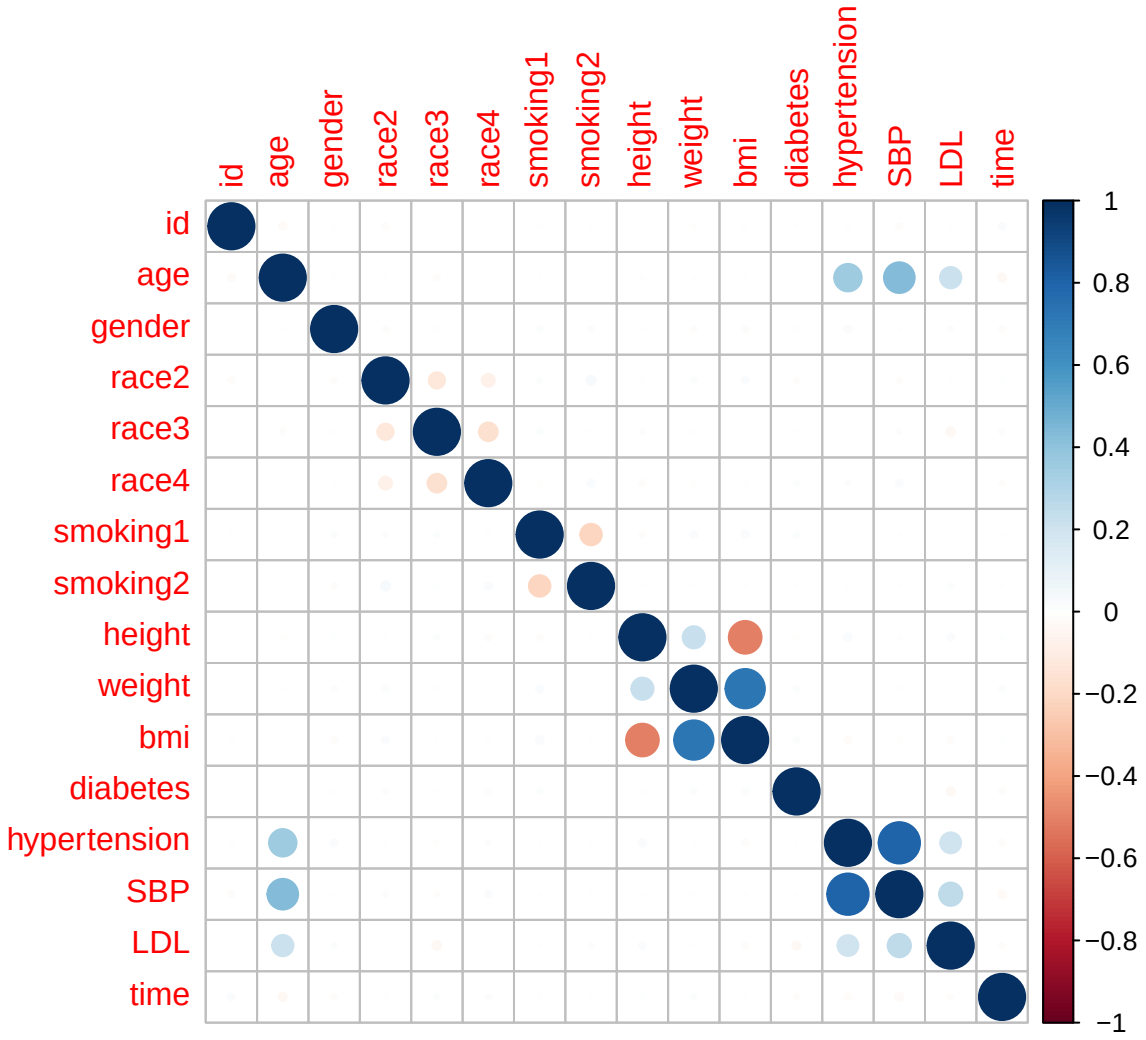
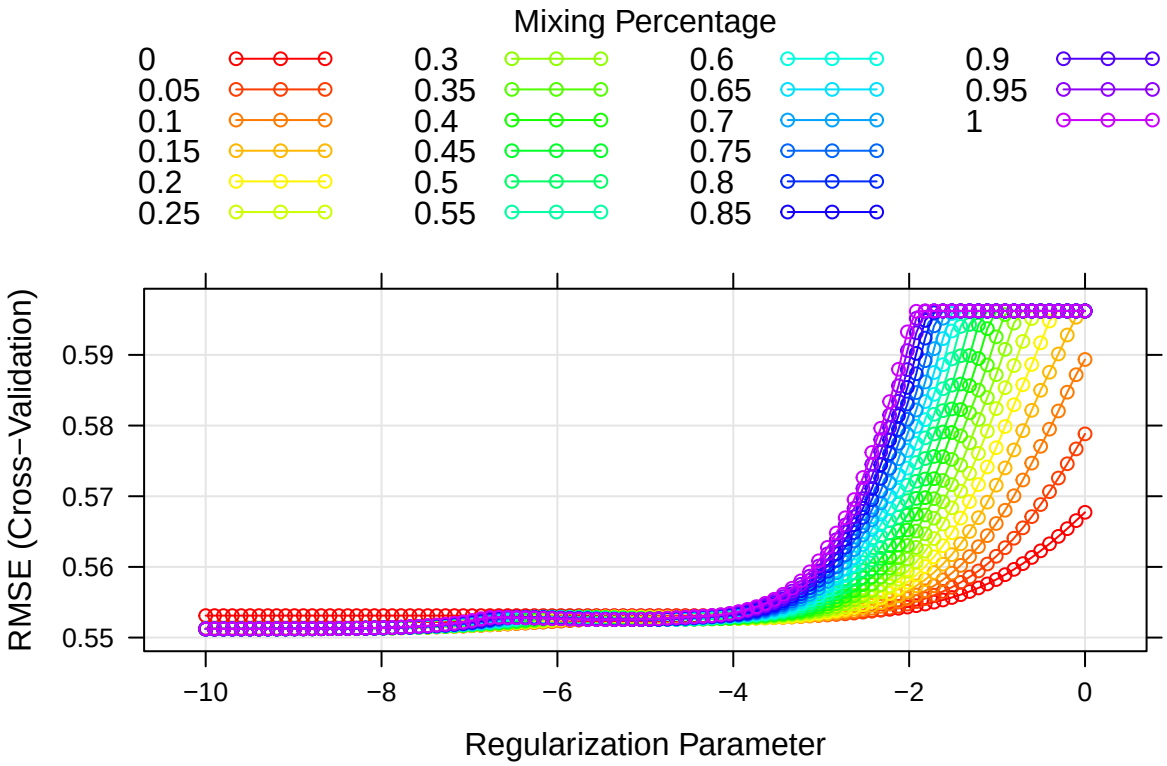

id
age
gender
race2
race3
race4
smoking1
smoking2
height
weight
bmi
diabetes
hypertension
SBP
LDL
time
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Figure 7

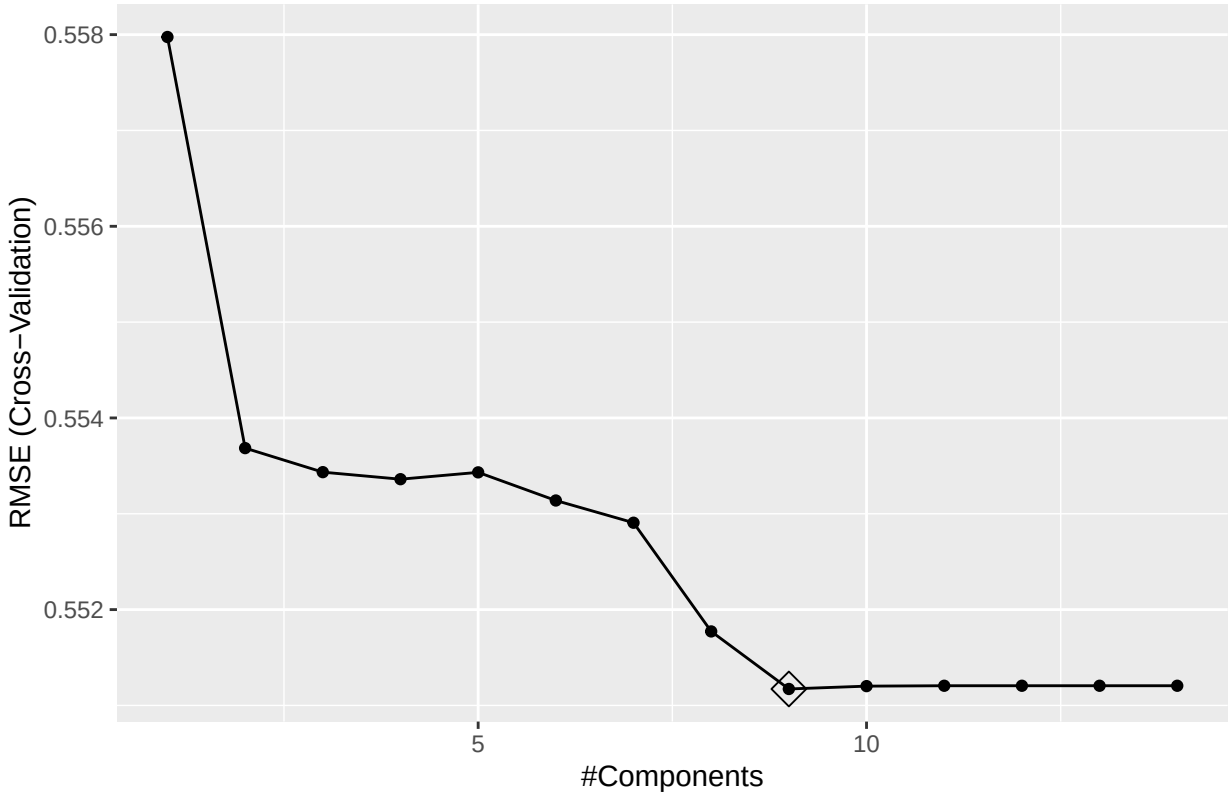


Elastic Net

PLS

```
## Data:    X dimension: 5000 15
##  Y dimension: 5000 1
## Fit method: oscorespls
## Number of components considered: 9
## TRAINING: % variance explained
##           1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X           10.77    19.88    31.77    38.37    43.94    47.54    53.19
## .outcome    12.83    14.29    14.38    14.41    14.42    14.45    14.51
##           8 comps  9 comps
## X           54.88    60.39
## .outcome    15.03    15.13
```

Figure 8: Optimal number of components



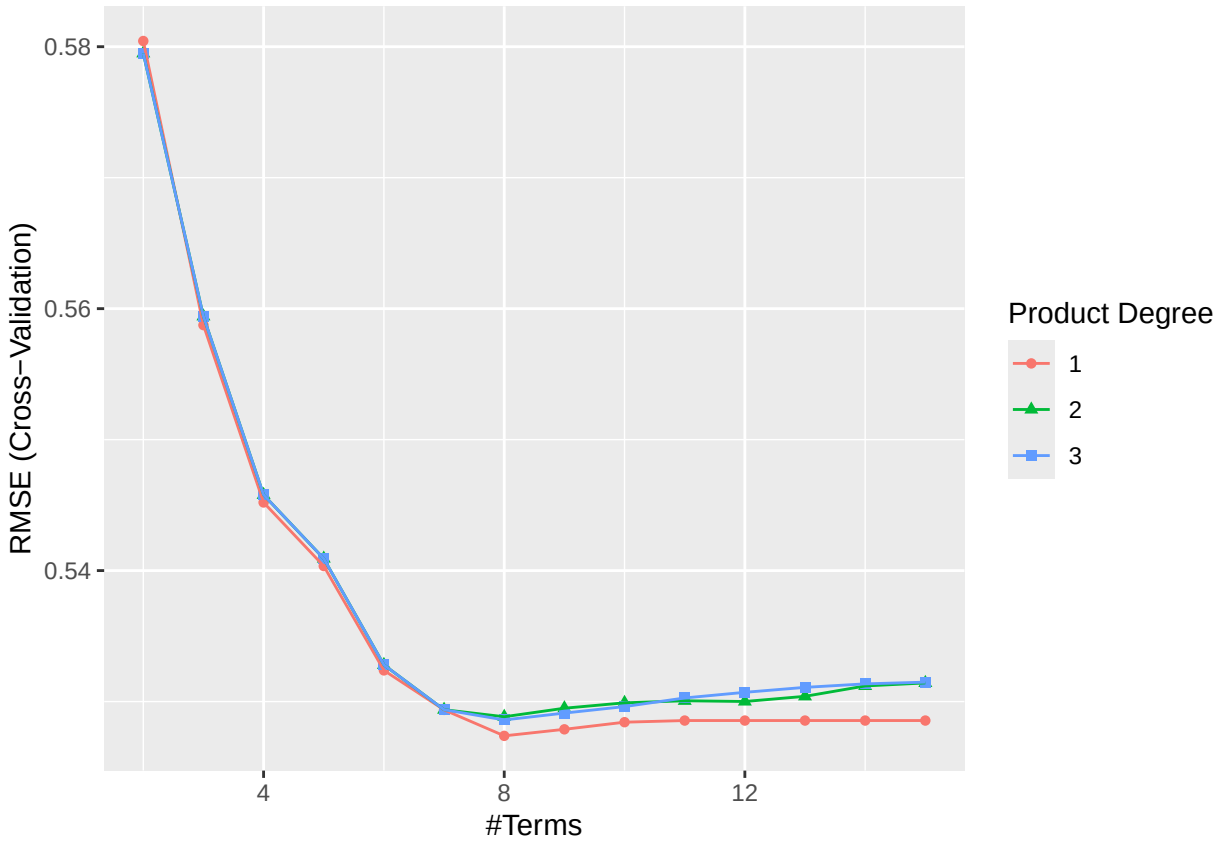

```
##   ncomp
## 9     9
```

GAM

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## .outcome ~ gender + race2 + race3 + race4 + smoking1 + smoking2 +
##     diabetes + hypertension + s(age) + s(SBP) + s(LDL) + s(bmi) +
##     s(time) + s(height) + s(weight)
##
## Parametric coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.228177   0.015328 667.269  < 2e-16 ***
## gender       -0.297837   0.014933 -19.945  < 2e-16 ***
## race2        -0.003296   0.033009  -0.100    0.920
## race3        -0.010509   0.018837  -0.558    0.577
## race4        -0.037424   0.026176  -1.430    0.153
## smoking1      0.022219   0.016660   1.334    0.182
## smoking2     -0.193175   0.025834  -7.478  8.9e-14 ***
## diabetes      0.014230   0.020640   0.689    0.491
## hypertension -0.007678   0.015995  -0.480    0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Approximate significance of smooth terms:
##                 edf Ref.df      F p-value
## s(age)    9.908e-01      9 13.733  <2e-16 ***
## s(SBP)    6.175e-07      9  0.000   0.765
## s(LDL)    6.648e-07      9  0.000   0.639
## s(bmi)    4.179e+00      9 41.897  <2e-16 ***
## s(time)   7.892e+00      9 44.960  <2e-16 ***
## s(height) 1.234e+00      9  0.278   0.121
## s(weight) 2.262e-06      9  0.000   0.666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.22   Deviance explained = 22.4%
## GCV = 0.27867  Scale est. = 0.27738   n = 5000
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## .outcome ~ gender + race2 + race3 + race4 + smoking1 + smoking2 +
##     diabetes + hypertension + s(age) + s(SBP) + s(LDL) + s(bmi) +
##     s(time) + s(height) + s(weight)
##
## Estimated degrees of freedom:
## 0.991 0.000 0.000 4.179 7.892 1.234 0.000
##  total = 23.3
##
## GCV score: 0.2786734
```

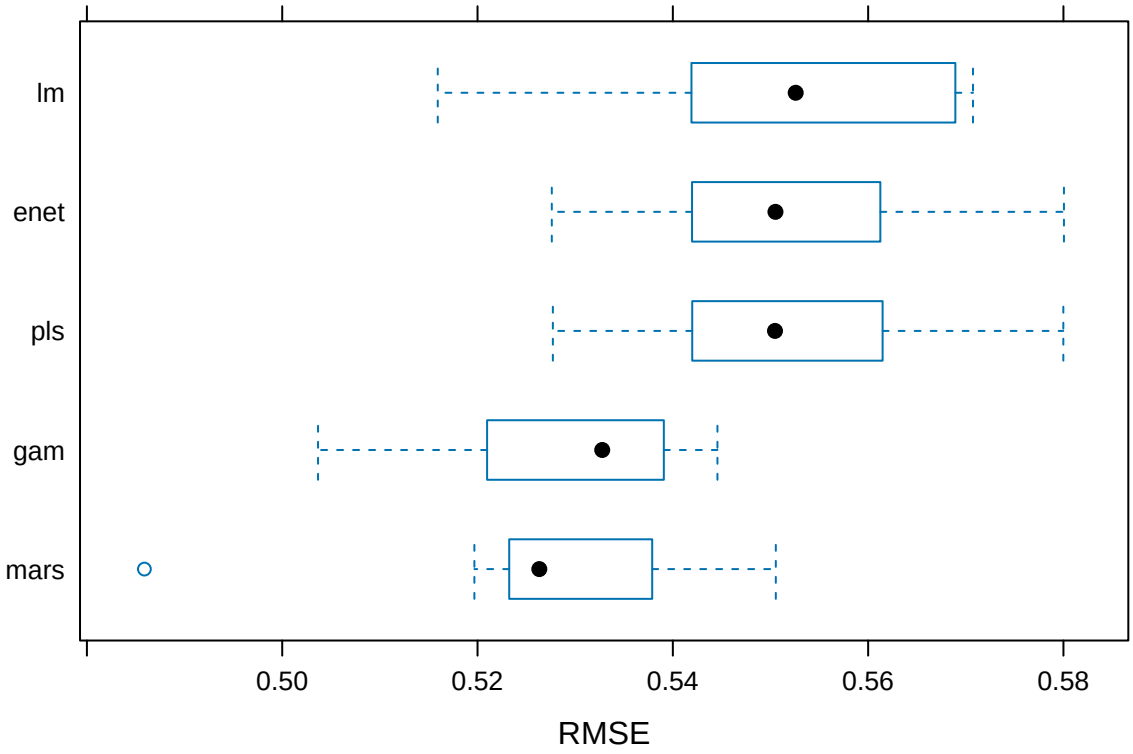


MARS

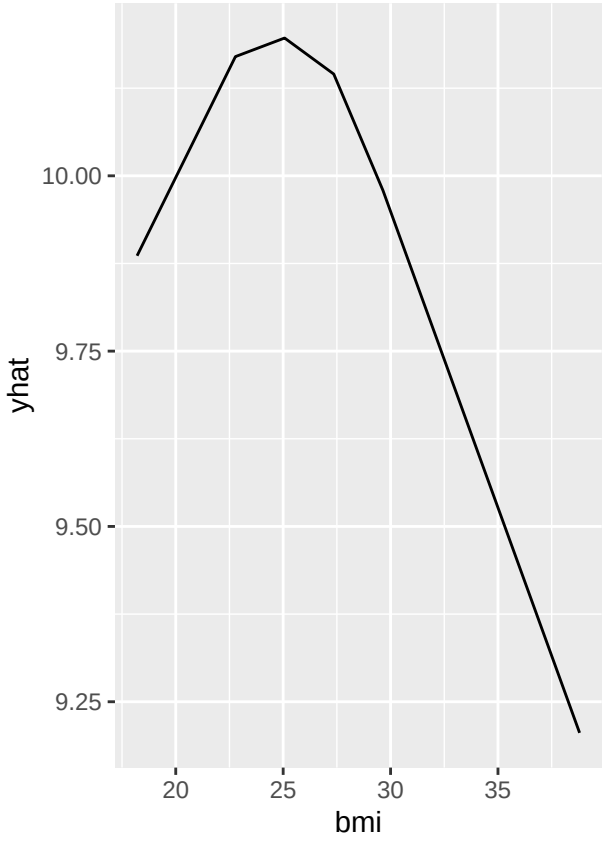
Linear regression

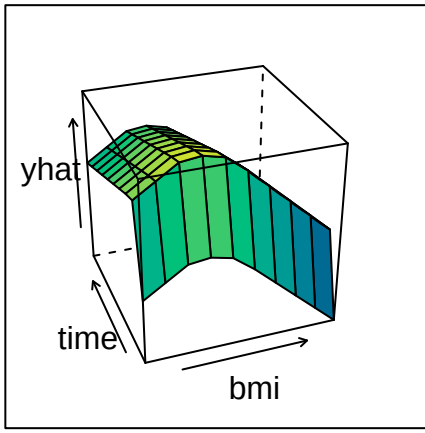
```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14396 -0.35840  0.02944  0.37802  1.65090
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.6751961  2.3149812  11.523  < 2e-16 ***
## age          -0.0205979  0.0019385 -10.626  < 2e-16 ***
## gender       -0.2974929  0.0155977 -19.073  < 2e-16 ***
## race2        -0.0060422  0.0344613  -0.175   0.8608
## race3        -0.0075295  0.0196815  -0.383   0.7021
## race4        -0.0417571  0.0273309  -1.528   0.1266
## smoking1      0.0219907  0.0173992   1.264   0.2063
## smoking2     -0.1934834  0.0269576  -7.177 8.15e-13 ***
## height       -0.0821381  0.0135622  -6.056 1.49e-09 ***
## weight        0.0859034  0.0143481   5.987 2.29e-09 ***
## bmi          -0.2977935  0.0412612  -7.217 6.10e-13 ***
## diabetes      0.0112795  0.0215643   0.523   0.6010
## hypertension -0.0179106  0.0260931  -0.686   0.4925
## SBP           0.0015181  0.0017049   0.890   0.3733
## LDL          -0.0001645  0.0004028  -0.409   0.6829
## time         -0.0003011  0.0001795  -1.677   0.0936 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5503 on 4984 degrees of freedom
## Multiple R-squared:  0.1513, Adjusted R-squared:  0.1488
## F-statistic: 59.25 on 15 and 4984 DF,  p-value: < 2.2e-16
```



CV RMSEs





Partial Dependence Plots

Test RMSE

```
## [1] 0.5349877
```