Examining Predictors of HDL Cholestrol using NHANES Data

Edward J. Lee, Yusuf Emre, Vincent Mazz

2025-04-05

Introduction

Cardiovascular disease remains a leading cause of death worldwide, with elevated cholesterol levels serving as an important and preventable risk factor. As prevention becomes a cornerstone of public health policy, understanding what drives changes in cholesterol is of the most importance.

This study investigates which factors significantly influence total cholesterol levels in the adult population, using data from the National Health and Nutrition Examination Survey (NHANES). Specifically, it examines the role of age, weight, height, blood pressure, smoking habits, and physical activity as potential predictors.

Previous research provides a useful foundation. For example, Ferrara et al. (1997) found that cholesterol levels tend to decline in older adults. However, this study observed a weak but significant positive association between age and cholesterol, suggesting that additional lifestyle or metabolic factors may be at play. Next, Henriksson et al. (2001) reported a negative correlation between BMI and HDL cholesterol. Although BMI was not included directly in this model, height and weight were analyzed independently. The findings showed that weight alone lacked a strong relationship with cholesterol, partially contradicting earlier work. Finally, Kim et al. (2011) linked high blood pressure with poorer lipid profiles—a pattern repeated here, as both systolic and diastolic blood pressure were positively associated with cholesterol levels.

However, there remains inconsistency in how these predictors interact across populations and within multifactorial health profiles. This study addresses this gap by assessing the influence of each factor using multivariable regression.

Linear regression was chosen for its ability to estimate the relationship between a continuous outcome—total cholesterol—and multiple predictors. Diagnostic checks were used to evaluate key assumptions, including linearity, homoscedasticity, and normality of residuals. While some violations were observed (e.g., non-normal residuals and heteroscedasticity), potential remedies such as Box-Cox transformations were explored. Despite these limitations, linear regression remains a strong baseline method for identifying statistically significant predictors of cholesterol.

By applying regression analysis to nationally representative NHANES data, this study provides data-driven insights into the factors most strongly associated with cholesterol levels—insights that can help shape future public health strategies.

Data Description

NHANES is a built-in data package from R containing survey data collected by the US National Center for Health Statistics (NCHS) from 2009-2011. The target population of the National Health and Nutrition Examination Surveys (NHANES) is the "the non-institutionalized civilian resident population of the United States". The sample population was obtained via a multistep sampling method, and intentionally oversamples certain minority groups to accurately weight their proportions. These efforts are made with the goal to assess and monitor the health and nutritional conditions of the US population via health checkups and interviews (Centers for Disease Control and Prevention, 2015).

Total high-density lipoprotein (HDL) cholesterol in mmol/L is a variable surveyed by NHANES, captured by the variable *TotChol* in the data package. The sample mean is 5.026 mmol/L and the median is 4.970

mmol/L based on 1289 observations. The nature of this variable being a continuous measurement which follows an approximately normal distribution allows it to be a suitable response variable for a linear regression model. Furthermore, the current literature has evidence that the predictor variables used in our model have an influence upon HDL levels.

Table 1: Summary Statistics of ***TotChol***

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.53	4.24	4.97	5.025725	5.64	10.29

The following is a brief summary of these predictor variables and what current research knows about its relationship to the HDL cholesterol levels in humans.

Age: Age of participant at time of screening in years. The approximately uniform distribution of this variable is indicative of good sampling design. Researchers find that HDL cholesterol levels tend to diminish with age, but there are many other factors that can contribute to this (Ferrera et al., 1997; Milman et al., 2014).

Height and ***Weight***: The standing height of the participant in centimetres, and the weight of the participant in kilograms respectively. The measurements of height follow an approximately standard normal distribution, once again indicative of good sampling techniques.

Weight follows moderate right skew with its measures of central tendency gathering around 80kg (M = 81.16kg, Mdn = 78.8). Seeing as how close the mean and median are to each other and how consistent these statistics are to population parameters captured by the NCHS (Fryar et al., 2021), it is probable that the skew is caused by outliers.

Researchers have shown that Body Mass Index (BMI) scores are negatively associated with HCL cholesterol levels (Lamon-Fava et al., 1996). Note that BMI is a function of height and weight, which both have their respective associations with cholesterol levels (Henriksson et al, 2001).

BPSysAve and ***BPDiaAve***: The systolic and diastolic blood pressure readings of participants in millimeters of mercury. Both follow approximately normal distributions, where BPSysAve (M = 121.69, Mdn = 119.0) has a right skew, and BPDiaAve (M = 70.38, Mdn = 71.0) has a slight left skew. Seeing as how the means and medians of these variables are very close to each other, these skews could be attributed to outliers or natural variation in the population. The current research shows evidence of healthier blood pressure readings in those with higher HDL cholesterol (Al-Jarallah, 2022; Kim et al., 2011, Lye et al., 2009).

SmokeNow: 41.48% of participants (799 participants) have indicated that they currently smoke tobacco, with the remaining 57.52% of participants claiming they do not (590 participants). Several studies have shown evidence of negative associations with smoking and HDL cholesterol levels (Garrison et al., 1978; Nakamura et al., 2021).

DaysPhysActive: The self-reported number of days per week a participant engages in moderate to high intensity physical activity. Most participants claim they generally engage in physical activity 3 times a week. The current research generally tends to agree that consistent physical exercise promotes higher levels of HDL cholesterol (Kodama et al., 1960).

Table 2: Summary Statistics of Continuous Predictors

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Age	21.0	34.0	47.0	47.68425	60.0	80.0
Height	141.3	163.7	170.8	170.57502	177.3	199.6
Weight	41.1	67.6	78.8	81.15741	93.0	172.5
BPSysAve	81.0	110.0	119.0	121.66796	130.0	202.0
BPDiaAve	31.0	64.0	71.0	70.38169	78.0	116.0

Prelminary Model Diagnostics

Model Selection

Preliminary model diagnostics indicated the model would benefit from modifications to improve model fit based on the indications of violated linear regression assumptions. With the primary objective of a predictive model in mind, certain changes were implemented into the model.

A distinct convex curvilinear relationship is evident in Figure Scatterplot Matrix and Figure Residuals vs Age, indicative of a severe violation in linearity. The additional polynomial term *Age2*, the square of the *Age* variable vector, was included to capture this non-linear relationship between the *Age* predictor and dependent variable *TotChol*.
Following this change, the Box-Cox transformation was applied to the dependent variable *TotChol*. This transformation aims to address violations in normality and homoscedasticity as indicated by the right-tailed skew seen in Figure QQ Plot, and fanning patterns of residuals shown by Figure Residuals vs Fitted. Maximum likelihood was used to derive a lambda value ($\lambda = 0.1414$) for the transformation by using functions from the R package *MASS* (Venables & Ripley, 2002) and default built-in algebraic operators. This transformation was not applied to the predictor variables to preserve interpretability.

Remarkable improvements in model assumptions were noticed in the diagnostic plots of the transformed model, such as residual plots showing approximately null relationships with residuals more evenly and widely dispersed across the fitted values. Figure QQ Plot Transformed also now shows the effectiveness of the Box-Cox transformation with its resulting nearly perfect normal distribution in the residuals.

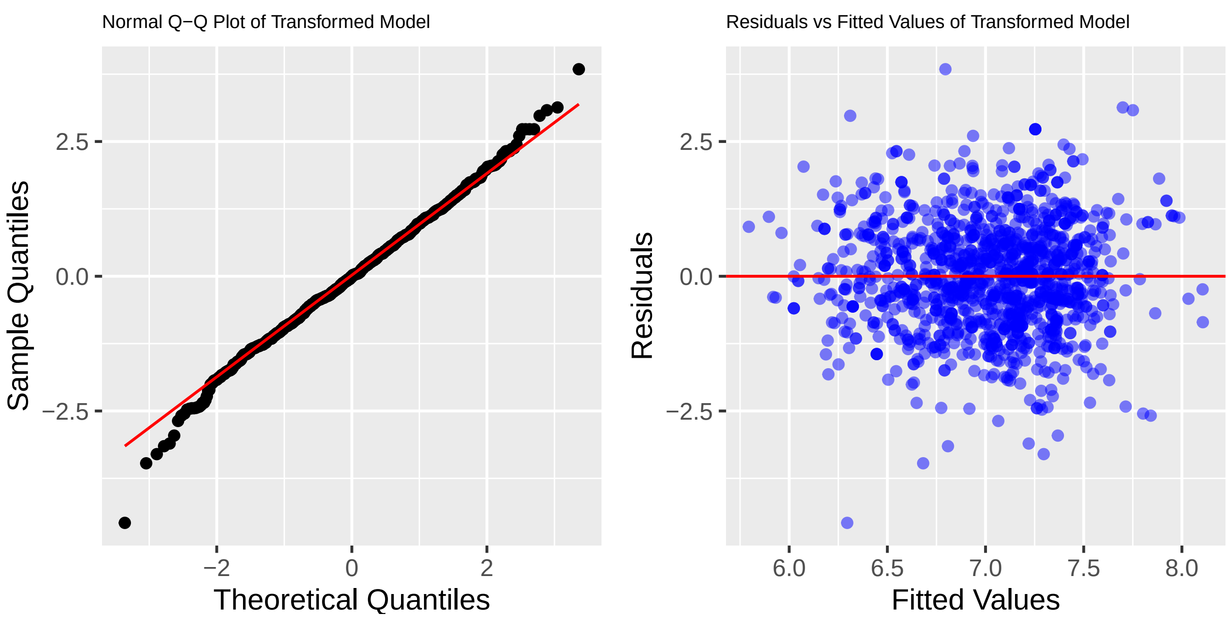


Figure 1: QQ-Plot and Residuals vs Fitted Plot of Transformed Model.

Metrics of model fit in the transformed model ($R^2 = 0.1264$, $adjR^2 = 0.121,$) also showed a large increase when compared to the preliminary model ($R^2 = 0.08085$, $adjR^2 = 0.07583$). Despite the polynomial untransformed model yielding higher fit, severe violations in normality motivated the use of the Box-Cox cocurrently. These metrics compared to other iterations of model transformations can be found in Table of Models.

Table 3: Comparison of Linear Regression Models

Model	R2	Adj_R2	F_value
Preliminary	0.081	0.076	16.10
Polynomial	0.127	0.121	23.20
Box-Cox	0.079	0.074	15.71
Poly and Box-Cox	0.126	0.121	23.15

Upon fitting the transformed model, the dataset was screened for problematic observations. Initial data cleaning ensured the dataset excludes null entries and obvious misinputs, thus the criteria for removing problematic observations was only a matter of measures of influence.

Outliers were identified by checking standardized residuals, and influential observations were identified based on their measurements of leverage, Cook's Distance, Difference in Fits (DFFITS) and Differences in Beta Coefficients (DFBETAS). If an observation had any of these measures surpass their respective thresholds and were concurrently highlighted by the *influenceIndexPlot()* function from the R package *car* (Fox & Weisberg, 2019) they were flagged as problematic observations.

Based on this criteria, five potentially problematic observations were identified, of which only observation 823 and 728 were removed from the data set. A summary of these observations and their measures of leverage are present in Table 2.

Table 4: Measures of Influence and Model Fit After Removing Selected Observations

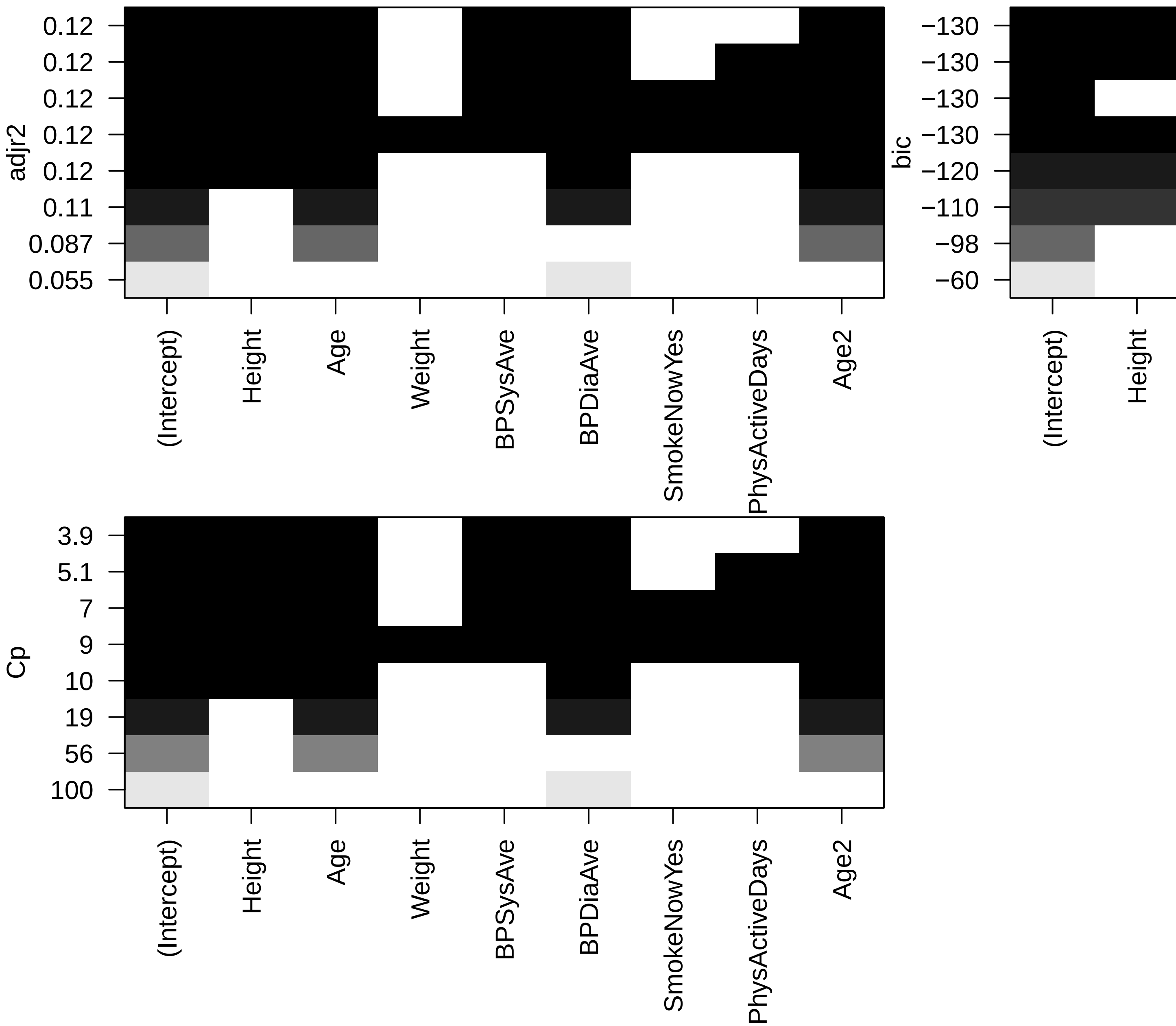
	St. Residual	Cook's Distance	Leverage	DFFITS	Adj.R2 with Obs Removed
10	1.818	0.01501	0.03927	0.3679	0.122
968	-4.662	0.01571	0.00646	-0.3791	0.120
724	-1.306	0.00588	0.03006	-0.2300	0.122
823	3.913	0.01094	0.00639	0.3156	0.125
728	-3.222	0.01468	0.01257	-0.3649	0.123

For each potentially problematic observation, the transformed model was fitted using the same dataset but with the exclusion of the observation under inspection. Models were then compared to determine which observations to remove for highest model fit. By this process, the exclusion of both observations 824 and 728 was found to induce the highest model fit ($R^2 = 0.129$, $adjR^2 = 0.1236$). The exclusion of any of the remaining problematic observations would decrease model fit (see Table 2), and thus with the motivation of a predictive model with high fit, the observations were retained in the dataset.

Several methods of variable selection were employed, such as full and partial F-tests, t-tests for individual predictors, and stepwise regression for AIC and BIC. All of these methods unanimously arrived at the same conclusion of finding the predictors *Age, Age2, Height, BPSysAve* and *BPDiaAve* to be statistically significant.

```
## Subset selection object
## Call: regsubsets.formula(pb.TotChol ~ ., data = clean.frame, nvmax = 8,
##     nbest = 1, really.big = TRUE, method = "exhaustive")
## 8 Variables  (and intercept)
##                Forced in Forced out
## Height             FALSE      FALSE
## Age                FALSE      FALSE
## Weight             FALSE      FALSE
## BPSysAve           FALSE      FALSE
## BPDiaAve           FALSE      FALSE
```

```
## SmokeNowYes         FALSE      FALSE
## PhysActiveDays      FALSE      FALSE
## Age2                FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          Height Age Weight BPSysAve BPDiaAve SmokeNowYes PhysActiveDays Age2
## 1  ( 1 ) " "    " " " "    " "      "*"      " "         " "            " "
## 2  ( 1 ) " "    "*" " "    " "      " "      " "         " "            "*"
## 3  ( 1 ) " "    "*" " "    " "      "*"      " "         " "            "*"
## 4  ( 1 ) "*"    "*" " "    " "      "*"      " "         " "            "*"
## 5  ( 1 ) "*"    "*" " "    "*"      "*"      " "         " "            "*"
## 6  ( 1 ) "*"    "*" " "    "*"      "*"      " "         "*"            "*"
## 7  ( 1 ) "*"    "*" " "    "*"      "*"      "*"         "*"            "*"
## 8  ( 1 ) "*"    "*" "*"    "*"      "*"      "*"         "*"            "*"
```



```
## Start:  AIC=-54.33
## pb.TotChol ~ Age + Age2 + Weight + Height + BPSysAve + BPDiaAve +
##     SmokeNow + PhysActiveDays
##
##                  Df Sum of Sq    RSS     AIC
## - Weight          1     0.000 1216.7 -56.334
## - SmokeNow        1     0.086 1216.8 -56.244
## - PhysActiveDays  1     0.811 1217.5 -55.476
## <none>                        1216.7 -54.335
## - BPSysAve        1     7.884 1224.5 -48.022
## - Height          1     8.333 1225.0 -47.550
## - BPDiaAve        1    19.475 1236.1 -35.897
## - Age2            1    65.377 1282.0  11.028
## - Age             1    74.647 1291.3  20.300
##
## Step:  AIC=-56.33
## pb.TotChol ~ Age + Age2 + Height + BPSysAve + BPDiaAve + SmokeNow +
##     PhysActiveDays
##
##                  Df Sum of Sq    RSS     AIC
## - SmokeNow        1     0.088 1216.8 -58.241
## - PhysActiveDays  1     0.811 1217.5 -57.476
## <none>                        1216.7 -56.334
## + Weight          1     0.000 1216.7 -54.335
## - BPSysAve        1     7.936 1224.6 -49.967
## - Height          1    10.536 1227.2 -47.237
## - BPDiaAve        1    19.546 1236.2 -37.823
## - Age2            1    65.904 1282.6   9.557
## - Age             1    75.216 1291.9  18.868
##
## Step:  AIC=-58.24
## pb.TotChol ~ Age + Age2 + Height + BPSysAve + BPDiaAve + PhysActiveDays
##
##                  Df Sum of Sq    RSS     AIC
## - PhysActiveDays  1     0.811 1217.6 -59.384
## <none>                        1216.8 -58.241
## + SmokeNow        1     0.088 1216.7 -56.334
## + Weight          1     0.003 1216.8 -56.244
## - BPSysAve        1     8.071 1224.8 -51.731
## - Height          1    10.615 1227.4 -49.062
## - BPDiaAve        1    19.459 1236.2 -39.821
## - Age2            1    66.037 1282.8   7.779
## - Age             1    75.131 1291.9  16.872
##
## Step:  AIC=-59.38
## pb.TotChol ~ Age + Age2 + Height + BPSysAve + BPDiaAve
##
##                  Df Sum of Sq    RSS     AIC
## <none>                        1217.6 -59.384
## + PhysActiveDays  1     0.811 1216.8 -58.241
## + SmokeNow        1     0.088 1217.5 -57.476
## + Weight          1     0.000 1217.6 -57.384
## - BPSysAve        1     7.982 1225.5 -52.974
## - Height          1    10.444 1228.0 -50.391
```

```
## - BPDiaAve        1    19.562 1237.1 -40.870
## - Age2            1    65.411 1283.0   5.965
## - Age             1    74.398 1292.0  14.949
```

```
##
## Call:
## lm(formula = pb.TotChol ~ Age + Age2 + Height + BPSysAve + BPDiaAve,
##     data = clean.frame)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5880 -0.6170 -0.0140  0.6438  3.1057
##
## Coefficients:
##               Estimate Std. Error t value             Pr(>|t|)
## (Intercept)  4.8098341  0.5741846   8.377 < 0.0000000000000002 ***
## Age          0.0974977  0.0110200   8.847 < 0.0000000000000002 ***
## Age2        -0.0009263  0.0001117  -8.296  0.00000000000000027 ***
## Height      -0.0098211  0.0029628  -3.315             0.000943 ***
## BPSysAve     0.0056469  0.0019487   2.898             0.003821 **
## BPDiaAve     0.0127101  0.0028016   4.537  0.00000624991387098 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9749 on 1281 degrees of freedom
## Multiple R-squared:  0.1284, Adjusted R-squared:  0.125
## F-statistic: 37.73 on 5 and 1281 DF,  p-value: < 0.00000000000000022
```

FINAL MODEL


```
##
## Call:
## lm(formula = pb.TotChol ~ Age + Age2 + Height + BPSysAve + BPDiaAve,
##     data = clean.frame)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5880 -0.6170 -0.0140  0.6438  3.1057
##
## Coefficients:
##               Estimate Std. Error t value             Pr(>|t|)
## (Intercept)  4.8098341  0.5741846   8.377 < 0.0000000000000002 ***
## Age          0.0974977  0.0110200   8.847 < 0.0000000000000002 ***
## Age2        -0.0009263  0.0001117  -8.296  0.00000000000000027 ***
## Height      -0.0098211  0.0029628  -3.315             0.000943 ***
## BPSysAve     0.0056469  0.0019487   2.898             0.003821 **
## BPDiaAve     0.0127101  0.0028016   4.537  0.00000624991387098 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9749 on 1281 degrees of freedom
## Multiple R-squared:  0.1284, Adjusted R-squared:  0.125
## F-statistic: 37.73 on 5 and 1281 DF,  p-value: < 0.00000000000000022
```


```
##                      2.5 %        97.5 %
```

```
## (Intercept)  3.683388750  5.9362795086
## Age          0.075878426  0.1191170717
## Age2        -0.001145301 -0.0007072109
## Height      -0.015633630 -0.0040086511
## BPSysAve     0.001823995  0.0094698317
## BPDiaAve     0.007213819  0.0182063184
```

Prediction Accuracy and Model Validation

```
## [1] 0.9542581
```

```
## [1] 0.9768613
```

```
## Linear Regression
##
## 1287 samples
##    5 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1158, 1159, 1158, 1158, 1158, 1159, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.9751559  0.1373103  0.7694394
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
## Linear Regression
##
## 1287 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1159, 1160, 1160, 1157, 1159, 1158, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.9751863  0.1249023  0.7700059
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo,
## : There were missing values in resampled performance measures.
```

```
## Linear Regression
##
## 1287 samples
##    1 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1159, 1159, 1159, 1159, 1158, 1158, ...
## Resampling results:
##
##   RMSE      Rsquared  MAE
```

```
##   1.040114  NaN       0.8219892
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
## Linear Regression
##
## 1289 samples
##    4 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1161, 1159, 1162, 1160, 1160, 1159, ...
## Resampling results:
##
##   RMSE      Rsquared    MAE
##   1.040968  0.08495365  0.811146
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
## Linear Regression
##
## 1289 samples
##    7 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1159, 1160, 1161, 1160, 1160, 1160, ...
## Resampling results:
##
##   RMSE      Rsquared    MAE
##   1.042695  0.07761385  0.8123572
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo,
## : There were missing values in resampled performance measures.
```

```
## glmnet
##
## 1287 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1157, 1159, 1159, 1159, 1157, 1157, ...
## Resampling results across tuning parameters:
##
##   lambda        RMSE       Rsquared    MAE
##    0.0001000000  0.9758988  0.12666823  0.7709795
##    0.0001123324  0.9758988  0.12666823  0.7709795
##    0.0001261857  0.9758999  0.12666666  0.7709812
##    0.0001417474  0.9759009  0.12666285  0.7709869
##    0.0001592283  0.9758998  0.12666193  0.7709929
##    0.0001788650  0.9758984  0.12666299  0.7709989
##    0.0002009233  0.9758963  0.12666452  0.7710054
##    0.0002257020  0.9758946  0.12666494  0.7710128
```

```
## 0.0002535364 0.9758924 0.12666557 0.7710213
## 0.0002848036 0.9758909 0.12666603 0.7710309
## 0.0003199267 0.9758887 0.12666674 0.7710411
## 0.0003593814 0.9758860 0.12666812 0.7710534
## 0.0004037017 0.9758836 0.12666920 0.7710672
## 0.0004534879 0.9758815 0.12666968 0.7710833
## 0.0005094138 0.9758796 0.12666983 0.7711006
## 0.0005722368 0.9758785 0.12666887 0.7711209
## 0.0006428073 0.9758784 0.12666621 0.7711437
## 0.0007220809 0.9758798 0.12666199 0.7711696
## 0.0008111308 0.9758819 0.12665717 0.7712009
## 0.0009111628 0.9758863 0.12665024 0.7712405
## 0.0010235310 0.9758936 0.12663926 0.7712890
## 0.0011497570 0.9759057 0.12662259 0.7713466
## 0.0012915497 0.9759227 0.12660071 0.7714115
## 0.0014508288 0.9759474 0.12656973 0.7714889
## 0.0016297508 0.9759794 0.12653072 0.7715771
## 0.0018307383 0.9760233 0.12647811 0.7716849
## 0.0020565123 0.9760816 0.12640925 0.7718162
## 0.0023101297 0.9761577 0.12631877 0.7719638
## 0.0025950242 0.9762578 0.12619962 0.7721312
## 0.0029150531 0.9763887 0.12604303 0.7723414
## 0.0032745492 0.9765584 0.12583663 0.7725954
## 0.0036783798 0.9767756 0.12557097 0.7728794
## 0.0041320124 0.9770538 0.12522643 0.7732245
## 0.0046415888 0.9774168 0.12476170 0.7736467
## 0.0052140083 0.9778836 0.12414338 0.7741313
## 0.0058570208 0.9784840 0.12331670 0.7747493
## 0.0065793322 0.9792591 0.12220102 0.7755570
## 0.0073907220 0.9802429 0.12072598 0.7765613
## 0.0083021757 0.9814818 0.11878954 0.7778023
## 0.0093260335 0.9829998 0.11633587 0.7792333
## 0.0104761575 0.9849164 0.11308017 0.7809314
## 0.0117681195 0.9873332 0.10877773 0.7829471
## 0.0132194115 0.9904062 0.10305807 0.7853466
## 0.0148496826 0.9943045 0.09554966 0.7883378
## 0.0166810054 0.9986372 0.08715637 0.7917896
## 0.0187381742 1.0000214 0.08473350 0.7929129
## 0.0210490414 1.0000562 0.08494733 0.7929042
## 0.0236448941 1.0001824 0.08502703 0.7929373
## 0.0265608778 1.0004079 0.08495207 0.7930512
## 0.0298364724 1.0007144 0.08481217 0.7932449
## 0.0335160265 1.0010958 0.08463082 0.7935107
## 0.0376493581 1.0015599 0.08441186 0.7938229
## 0.0422924287 1.0021305 0.08413908 0.7942505
## 0.0475081016 1.0028265 0.08379651 0.7947587
## 0.0533669923 1.0036968 0.08331091 0.7953587
## 0.0599484250 1.0047976 0.08259270 0.7960690
## 0.0673415066 1.0061616 0.08156668 0.7969913
## 0.0756463328 1.0076840 0.08041895 0.7980038
## 0.0849753436 1.0092618 0.07970835 0.7991879
## 0.0954548457 1.0110954 0.07890115 0.8006928
## 0.1072267222 1.0134043 0.07743096 0.8026675
## 0.1204503540 1.0163388 0.07457984 0.8051721
```

```
##    0.1353047775  1.0197924  0.07010347  0.8080000
##    0.1519911083  1.0230368  0.06637875  0.8105885
##    0.1707352647  1.0260602  0.06605730  0.8126892
##    0.1917910262  1.0297810  0.06605634  0.8152489
##    0.2154434690  1.0344526  0.06605634  0.8184422
##    0.2420128265  1.0395529  0.03679097  0.8215551
##    0.2718588243  1.0403394         NaN  0.8219620
##    0.3053855509  1.0403394         NaN  0.8219620
##    0.3430469286  1.0403394         NaN  0.8219620
##    0.3853528594  1.0403394         NaN  0.8219620
##    0.4328761281  1.0403394         NaN  0.8219620
##    0.4862601580  1.0403394         NaN  0.8219620
##    0.5462277218  1.0403394         NaN  0.8219620
##    0.6135907273  1.0403394         NaN  0.8219620
##    0.6892612104  1.0403394         NaN  0.8219620
##    0.7742636827  1.0403394         NaN  0.8219620
##    0.8697490026  1.0403394         NaN  0.8219620
##    0.9770099573  1.0403394         NaN  0.8219620
##    1.0974987655  1.0403394         NaN  0.8219620
##    1.2328467394  1.0403394         NaN  0.8219620
##    1.3848863714  1.0403394         NaN  0.8219620
##    1.5556761439  1.0403394         NaN  0.8219620
##    1.7475284000  1.0403394         NaN  0.8219620
##    1.9630406500  1.0403394         NaN  0.8219620
##    2.2051307399  1.0403394         NaN  0.8219620
##    2.4770763560  1.0403394         NaN  0.8219620
##    2.7825594022  1.0403394         NaN  0.8219620
##    3.1257158497  1.0403394         NaN  0.8219620
##    3.5111917342  1.0403394         NaN  0.8219620
##    3.9442060594  1.0403394         NaN  0.8219620
##    4.4306214576  1.0403394         NaN  0.8219620
##    4.9770235643  1.0403394         NaN  0.8219620
##    5.5908101825  1.0403394         NaN  0.8219620
##    6.2802914418  1.0403394         NaN  0.8219620
##    7.0548023107  1.0403394         NaN  0.8219620
##    7.9248289835  1.0403394         NaN  0.8219620
##    8.9021508545  1.0403394         NaN  0.8219620
##   10.0000000000  1.0403394         NaN  0.8219620
##
## Tuning parameter 'alpha' was held constant at a value of 1
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 1 and lambda = 0.0006428073.
```

```
##                  Coefficient       Variable
## (Intercept)     4.9053321709    (Intercept)
## Height         -0.0097653068         Height
## Age             0.0941196741            Age
## BPSysAve        0.0054486565       BPSysAve
## BPDiaAve        0.0130758073       BPDiaAve
## SmokeNowYes     0.0163610102    SmokeNowYes
## PhysActiveDays -0.0133636084 PhysActiveDays
## Age2           -0.0008885573           Age2
```

```
## Linear Regression
##
```

```
## 1287 samples
##    7 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1158, 1158, 1158, 1159, 1159, 1159, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.9772544  0.1289544  0.7713758
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```