Parent and Provider Perceptions of Behavioral Healthcare in Pediatric Primary Care (PI: Andrew Riley; BDP2-262)

Benjamin Chan (chanb@ohsu.edu)

2018-06-27

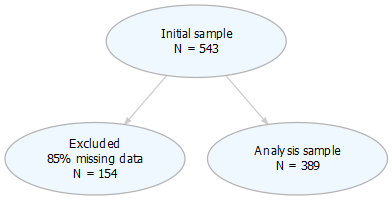
# Import Andrew’s SPSS data

Map new names to variables.

|  |  |
| --- | --- |
| oldnames | newnames |
| record\_id | id |
| eng\_span | languageSurvey |
| children\_totv\_1 | totalChildren |
| oldest\_middle\_youngest | birthOrder |
| child\_sexv\_1 | childSex |
| child\_age\_years | childAge |
| child\_ethnicity | childEthnicity |
| child\_racev\_1\_\_\_1 | childRaceWhite |
| child\_racev\_1\_\_\_2 | childRaceAsian |
| child\_racev\_1\_\_\_3 | childRaceAfrAm |
| child\_racev\_1\_\_\_4 | childRaceAIAN |
| child\_racev\_1\_\_\_5 | childRaceNHPI |
| child\_racev\_1\_\_\_6 | childRaceOther |
| child\_racev\_1\_\_\_7 | childRaceNoResp |
| related\_child | childRelationship |
| gender | parentGender |
| parent\_sexv\_1 | parentSex |
| parent\_agev\_1 | parentAge |
| parent\_ethnicity | parentEthnicity |
| parent\_race\_\_\_1 | parentRaceWhite |
| parent\_race\_\_\_2 | parentRaceAsian |
| parent\_race\_\_\_3 | parentRaceAfrAm |
| parent\_race\_\_\_4 | parentRaceAIAN |
| parent\_race\_\_\_5 | parentRaceNHPI |
| parent\_race\_\_\_6 | parentRaceOther |
| parent\_race\_\_\_7 | parentRaceNoResp |
| marital\_status | parentMaritalStatus |
| parenting\_situationv\_1 | parentSituation |
| number\_parents | parentsNumber |
| parent\_to\_child\_ratio | parentChildRatio |
| zipcode\_classification\_combined | zipcodeClass |
| zipcode | zipcode |
| community\_type | community |
| distance | distance |
| parent\_educationv\_1 | parentEducation |
| annual\_income | income |
| internet | internet |
| ECBI\_intensity\_raw\_score | ECBI\_intensity\_raw\_score |
| ECBI\_intensity\_T\_score | ECBI\_intensity\_T\_score |
| ECBI\_intensity\_clinical\_cutoff | ECBI\_intensity\_clinical\_cutoff |
| ECBI\_problem\_raw\_score | ECBI\_problem\_raw\_score |
| ECBI\_problem\_T\_score | ECBI\_problem\_T\_score |
| ECBI\_problem\_clinical\_cutoff | ECBI\_problem\_clinical\_cutoff |
| ECBI\_Opp | ECBI\_Opp |
| ECBI\_Inatt | ECBI\_Inatt |
| ECBI\_Cond | ECBI\_Cond |
| MAPS\_PP | MAPS\_PP |
| MAPS\_PR | MAPS\_PR |
| MAPS\_WM | MAPS\_WM |
| MAPS\_SP | MAPS\_SP |
| MAPS\_HS | MAPS\_HS |
| MAPS\_LC | MAPS\_LC |
| MAPS\_PC | MAPS\_PC |
| MAPS\_POS | MAPS\_POS |
| MAPS\_NEG | MAPS\_NEG |
| SEPTI\_nurturance | SEPTI\_nurturance |
| SEPTI\_n\_clinical\_cutoff | SEPTI\_n\_clinical\_cutoff |
| SEPTI\_discipline | SEPTI\_discipline |
| SEPTI\_d\_clinical\_cutoff | SEPTI\_d\_clinical\_cutoff |
| SEPTI\_play | SEPTI\_play |
| SEPTI\_p\_clinical\_cutoff | SEPTI\_p\_clinical\_cutoff |
| SEPTI\_routine | SEPTI\_routine |
| SEPTI\_r\_clinical\_cutoff | SEPTI\_r\_clinical\_cutoff |
| SEPTI\_total | SEPTI\_total |
| SEPTI\_total\_clin\_cutoff | SEPTI\_total\_clin\_cutoff |
| PCB1\_Total | PCB1\_Total |
| PCB1\_CondEmot | PCB1\_CondEmot |
| PCB1\_DevHab | PCB1\_DevHab |
| PCB2\_Tot | PCB2\_Tot |
| PCB3\_Total | PCB3\_Total |
| PBC3\_PCPonly | PCB3\_PCPonly |
| PCB3\_Person | PCB3\_Person |
| PCB3\_Resource | PCB3\_Resource |

Build analysis data set. Exclude if missing any dependent variable, PCB1\_Total, PCB2\_Tot, PCB3\_Total. Exclude rows if there are a high proportion of row-wise NA.

## PCB1\_Total PCB2\_Tot PCB3\_Total   
## Min. :18.00 Min. : 6.00 Min. :15.0   
## 1st Qu.:58.00 1st Qu.:22.00 1st Qu.:39.0   
## Median :71.00 Median :25.00 Median :48.0   
## Mean :67.89 Mean :24.54 Mean :47.6   
## 3rd Qu.:81.00 3rd Qu.:28.00 3rd Qu.:57.0   
## Max. :90.00 Max. :30.00 Max. :75.0



figures/flowChart.png

# Preprocess data

Initial preprocesssing that needs to be done that is common to PCB1\_Total, PCB2\_Tot, and PCB3\_Total.

p <- 0.75

Split data set into 75:25 training:validation samples.

inTrain <- createDataPartition(df$id, p = p)  
dfTrain <- df[inTrain$Resample1, ]  
dfValid <- df[-inTrain$Resample1, ]

Preprocess the training sample.

1. Exclude near-zero variance predictors
2. Impute missing values using k-nearest neighbor

message(sprintf("Number of complete cases before imputation = %d",  
 complete.cases(dfTrain) %>% sum()))

## Number of complete cases before imputation = 264

nzv <-   
 dfTrain %>%   
 select(-c(id,   
 PCB1\_Total, PCB1\_CondEmot, PCB1\_DevHab,   
 PCB2\_Tot,   
 PCB3\_Total, PCB3\_PCPonly, PCB3\_Person, PCB3\_Resource)) %>%   
 nearZeroVar(names = TRUE, saveMetric = TRUE) %>%  
 mutate(varname = row.names(.)) %>%   
 filter(nzv == TRUE) %>%   
 select(varname, freqRatio, percentUnique, zeroVar, nzv)   
nzv %>% kable()

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| varname | freqRatio | percentUnique | zeroVar | nzv |
| languageSurvey | 72.25000 | 0.6825939 | FALSE | TRUE |
| childRaceAfrAm | 19.85714 | 0.6825939 | FALSE | TRUE |
| childRaceAIAN | 35.50000 | 0.6825939 | FALSE | TRUE |
| childRaceNHPI | 47.66667 | 0.6825939 | FALSE | TRUE |
| childRaceOther | 21.46154 | 0.6825939 | FALSE | TRUE |
| parentRaceAfrAm | 40.71429 | 0.6825939 | FALSE | TRUE |
| parentRaceAIAN | 35.50000 | 0.6825939 | FALSE | TRUE |
| parentRaceNHPI | 57.40000 | 0.6825939 | FALSE | TRUE |
| parentRaceOther | 25.54545 | 0.6825939 | FALSE | TRUE |
| internet | 35.50000 | 0.6825939 | FALSE | TRUE |

dfTrainPreProc1 <-  
 dfTrain %>%   
 select(-one\_of(nzv$varname))  
dfOutcomes <-   
 dfTrainPreProc1 %>%   
 select(c(id,  
 PCB1\_Total, PCB1\_CondEmot, PCB1\_DevHab,   
 PCB2\_Tot,   
 PCB3\_Total, PCB3\_PCPonly, PCB3\_Person, PCB3\_Resource))  
dfTrainPreProc2 <-   
 dfTrainPreProc1 %>%   
 select(-c(id,  
 PCB1\_Total, PCB1\_CondEmot, PCB1\_DevHab,   
 PCB2\_Tot,   
 PCB3\_Total, PCB3\_PCPonly, PCB3\_Person, PCB3\_Resource))  
preProc <-  
 dfTrainPreProc2 %>%   
 preProcess(method = c("nzv", "corr", "knnImpute"), verbose = TRUE)

## 2 highly correlated predictors were removed.  
## Calculating 32 means for centering  
## Calculating 32 standard deviations for scaling

preProc

## Created from 264 samples and 54 variables  
##   
## Pre-processing:  
## - centered (32)  
## - ignored (20)  
## - 5 nearest neighbor imputation (32)  
## - removed (2)  
## - scaled (32)

dfTrainPreProc3 <- predict(preProc, dfTrainPreProc2)  
dfTrainPreProc <- bind\_cols(dfOutcomes, dfTrainPreProc3)  
message(sprintf("Number of complete cases after imputation = %d",  
 complete.cases(dfTrainPreProc) %>% sum()))

## Number of complete cases after imputation = 291

save(dfTrainPreProc, dfValid, dfTrain, nzv, preProc, df, file = "data/processed/dataframes.RData")  
rm(dfTrainPreProc1, dfTrainPreProc2, dfTrainPreProc3)

Set the control parameters.

ctrl <- trainControl(method = "repeatedcv",  
 number = 10,  
 repeats = 10,  
 savePredictions = TRUE,  
 allowParallel = TRUE,  
 search = "random")  
cores <- 24

Set the model and tuning parameter grid.

library(earth)

## Loading required package: plotmo

## Loading required package: plotrix

## Loading required package: TeachingDemos

citation("earth")

##   
## To cite package 'earth' in publications use:  
##   
## Stephen Milborrow. Derived from mda:mars by Trevor Hastie and  
## Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas  
## Lumley's leaps wrapper. (2018). earth: Multivariate Adaptive  
## Regression Splines. R package version 4.6.2.  
## https://CRAN.R-project.org/package=earth  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {earth: Multivariate Adaptive Regression Splines},  
## author = {Stephen Milborrow. Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper.},  
## year = {2018},  
## note = {R package version 4.6.2},  
## url = {https://CRAN.R-project.org/package=earth},  
## }  
##   
## ATTENTION: This citation information has been auto-generated from  
## the package DESCRIPTION file and may need manual editing, see  
## 'help("citation")'.

method <- "bagEarth"  
modelLookup(method) %>% kable()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| model | parameter | label | forReg | forClass | probModel |
| bagEarth | nprune | #Terms | TRUE | TRUE | TRUE |
| bagEarth | degree | Product Degree | TRUE | TRUE | TRUE |

grid <- expand.grid(nprune = c(seq(2, 9, 1), seq(10, 30, 5)),  
 degree = seq(1, 3))  
grid %>% kable()

|  |  |
| --- | --- |
| nprune | degree |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |
| 15 | 1 |
| 20 | 1 |
| 25 | 1 |
| 30 | 1 |
| 2 | 2 |
| 3 | 2 |
| 4 | 2 |
| 5 | 2 |
| 6 | 2 |
| 7 | 2 |
| 8 | 2 |
| 9 | 2 |
| 10 | 2 |
| 15 | 2 |
| 20 | 2 |
| 25 | 2 |
| 30 | 2 |
| 2 | 3 |
| 3 | 3 |
| 4 | 3 |
| 5 | 3 |
| 6 | 3 |
| 7 | 3 |
| 8 | 3 |
| 9 | 3 |
| 10 | 3 |
| 15 | 3 |
| 20 | 3 |
| 25 | 3 |
| 30 | 3 |

# Model PCB1

## PCB1 Total

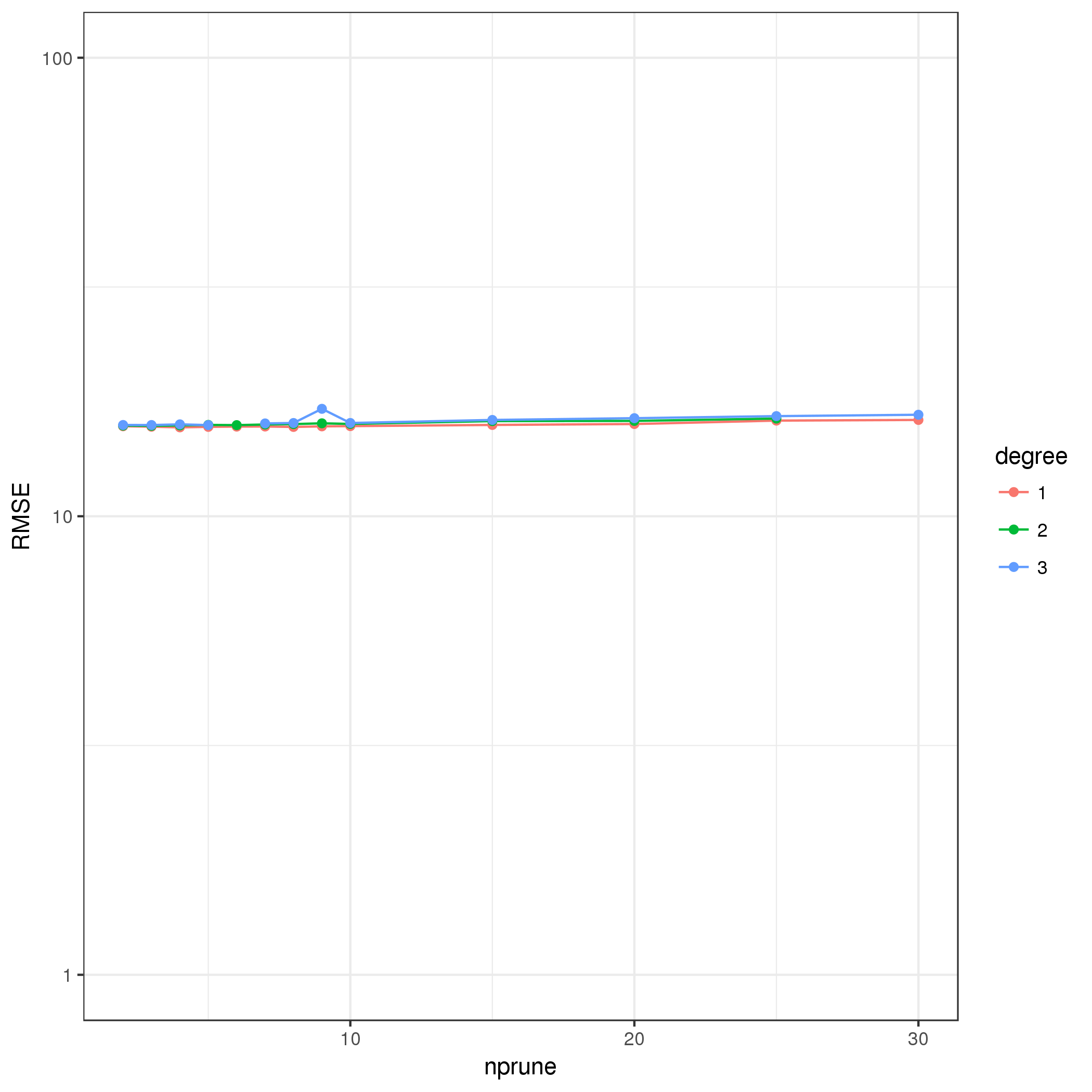
Prediction model for PCB1.

Train model over the tuning parameters.

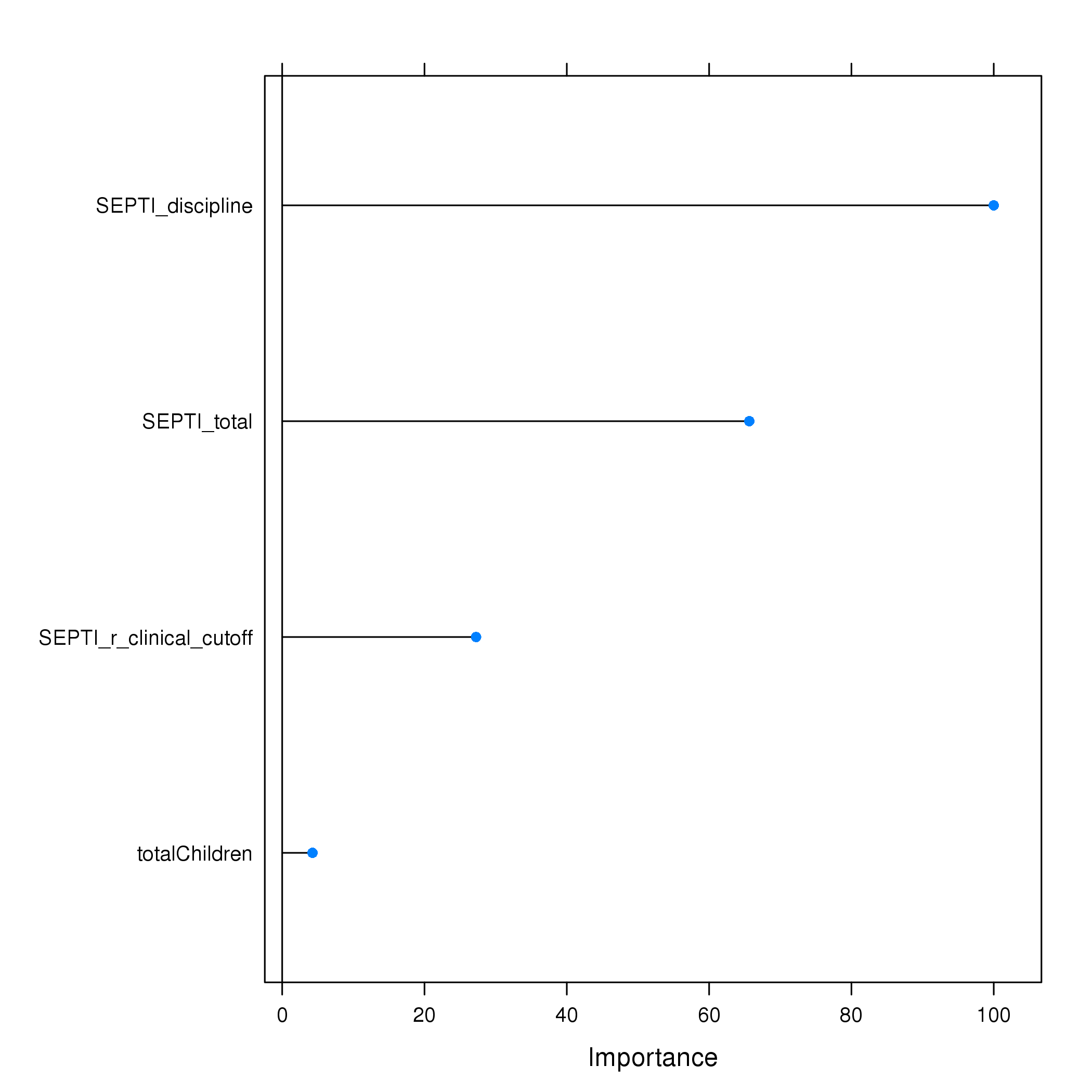
## Bagged MARS   
##   
## 293 samples  
## 52 predictor  
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold, repeated 10 times)   
## Summary of sample sizes: 261, 263, 263, 261, 261, 262, ...   
## Resampling results across tuning parameters:  
##   
## degree nprune RMSE Rsquared MAE   
## 1 2 1.571482e+01 0.04320407 1.284548e+01  
## 1 3 1.567991e+01 0.05145719 1.278524e+01  
## 1 4 1.562755e+01 0.05651002 1.274021e+01  
## 1 5 1.566533e+01 0.05541351 1.277778e+01  
## 1 6 1.568706e+01 0.05517188 1.277360e+01  
## 1 7 1.569454e+01 0.05885924 1.277500e+01  
## 1 8 1.566087e+01 0.06245054 1.274509e+01  
## 1 9 1.570695e+01 0.05630420 1.276691e+01  
## 1 10 1.572521e+01 0.05725950 1.274594e+01  
## 1 15 1.581852e+01 0.06600275 1.275550e+01  
## 1 20 1.589211e+01 0.06726229 1.272955e+01  
## 1 25 1.616667e+01 0.06474451 1.290041e+01  
## 1 30 1.622059e+01 0.06578394 1.292120e+01  
## 2 2 1.576334e+01 0.04481756 1.292294e+01  
## 2 3 1.574651e+01 0.04180724 1.288224e+01  
## 2 4 1.578125e+01 0.04323332 1.291120e+01  
## 2 5 1.582187e+01 0.04828768 1.292781e+01  
## 2 6 1.580352e+01 0.04944371 1.290291e+01  
## 2 7 1.585343e+01 0.04745533 1.293071e+01  
## 2 8 1.588597e+01 0.04639400 1.293602e+01  
## 2 9 1.594741e+01 0.04281639 1.297088e+01  
## 2 10 1.589872e+01 0.05401455 1.291907e+01  
## 2 15 1.612759e+01 0.04880661 1.298727e+01  
## 2 20 1.614137e+01 0.05576970 1.295629e+01  
## 2 25 1.633253e+01 0.05678335 1.309085e+01  
## 2 30 2.746094e+13 0.05023996 5.013660e+12  
## 3 2 1.581464e+01 0.03933084 1.296712e+01  
## 3 3 1.581118e+01 0.03953727 1.292969e+01  
## 3 4 1.586520e+01 0.03681780 1.298052e+01  
## 3 5 1.580509e+01 0.04897981 1.292875e+01  
## 3 6 1.771734e+12 0.04255419 3.234729e+11  
## 3 7 1.593850e+01 0.04323598 1.303057e+01  
## 3 8 1.597212e+01 0.05078073 1.299998e+01  
## 3 9 1.715234e+01 0.04101746 1.340192e+01  
## 3 10 1.597406e+01 0.04571486 1.294588e+01  
## 3 15 1.621960e+01 0.04910649 1.307904e+01  
## 3 20 1.636452e+01 0.05537843 1.312913e+01  
## 3 25 1.653522e+01 0.05590162 1.320590e+01  
## 3 30 1.664848e+01 0.05316593 1.330270e+01  
##   
## RMSE was used to select the optimal model using the smallest value.  
## The final values used for the model were nprune = 4 and degree = 1.

## Warning: Removed 1 rows containing missing values (geom\_path).

## Warning: Removed 2 rows containing missing values (geom\_point).



plot of chunk PCB1\_Total\_Training

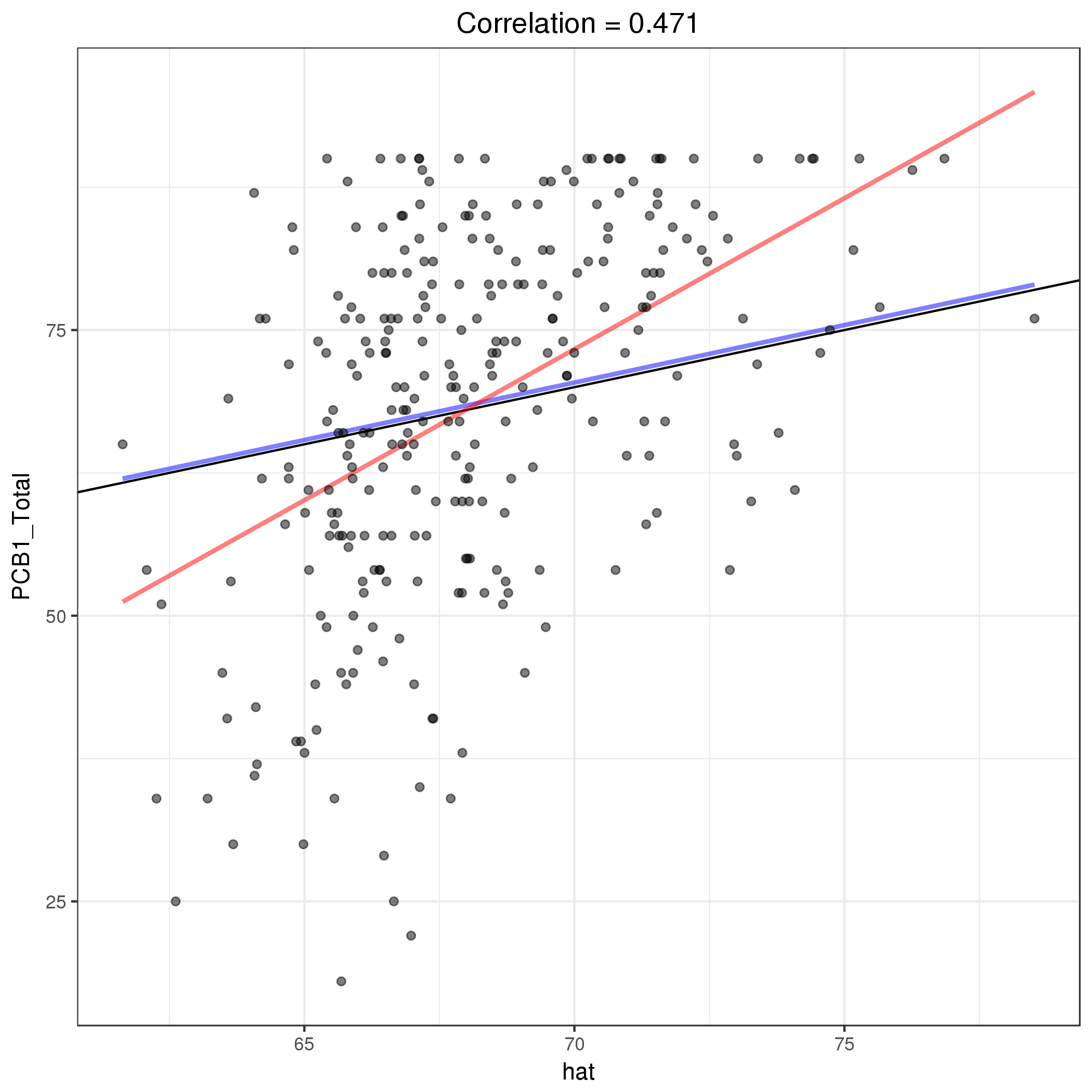


plot of chunk PCB1\_Total\_Training-varImp

|  |  |
| --- | --- |
| variable | Overall |
| SEPTI\_discipline | 100.000000 |
| SEPTI\_total | 65.646693 |
| SEPTI\_r\_clinical\_cutoff | 27.254022 |
| totalChildren | 4.260233 |

## RMSE Rsquared MAE   
## 14.6958975 0.2222885 11.9854896

## PCB1\_Total hat  
## PCB1\_Total 1.0000000 0.4714748  
## hat 0.4714748 1.0000000

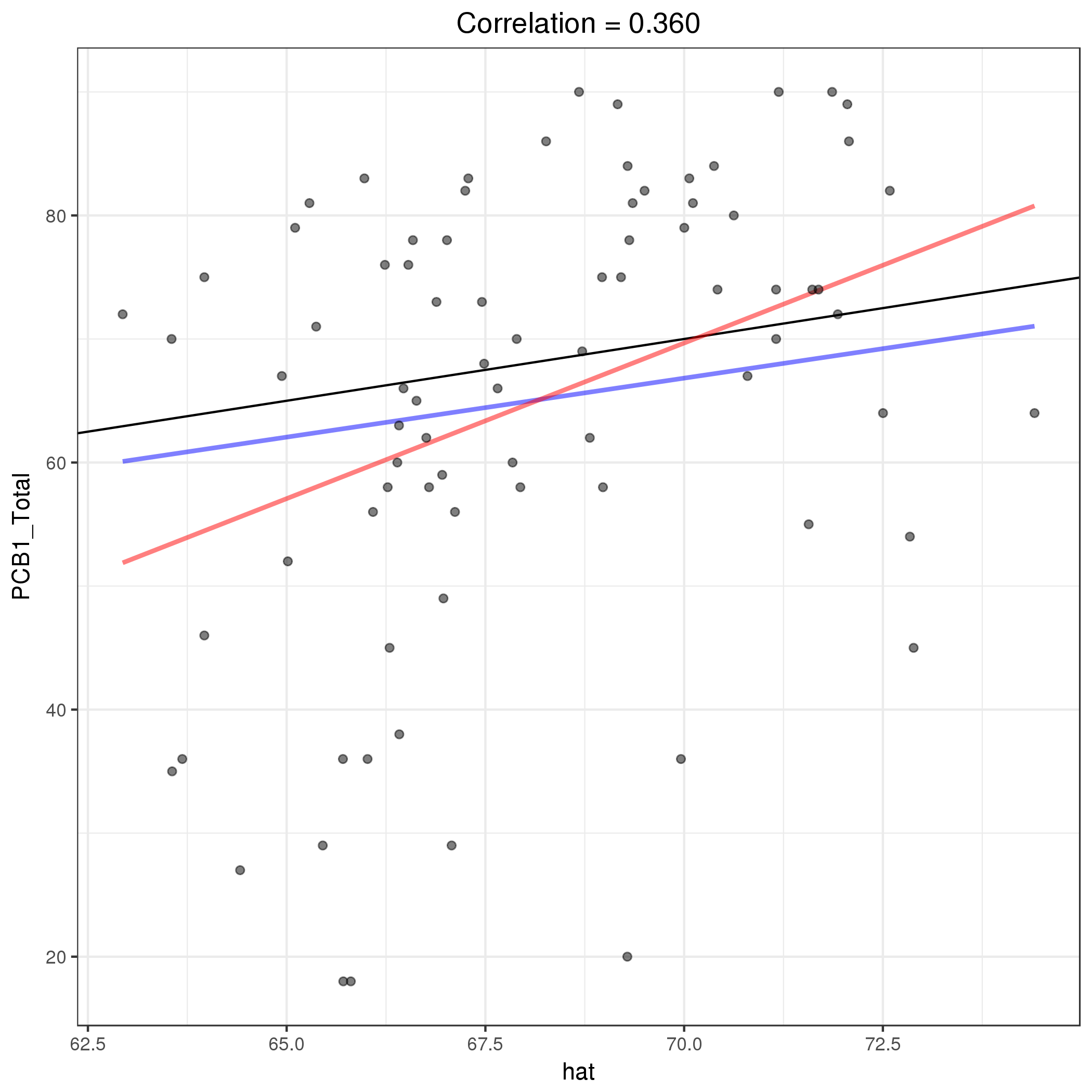


plot of chunk PCB1\_Total\_Training-predict

Evaluate model on the validation sample.

## RMSE Rsquared MAE   
## 17.8941724 0.1298564 13.8599137

## PCB1\_Total hat  
## PCB1\_Total 1.0000000 0.3603559  
## hat 0.3603559 1.0000000



plot of chunk PCB1\_Total\_Validation-predict

# Model PCB2

## Total PCB2

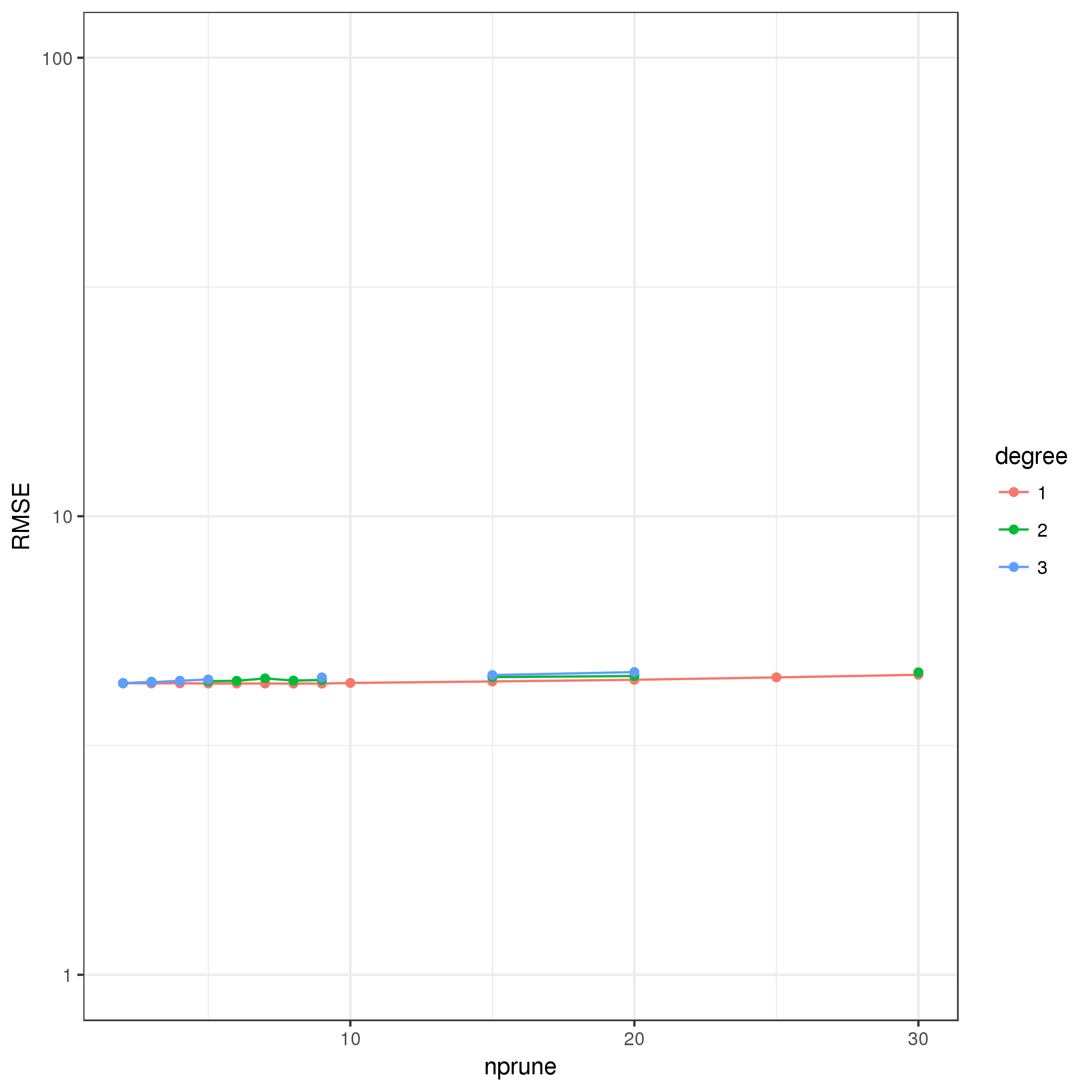
Prediction model for PCB2\_Total.

Train model over the tuning parameters.

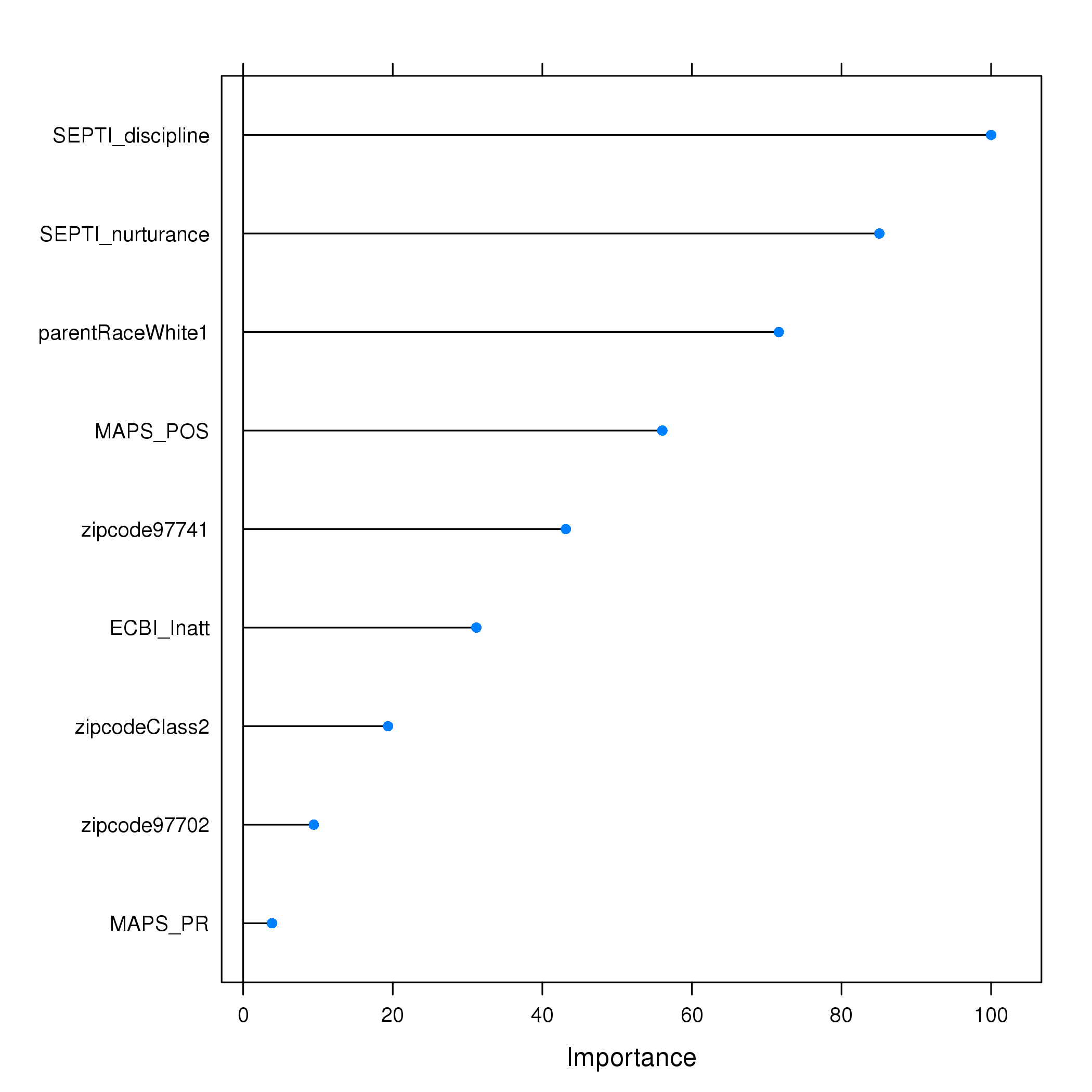
## Bagged MARS   
##   
## 293 samples  
## 52 predictor  
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold, repeated 10 times)   
## Summary of sample sizes: 263, 262, 262, 262, 262, 262, ...   
## Resampling results across tuning parameters:  
##   
## degree nprune RMSE Rsquared MAE   
## 1 2 4.326906e+00 0.03340834 3.298945e+00  
## 1 3 4.319109e+00 0.04858244 3.290790e+00  
## 1 4 4.322637e+00 0.05248578 3.284097e+00  
## 1 5 4.317835e+00 0.05373481 3.287895e+00  
## 1 6 4.317654e+00 0.05590818 3.282685e+00  
## 1 7 4.317478e+00 0.05946571 3.277796e+00  
## 1 8 4.316295e+00 0.06109812 3.275342e+00  
## 1 9 4.315981e+00 0.06059385 3.282386e+00  
## 1 10 4.329833e+00 0.05518009 3.279910e+00  
## 1 15 4.362482e+00 0.06387532 3.313125e+00  
## 1 20 4.399549e+00 0.06267637 3.335373e+00  
## 1 25 4.452408e+00 0.06088870 3.379584e+00  
## 1 30 4.511364e+00 0.06107648 3.430025e+00  
## 2 2 4.320572e+00 0.04109506 3.301124e+00  
## 2 3 4.354108e+00 0.04046534 3.328897e+00  
## 2 4 9.201627e+11 0.04217304 2.128289e+11  
## 2 5 4.369004e+00 0.04602975 3.342857e+00  
## 2 6 4.376198e+00 0.04619965 3.347672e+00  
## 2 7 4.429904e+00 0.03776036 3.368047e+00  
## 2 8 4.381312e+00 0.05053456 3.350606e+00  
## 2 9 4.395099e+00 0.04939043 3.359015e+00  
## 2 10 3.025465e+11 0.04295402 5.822511e+10  
## 2 15 4.458666e+00 0.04639679 3.412187e+00  
## 2 20 4.483655e+00 0.05107595 3.435342e+00  
## 2 25 4.032199e+10 0.04290887 7.487606e+09  
## 2 30 4.565993e+00 0.04949845 3.498730e+00  
## 3 2 4.328941e+00 0.04847328 3.301778e+00  
## 3 3 4.344888e+00 0.04161013 3.328367e+00  
## 3 4 4.378336e+00 0.04015509 3.349420e+00  
## 3 5 4.405762e+00 0.04231529 3.367112e+00  
## 3 6 1.119803e+13 0.05011717 2.116229e+12  
## 3 7 8.174627e+11 0.03181318 1.492476e+11  
## 3 8 5.967251e+11 0.03735934 1.545755e+11  
## 3 9 4.451882e+00 0.04306249 3.410233e+00  
## 3 10 3.912912e+11 0.04308901 7.266094e+10  
## 3 15 4.505070e+00 0.04236338 3.461924e+00  
## 3 20 4.571760e+00 0.05546553 3.487775e+00  
## 3 25 9.023605e+11 0.04262326 1.675641e+11  
## 3 30 3.343157e+11 0.04375086 6.203404e+10  
##   
## RMSE was used to select the optimal model using the smallest value.  
## The final values used for the model were nprune = 9 and degree = 1.

## Warning: Removed 2 rows containing missing values (geom\_path).

## Warning: Removed 9 rows containing missing values (geom\_point).



plot of chunk PCB2\_Tot\_Training

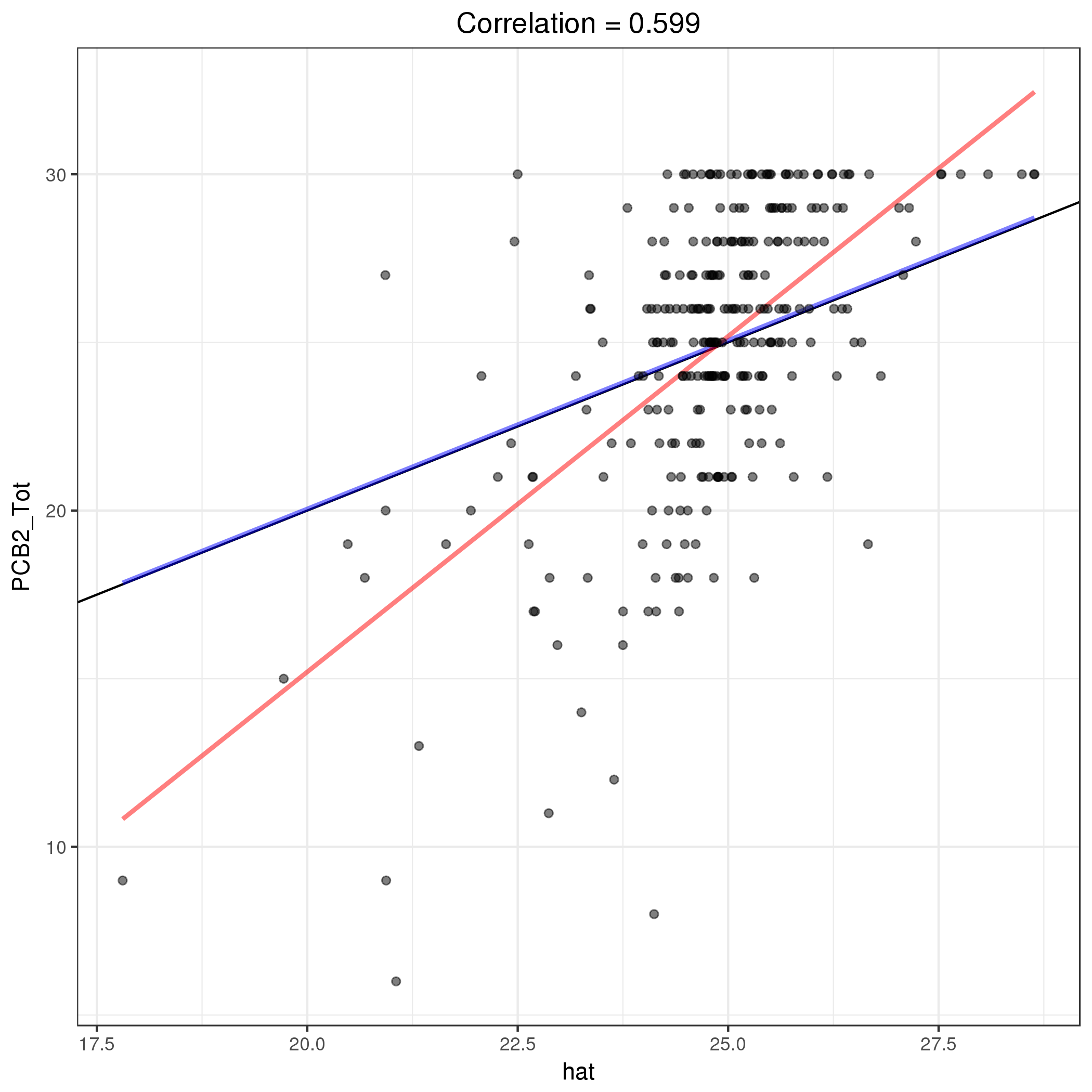


plot of chunk PCB2\_Total\_Training-varImp

|  |  |
| --- | --- |
| variable | Overall |
| SEPTI\_discipline | 100.000000 |
| SEPTI\_nurturance | 85.061688 |
| parentRaceWhite1 | 71.614498 |
| MAPS\_POS | 56.040444 |
| zipcode97741 | 43.142712 |
| ECBI\_Inatt | 31.176062 |
| zipcodeClass2 | 19.364705 |
| zipcode97702 | 9.444855 |
| MAPS\_PR | 3.852440 |

## RMSE Rsquared MAE   
## 3.7239749 0.3586612 2.8582647

## PCB2\_Tot hat  
## PCB2\_Tot 1.0000000 0.5988833  
## hat 0.5988833 1.0000000

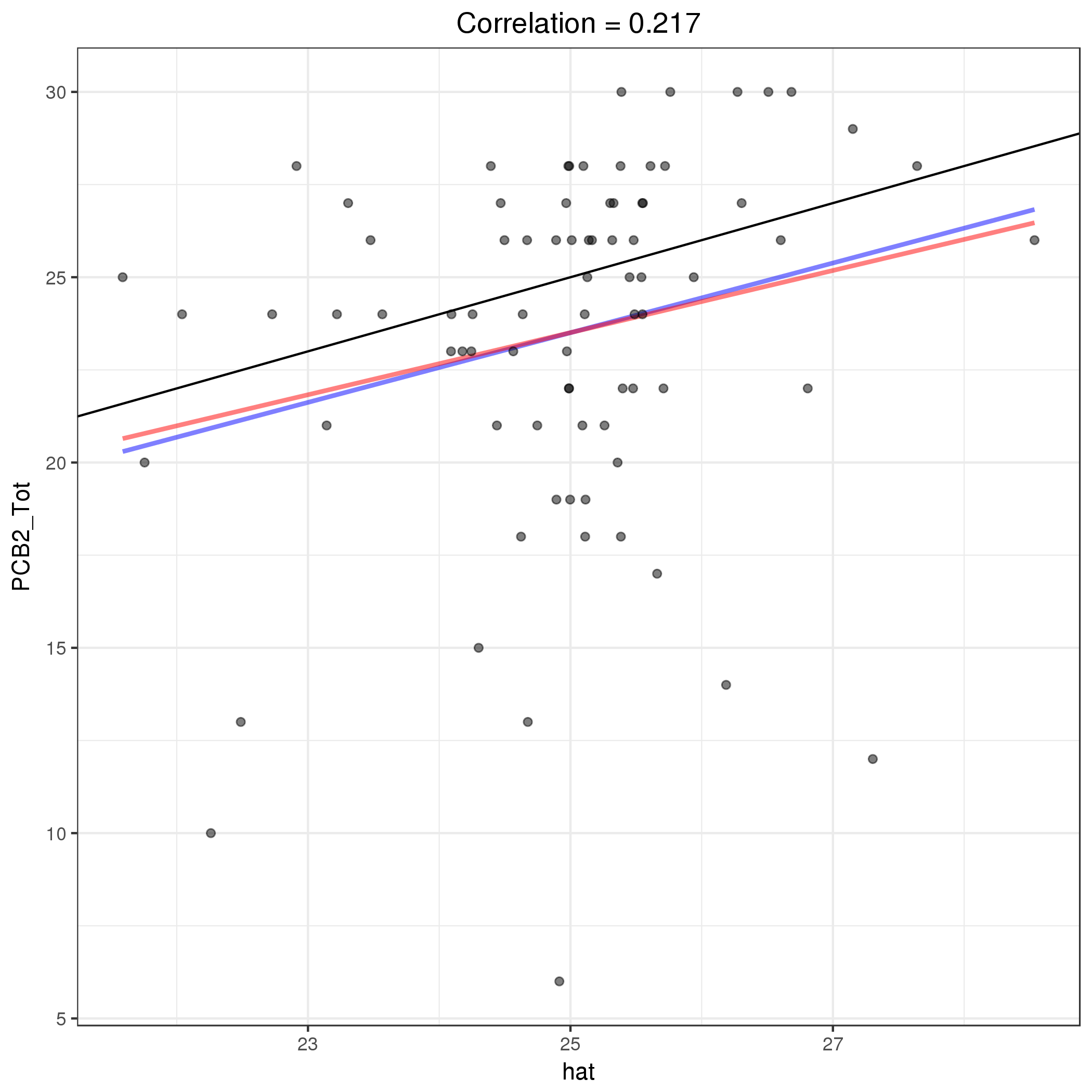


plot of chunk PCB2\_Tot\_Training-predict

Evaluate model on the validation sample.

## RMSE Rsquared MAE   
## 4.92728210 0.04714749 3.46159461

## PCB2\_Tot hat  
## PCB2\_Tot 1.0000000 0.2171347  
## hat 0.2171347 1.0000000



plot of chunk PCB2\_Tot\_Validation-predict

# Model PCB3

## Total PCB3

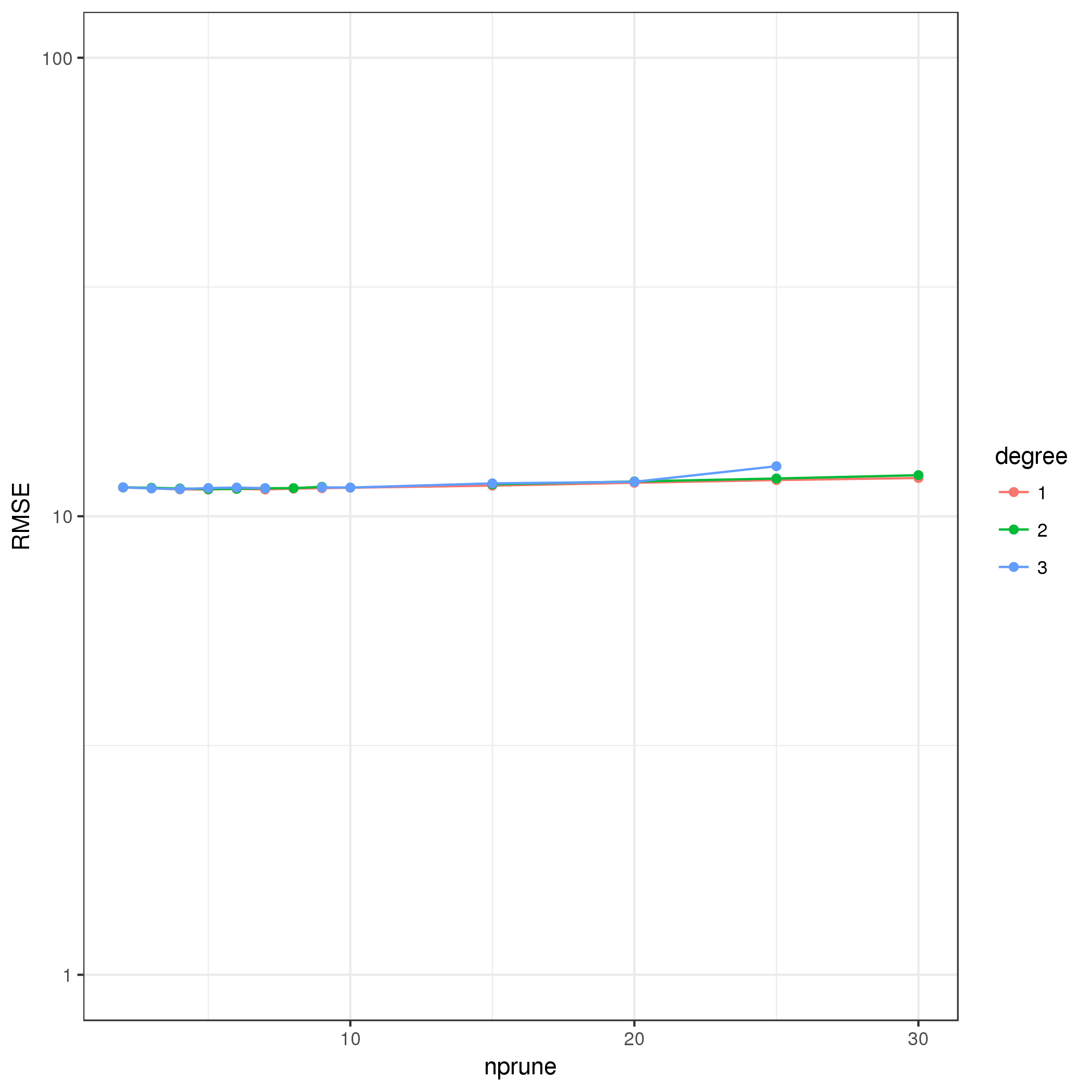
Prediction model for PCB3\_Total.

Train model over the tuning parameters.

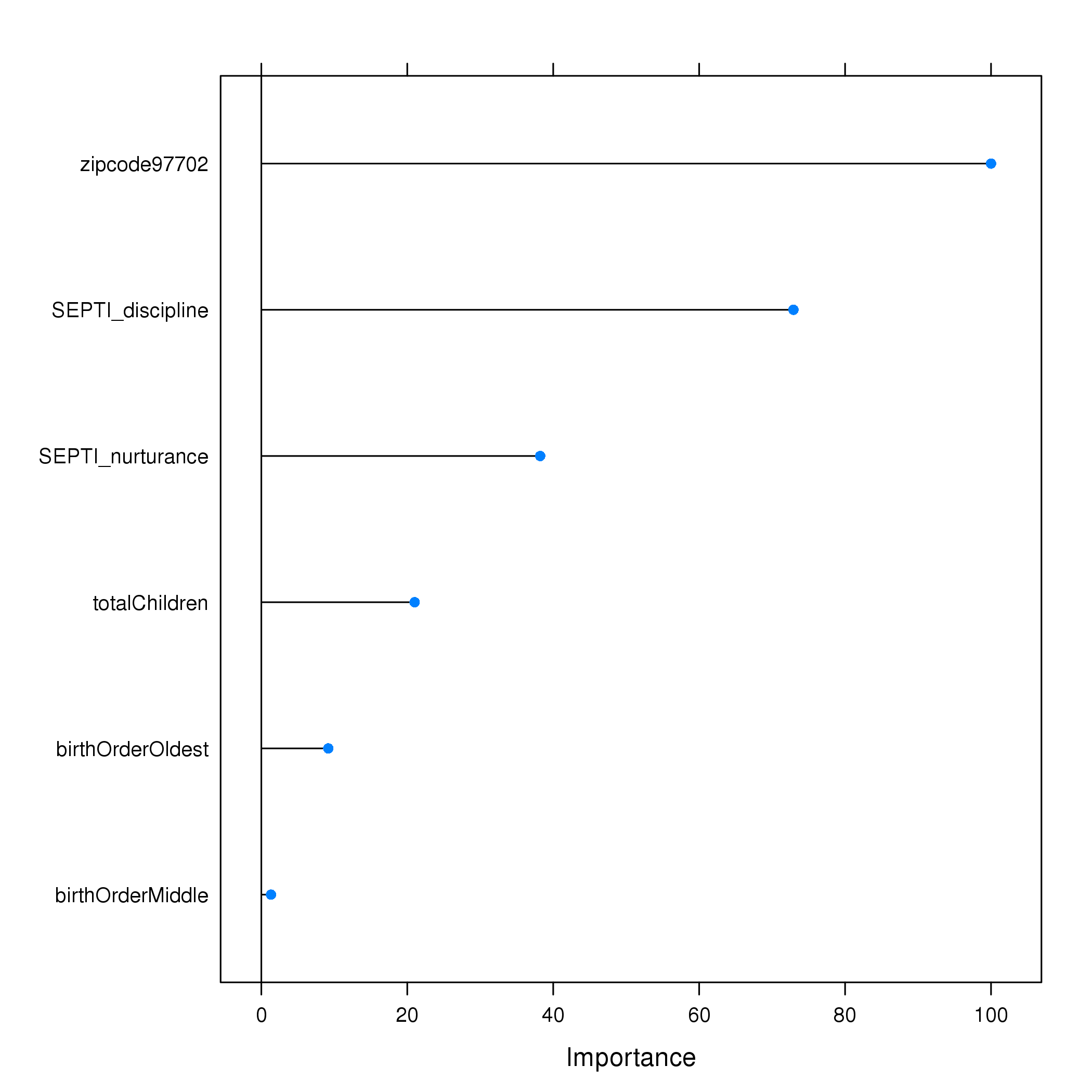
## Bagged MARS   
##   
## 293 samples  
## 52 predictor  
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold, repeated 10 times)   
## Summary of sample sizes: 264, 263, 262, 260, 261, 262, ...   
## Resampling results across tuning parameters:  
##   
## degree nprune RMSE Rsquared MAE   
## 1 2 1.155186e+01 0.09121569 9.588241e+00  
## 1 3 1.151347e+01 0.09007050 9.547837e+00  
## 1 4 1.144426e+01 0.09854393 9.457875e+00  
## 1 5 1.144068e+01 0.09668763 9.440352e+00  
## 1 6 1.147772e+01 0.09406201 9.459848e+00  
## 1 7 1.144264e+01 0.09994678 9.417214e+00  
## 1 8 1.148913e+01 0.09800851 9.442785e+00  
## 1 9 1.151576e+01 0.09819661 9.444117e+00  
## 1 10 1.154247e+01 0.09462947 9.464077e+00  
## 1 15 1.166429e+01 0.09344122 9.509728e+00  
## 1 20 1.183185e+01 0.08898440 9.614596e+00  
## 1 25 1.199443e+01 0.08601278 9.723684e+00  
## 1 30 1.211864e+01 0.08069752 9.781465e+00  
## 2 2 1.156130e+01 0.09276383 9.594094e+00  
## 2 3 1.153483e+01 0.09197486 9.543339e+00  
## 2 4 1.149728e+01 0.08670606 9.511574e+00  
## 2 5 1.145706e+01 0.09452997 9.457835e+00  
## 2 6 1.147918e+01 0.09197558 9.465717e+00  
## 2 7 1.151060e+01 0.09177462 9.466222e+00  
## 2 8 1.151764e+01 0.08771254 9.442970e+00  
## 2 9 1.159369e+01 0.08525993 9.508488e+00  
## 2 10 4.619929e+13 0.08733468 1.205624e+13  
## 2 15 1.173873e+01 0.08515665 9.557688e+00  
## 2 20 1.190960e+01 0.07983226 9.652294e+00  
## 2 25 1.208812e+01 0.06810859 9.780269e+00  
## 2 30 1.229209e+01 0.06640448 9.884609e+00  
## 3 2 1.156140e+01 0.09291263 9.606094e+00  
## 3 3 1.149704e+01 0.09230624 9.552309e+00  
## 3 4 1.146244e+01 0.09168361 9.487952e+00  
## 3 5 1.152220e+01 0.08458671 9.512262e+00  
## 3 6 1.155166e+01 0.08828323 9.509047e+00  
## 3 7 1.151648e+01 0.09632282 9.486651e+00  
## 3 8 1.157512e+13 0.09071222 2.149446e+12  
## 3 9 1.156105e+01 0.08395228 9.489769e+00  
## 3 10 1.155130e+01 0.08751224 9.469815e+00  
## 3 15 1.179592e+01 0.08367928 9.590294e+00  
## 3 20 1.188937e+01 0.08213511 9.631434e+00  
## 3 25 1.285698e+01 0.07132946 9.974221e+00  
## 3 30 5.918788e+12 0.06450496 1.121730e+12  
##   
## RMSE was used to select the optimal model using the smallest value.  
## The final values used for the model were nprune = 5 and degree = 1.

## Warning: Removed 1 rows containing missing values (geom\_path).

## Warning: Removed 3 rows containing missing values (geom\_point).



plot of chunk PCB3\_Total\_Training

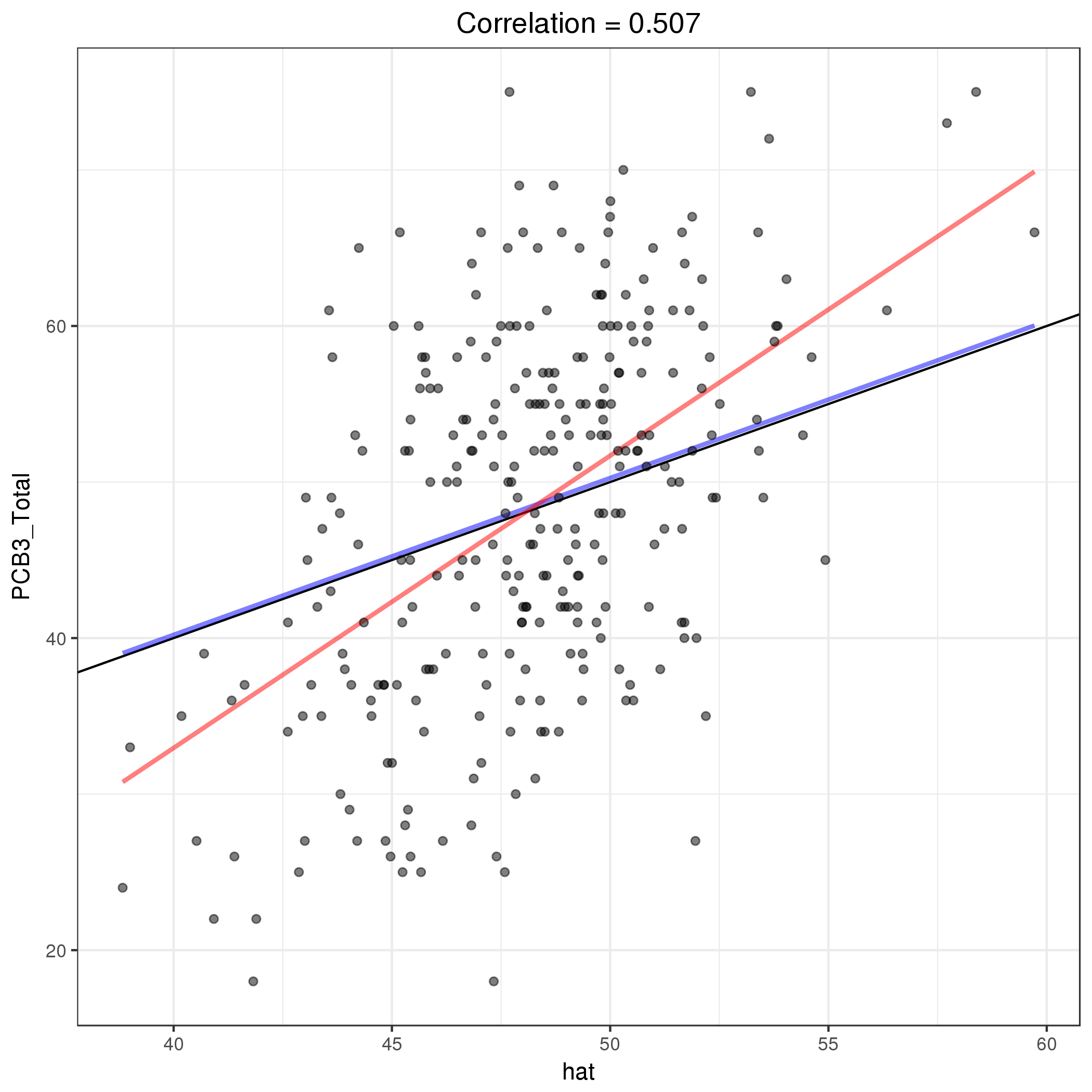


plot of chunk PCB3\_Total\_Training-varImp

|  |  |
| --- | --- |
| variable | Overall |
| zipcode97702 | 100.000000 |
| SEPTI\_discipline | 72.920401 |
| SEPTI\_nurturance | 38.219334 |
| totalChildren | 21.012094 |
| birthOrderOldest | 9.172569 |
| birthOrderMiddle | 1.325648 |

## RMSE Rsquared MAE   
## 10.574054 0.257276 8.803982

## PCB3\_Total hat  
## PCB3\_Total 1.0000000 0.5072239  
## hat 0.5072239 1.0000000

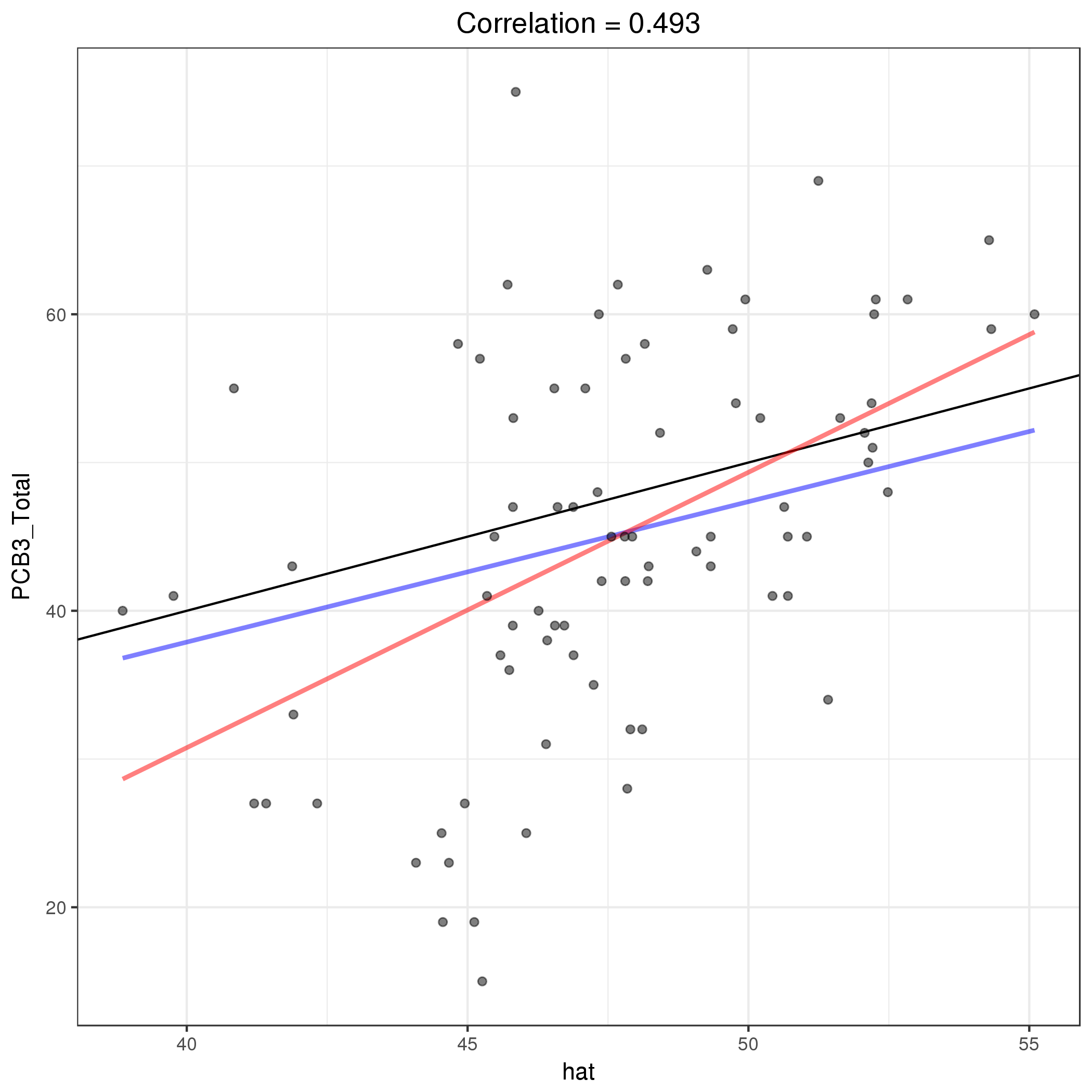


plot of chunk PCB3\_Total\_Training-predict

Evaluate model on the validation sample.

## RMSE Rsquared MAE   
## 11.7605361 0.2433116 9.4118031

## PCB3\_Total hat  
## PCB3\_Total 1.0000000 0.4932663  
## hat 0.4932663 1.0000000



plot of chunk PCB3\_TotalValidation-predict