data\_science2\_midterm

Amanda Howarth

3/29/2020

library(tidyverse)  
library(dplyr)  
library(sqldf)  
library(caret)  
library(ModelMetrics)  
library(ISLR)  
library(pls)  
library(mgcv)  
library(microbenchmark)  
library(stargazer)  
library(viridis)  
library(knitr)

# Background / Motivation

Total hospital charges vary greatly for patients across the United States. In 2012, it was estimated that the mean hospital cost per stay was $10,400 (Moore et al., 2014). However, the cost per stay varies by different factors, including a patient’s demographic information, the severity of their health condition, and type of admission. The costs of having a baby and the costs of infant hospitalizations also differ based on many different factors. In 2004, March of Dimes reported that the average cost of having a baby was $8,802. This cost varied by factors such as type of birth, whether it be a Cesarean-section of vaginal birth ($10,958 vs $7,737) (March of Dimes, 2007). This cost was also affected by the birthweight of the newborn, with the average cost per stay increasing to $15,100 for pre-term/low birthweight babies (Russell et al, 2007). For babies born at a weight less than 1000 grams, the average cost of hospital stay was reported to be $65,600 (Russell et al, 2007).

The ability to predict the total cost of a patient’s stay at a hospital based on known patient characteristics, hospital information, and diagnosis would provide significant benefits for health insurance companies, patients, and hospital systems. If a soon-to-be mother was knowledgeable of the predicted charges of her upcoming birth at different hospitals, she would be better informed in choosing which hospital she would like to be admitted to. Additionally, a newborn/infant’s parent can choose a hospital that is best for their family economically if their child is having health complications.

The goal of this analysis is to determine the best model to predict the total healthcare charges of newborn and infant patients who were admitted to hospitals in New York City. We used the Hospital Inpatient Discharges (SPARCS De-Identified) 2013 data set. This data was provided by the New York State Department of Health’s Office of Quality and Patient Safety and includes 2.43 million observations and 34 variables, such as features for demographic information, hospital stay, payment typology, cost information.

In order to load the dataset to my computer, this dataset was limited (through the NY State’s website filter) to the hospital service area of New York City and includes information from hospitals in all five boroughs. The dataset was limited to include only patients who had birthweight information available, recorded during their hospital stay. Over 150+ CCS diagnosis codes were included in the original dataset. The CCS diagnosis codes were categorized into 17 diagnosis types, such as “infectious diseases” and" pregnancy and childbirth complications." Individuals with missing values (n = 119) were dropped from the dataset, totaling 121,380 patients in the final dataset.

Nineteen variables were dropped from the dataset. Variables were dropped if they were redundant of other variables or if they were not deemed to be clinically meaningful in predicting healthcare hospitalization charges. All categorical variables were turned into dummy variables with referent groups. The final dataset used for models in this analysis (with dummy variables) included 92 predictors and 1 response variable (total charges).

Our research questions were the following: • Which approach (linear regression, lasso regression, ridge regression, principle component regression, or generalized additive model) best predicts a patient’s total healthcare charges per hospital stay in New York? • What is the root mean square error (RMSE) of each model? • What percent of the variance in total healthcare charges is explained by the predictors available in SPARCS?

# Import and clean data

discharge\_data = read\_csv("./data/Hospital\_Inpatient\_Discharges\_\_SPARCS\_De-Identified\_\_\_2013-4.csv") %>%  
 janitor::clean\_names() %>%  
 mutate(birth\_weight = as.numeric(birth\_weight)) %>%   
 mutate(length\_of\_stay = as.numeric(length\_of\_stay)) %>%   
 janitor::clean\_names() %>%  
select(-ccs\_diagnosis\_description, -ccs\_procedure\_code, -ccs\_procedure\_description, -apr\_drg\_description, -apr\_mdc\_description, -apr\_severity\_of\_illness\_description, -facility\_id, -payment\_typology\_2, -payment\_typology\_3, -zip\_code\_3\_digits, -discharge\_year, -health\_service\_area, -operating\_certificate\_number, -abortion\_edit\_indicator, -patient\_disposition, -age\_group, -apr\_drg\_code, -apr\_mdc\_code) %>%   
 filter(type\_of\_admission != "Not Available", payment\_typology\_1 != "Unknown", payment\_typology\_1 != "Miscellaneous/Other", apr\_medical\_surgical\_description != "Not Applicable") %>%   
 mutate(ccs = ifelse(ccs\_diagnosis\_code %in% 1:10, "infectious\_disease",   
 ifelse(ccs\_diagnosis\_code %in% 11:47, "cancer",  
 ifelse(ccs\_diagnosis\_code %in% 48:58, "endocrine\_metabolic\_disease",  
 ifelse(ccs\_diagnosis\_code %in% 59:64, "blood\_diseases",  
 ifelse(ccs\_diagnosis\_code %in% 76:95, "nervous\_system\_disease",  
 ifelse(ccs\_diagnosis\_code %in% 96:121, "circulatory\_sysytem\_disease",  
 ifelse(ccs\_diagnosis\_code %in% 122:134, "respiratory\_disease",  
 ifelse(ccs\_diagnosis\_code %in% 135:155, "digestive\_disease",   
 ifelse(ccs\_diagnosis\_code %in% 156:166, "genitourinary\_disease",   
 ifelse(ccs\_diagnosis\_code %in% 167:196, "pregnancy\_childbirth\_complication",  
 ifelse(ccs\_diagnosis\_code %in% 197:200, "skin\_disease",   
 ifelse(ccs\_diagnosis\_code %in% 201:212, "musculoskeletal\_disease",  
 ifelse(ccs\_diagnosis\_code %in% 213:217, "congenital\_anomalies",  
 ifelse(ccs\_diagnosis\_code %in% 218:224, "perinatal\_condition",  
 ifelse(ccs\_diagnosis\_code %in% 225:244, "injury\_poisoning",  
 ifelse(ccs\_diagnosis\_code %in% 650:663, "mental\_disorder", "other"))))))))))))))))) %>%   
 select(-ccs\_diagnosis\_code) %>%   
 na.omit()

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## `Operating Certificate Number` = col\_double(),  
## `Facility ID` = col\_double(),  
## `Zip Code - 3 digits` = col\_double(),  
## `Discharge Year` = col\_double(),  
## `CCS Diagnosis Code` = col\_double(),  
## `CCS Procedure Code` = col\_double(),  
## `APR DRG Code` = col\_double(),  
## `APR MDC Code` = col\_double(),  
## `APR Severity of Illness Code` = col\_double(),  
## `Birth Weight` = col\_double(),  
## `Total Charges` = col\_double(),  
## `Total Costs` = col\_double()  
## )

## See spec(...) for full column specifications.

## Warning: 3301 parsing failures.  
## row col expected actual file  
## 2649 Zip Code - 3 digits a double OOS './data/Hospital\_Inpatient\_Discharges\_\_SPARCS\_De-Identified\_\_\_2013-4.csv'  
## 3543 Zip Code - 3 digits a double OOS './data/Hospital\_Inpatient\_Discharges\_\_SPARCS\_De-Identified\_\_\_2013-4.csv'  
## 3670 Zip Code - 3 digits a double OOS './data/Hospital\_Inpatient\_Discharges\_\_SPARCS\_De-Identified\_\_\_2013-4.csv'  
## 3838 Zip Code - 3 digits a double OOS './data/Hospital\_Inpatient\_Discharges\_\_SPARCS\_De-Identified\_\_\_2013-4.csv'  
## 3937 Zip Code - 3 digits a double OOS './data/Hospital\_Inpatient\_Discharges\_\_SPARCS\_De-Identified\_\_\_2013-4.csv'  
## .... ................... ........ ...... .........................................................................  
## See problems(...) for more details.

## Warning: NAs introduced by coercion

# Exploratory Analysis Findings / Visualization

This dataset includes 121,380 patients who had their birthweight recorded at the time of their hospital visit. The mean total charges for a patient at discharge across the 49 hospital facilities was $19,787.95. The median total charges for a patient was $6,590.02. For exploratory analysis, we compared the median total charges because there were extreme outliers with very high charges. The median is a better measure of central tendency than the mean. The median total charges varied widely across the 49 facilities included in this analysis. The median total charges for a patient at discharge were highest at Montefiore Medical Center - Henry & Lucy Moses Div ($25,160.36) and lowest at North Central Bronx Hospital ($2,407.01) (Table 1).

In Figure 1, there is a positive correlation between length of stay and the total charges at discharge. The relationship between length of stay and total charges demonstrates linearity. Thus, as the length of stay increases, the total charges increased. The charges appear to be relatively lower at hospitals in Richmond County (Staten Island), even as the number of days spent in the hospital increases. Additionally, it appears that the charges at hospitals in Manhattan are typically higher compared to other counties across the lengths of stay.

Figure 2 demonstrates that the total charges generally decrease across increasing birthweight. There is a moderately negative correlation between baby birthweight and total charges. Thus, patients with lower birthweights pay the highest total charges at discharge.

# Table 1

#Table 1. Median Total Charges by Hospital Facility  
charges\_by\_hospital = discharge\_data %>%   
 select(facility\_name, total\_charges, hospital\_county) %>%   
 group\_by(facility\_name, hospital\_county) %>%   
 summarize(median\_charge = median(total\_charges)) %>%  
 arrange(desc(median\_charge)) %>%   
 knitr::kable()  
charges\_by\_hospital

|  |  |  |
| --- | --- | --- |
| facility\_name | hospital\_county | median\_charge |
| Montefiore Medical Center - Henry & Lucy Moses Div | Bronx | 25160.36 |
| Brookdale Hospital Medical Center | Kings | 11561.00 |
| Long Island Jewish Schneiders Children’s Hospital Division | Queens | 11498.00 |
| NYU Hospitals Center | Manhattan | 10573.70 |
| New York Presbyterian Hospital - Columbia Presbyterian Center | Manhattan | 9383.75 |
| NewYork-Presbyterian/Queens | Queens | 9361.96 |
| New York Presbyterian Hospital - Allen Hospital | Manhattan | 9335.04 |
| New York Presbyterian Hospital - New York Weill Cornell Center | Manhattan | 8879.03 |
| Lenox Hill Hospital | Manhattan | 8868.30 |
| Montefiore Medical Center-Wakefield Hospital | Bronx | 8520.85 |
| Montefiore Medical Center - North Division | Bronx | 8083.35 |
| Forest Hills Hospital | Queens | 7946.70 |
| Mount Sinai West | Manhattan | 7897.00 |
| New York Hospital Medical Center of Queens | Queens | 7592.46 |
| New York Hospital Medical Center of Queens | Queens | 7522.46 |
| Montefiore Med Center - Jack D Weiler Hosp of A Einstein College Div | Bronx | 7136.95 |
| University Hospital of Brooklyn | Kings | 6771.00 |
| New York-Presbyterian/Queens | Queens | 6744.78 |
| Coney Island Hospital | Kings | 6331.11 |
| Mount Sinai Roosevelt | Manhattan | 6314.50 |
| Lutheran Medical Center | Kings | 6286.70 |
| St Lukes Roosevelt Hospital Center - Roosevelt Hospital Division | Manhattan | 5980.50 |
| New York Methodist Hospital | Kings | 5719.94 |
| Beth Israel Medical Center/Petrie Campus | Manhattan | 5676.00 |
| Mount Sinai Beth Israel | Manhattan | 5673.40 |
| SUNY Downstate Medical Center at LICH | Kings | 5367.00 |
| Maimonides Medical Center | Kings | 4993.00 |
| St Johns Episcopal Hospital So Shore | Queens | 4910.00 |
| Staten Island University Hosp-North | Richmond | 4820.00 |
| Mount Sinai Hospital | Manhattan | 4787.20 |
| Woodhull Medical & Mental Health Center | Kings | 4547.34 |
| Kings County Hospital Center | Kings | 4540.02 |
| Queens Hospital Center | Queens | 4527.31 |
| Bellevue Hospital Center | Manhattan | 4527.26 |
| Brooklyn Hospital Center - Downtown Campus | Kings | 4470.44 |
| Jacobi Medical Center | Bronx | 4461.91 |
| Richmond University Medical Center | Richmond | 4438.52 |
| St Barnabas Hospital | Bronx | 4394.39 |
| SBH Health System | Bronx | 4391.21 |
| Lincoln Medical & Mental Health Center | Bronx | 4376.42 |
| Harlem Hospital Center | Manhattan | 4361.67 |
| Wyckoff Heights Medical Center | Kings | 4335.00 |
| Metropolitan Hospital Center | Manhattan | 4299.00 |
| Elmhurst Hospital Center | Queens | 4285.44 |
| Jamaica Hospital Medical Center | Queens | 3692.00 |
| New York-Presbyterian/Lower Manhattan Hospital | Manhattan | 3498.26 |
| Flushing Hospital Medical Center | Queens | 3397.00 |
| Bronx-Lebanon Hospital Center - Concourse Division | Bronx | 3370.57 |
| North Central Bronx Hospital | Bronx | 2407.01 |

#mean total charges = $19,787.95  
mean(discharge\_data$total\_charges)

## [1] 19787.95

#median total charges =$6590.02  
median(discharge\_data$total\_charges)

## [1] 6590.015

# Figures 1 and 2

ggplot(discharge\_data, aes(x = length\_of\_stay, y = total\_charges, group = hospital\_county, color = hospital\_county)) +   
 geom\_point(na.rm = TRUE) +  
 labs(title = "Figure 1. Total Charges by Length of Stay in NYC Hospitals",  
 x = "Length of Stay",  
 y = "Total Charges")

ggplot(discharge\_data, aes(x = birth\_weight, y = total\_charges, group = emergency\_department\_indicator)) +   
 geom\_point(na.rm = TRUE) +  
 labs(title = "Figure 2. Total Charges Across Values of Baby Birthweight in NYC Hospitals",  
 x = "Baby Birthweight",  
 y = "Total Charges")

# Making categorical varaibles into dummy variables

cat\_dummies = dummyVars(" ~ .", data = discharge\_data, fullRank = T)  
discharge\_data2 = data.frame(predict(cat\_dummies, newdata = discharge\_data))

# Partitioning data into test and training data

set.seed(100)  
trRows <- createDataPartition(discharge\_data2$total\_charges,  
 p = .75,  
 list = FALSE)  
## p = 0.75 means u are taking 75% of the data and thus, 25% is the test set and 75% is ur training set  
  
# matrix of predictors & vector of response (training data)  
x <- model.matrix(total\_charges~.,discharge\_data2)[trRows,-1]  
y <- discharge\_data2$total\_charges[trRows]  
  
# matrix of predictors & vector of response (test data)  
x2 <- model.matrix(total\_charges~., discharge\_data2)[-trRows,-1]  
y2 <- discharge\_data2$total\_charges[-trRows]

# MODELS

Models We fit 5 different models using cross validation to determine which model would fit the data best: • One linear model, • Three regularized linear models (Lasso regression, Ridge regression, PCR), and • One non-linear model (Generalized Additive Model).

Predictor variables included: 1) hospital county (Manhattan, Bronx, King’s County, Richmond County, and Queens), 2) hospital facility, 3)gender, 4) race, 5) ethnicity, 6) length of stay, 7) type of admission, 8) APR severity of illness code, 9) APR medical-surgical description (medical or surgical), 10) type of payment, 11) birth weight, 12) emergency department indicator, 13) APR risk of mortality, 14) total costs, and 15) type of CCS diagnosis. All categorical variables were made into dummy variables. Three predictors were continuous (length of stay, birthweight, and total costs) as well as the outcome variable (total charges).

Our dataset (92 predictors total, with dummy variables) was partitioned into a training and test data set. Five models were fitted using the training data and the mean squared error (MSE) was calculated for each model using the test data. We measured the mean squared error to quantify the extent to which the predicted response value for a given observation is close to the true response value for that observation.

First, we fit a linear model using least squares on all the predictors in the training data. We found that the MSE calculated on the test data was 937604771. By looking at the coefficient values for the linear model, some of the dummy variables were not significant at an alpha value of 0.05 in predicting total charges, including ethnicity, emergency department indicator, and certain CCS diagnosis types.

Next, we fit two models on all the predictor variables using two different techniques that “shrink” the coefficient estimates towards zero, which reduces variance. We fit a ridge regression model on the training data. Alpha was held at a value of 0 and our final lambda value chosen by cross-validation was 5355.389. Using our test data, we found that test error (MSE) was 9790669585. Next, we fit a lasso model on the training data. Alpha was held at a value of 1 and our final lambda value chosen by cross-validation was 28.03162. Our test error was 937116653.

We also fit a principle component regression (PCR) model on the training data with M chosen by cross validation. The PCR method constructs the first M principal components. Our M-value was 87 and our test error was 937635353. A benefit of using PCR is that this method avoids multicollinearity between variables in the dataset. A limitation of PCR is that there is no guarantee that Z¬m¬ are the best linear combinations of the variables in predicting the response.

Lastly, we fit a GAM model to our training dataset. The benefit of using a GAM model is that it will automatically model non-linear relationships that standard linear regression would miss. S functions were applied to the three continuous variables in the model (total costs, length of stay, and birthweight).

The best model for predicting total charges was PCR. This model was chosen because it has the smallest cross validation RMSE, a value of 28173.68 (Figure 3). For the PCR model, 85.69% of the variance in total charges is explained by the predictors in the model. The mean RMSE value of the lasso regression model (28174.24) and linear regression model (28180.84) were very close to the PCR model (depicted in Figure 4). Overall, GAM had the lowest RMSE (29076.52), which was unexpected because it was the only model to account for non-linear relationships that a standard linear regression model would miss.

For the PCR model, the top 10 variables that played important roles in predicting a patient’s total charges are displayed in Figure 4. These variables were chosen using the varImp() function, which automatically chooses a measure of variable importance that is appropriate for given algorithms. The top five variables included length of stay, total costs, APR risk of mortality (major), APR risk of mortality (minor), and the APR severity of illness code. The greatest importance in predicting total healthcare charges was length of stay in the hospital, which aligns with the linear relationship depicted in Figure 1.

Limitations of this analysis included that non-significant variables were included in models and not dropped to improve the model. This was done in an effort to compare models, by first including all predictors in the model building phase. Thus, if we were to drop variables that are neither clinically meaningful nor significant, this may have created more parsimonious and efficient models. Additionally, the highest R^2 value was 85.69%. The remaining variance of total charges are thus unexplained and may be due to variables that were dropped (such as patient’s zip code) or by features not collected in this dataset.

# LINEAR MODEL

Fitting a linear model using least squares on the training data and calculating the mean square error using the test data.

set.seed(100)  
ctrl1 <- trainControl(method = "cv", number = 10)  
  
fit\_lm <- train(total\_charges~.,   
 data = discharge\_data2[trRows,-1],   
 method = "lm",   
 trControl = ctrl1)

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading  
  
## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading

fit\_lm

## Linear Regression   
##   
## 91036 samples  
## 91 predictor  
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 81932, 81933, 81933, 81933, 81932, 81933, ...   
## Resampling results:  
##   
## RMSE Rsquared MAE   
## 28174.24 0.8568802 9541.894  
##   
## Tuning parameter 'intercept' was held constant at a value of TRUE

summary(fit\_lm)

##   
## Call:  
## lm(formula = .outcome ~ ., data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -726284 -4513 436 5124 851654   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate  
## (Intercept) -7.858e+03  
## hospital\_countyManhattan -1.095e+03  
## hospital\_countyQueens 3.870e+03  
## hospital\_countyRichmond 2.663e+03  
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus 1.445e+04  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division -9.793e+03  
## facility\_nameBrookdale.Hospital.Medical.Center 1.176e+03  
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus 1.497e+03  
## facility\_nameConey.Island.Hospital 5.707e+03  
## facility\_nameElmhurst.Hospital.Center -3.334e+02  
## facility\_nameFlushing.Hospital.Medical.Center 2.240e+03  
## facility\_nameForest.Hills.Hospital 7.721e+03  
## facility\_nameHarlem.Hospital.Center -2.544e+03  
## facility\_nameJacobi.Medical.Center -1.627e+03  
## facility\_nameJamaica.Hospital.Medical.Center -2.184e+03  
## facility\_nameKings.County.Hospital.Center 8.359e+02  
## facility\_nameLenox.Hill.Hospital 2.030e+04  
## facility\_nameLincoln.Medical...Mental.Health.Center -5.839e+02  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 1.920e+04  
## facility\_nameLutheran.Medical.Center 1.169e+04  
## facility\_nameMaimonides.Medical.Center 1.436e+04  
## facility\_nameMetropolitan.Hospital.Center 1.854e+03  
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div 2.304e+04  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div 3.752e+04  
## facility\_nameMontefiore.Medical.Center...North.Division 2.113e+04  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital 2.002e+04  
## facility\_nameMount.Sinai.Beth.Israel 1.403e+04  
## facility\_nameMount.Sinai.Hospital 9.239e+03  
## facility\_nameMount.Sinai.Roosevelt 4.829e+03  
## facility\_nameMount.Sinai.West -1.559e+04  
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens 9.083e+03  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens 9.644e+03  
## facility\_nameNew.York.Methodist.Hospital 9.626e+03  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital 1.488e+04  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center 2.382e+04  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center 1.835e+04  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital 1.119e+04  
## facility\_nameNew.York.Presbyterian.Queens 9.595e+03  
## facility\_nameNewYork.Presbyterian.Queens 7.077e+03  
## facility\_nameNorth.Central.Bronx.Hospital 1.853e+03  
## facility\_nameNYU.Hospitals.Center 1.479e+04  
## facility\_nameQueens.Hospital.Center -1.758e+03  
## facility\_nameRichmond.University.Medical.Center 5.886e+03  
## facility\_nameSBH.Health.System -3.406e+03  
## facility\_nameSt.Barnabas.Hospital -9.094e+02  
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore NA  
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division 8.133e+03  
## facility\_nameStaten.Island.University.Hosp.North NA  
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH -4.371e+03  
## facility\_nameUniversity.Hospital.of.Brooklyn -1.109e+04  
## facility\_nameWoodhull.Medical...Mental.Health.Center 1.790e+02  
## facility\_nameWyckoff.Heights.Medical.Center NA  
## genderM -3.604e+02  
## raceOther.Race 7.354e+02  
## raceUnknown 3.197e+03  
## raceWhite 1.541e+02  
## ethnicitySpanish.Hispanic 2.984e+02  
## ethnicityUnknown 1.596e+03  
## length\_of\_stay 5.717e+03  
## type\_of\_admissionEmergency 4.938e+03  
## type\_of\_admissionNewborn 5.850e+03  
## type\_of\_admissionUrgent 3.143e+04  
## apr\_severity\_of\_illness\_code -1.800e+02  
## apr\_risk\_of\_mortalityMajor -3.026e+04  
## apr\_risk\_of\_mortalityMinor -5.904e+04  
## apr\_risk\_of\_mortalityModerate -6.198e+04  
## apr\_medical\_surgical\_descriptionSurgical 7.096e+04  
## payment\_typology\_1Federal.State.Local.VA -6.991e+02  
## payment\_typology\_1Managed.Care..Unspecified -1.078e+04  
## payment\_typology\_1Medicaid 6.376e+02  
## payment\_typology\_1Medicare -1.497e+03  
## payment\_typology\_1Private.Health.Insurance -1.245e+02  
## payment\_typology\_1Self.Pay 2.530e+03  
## birth\_weight 2.079e+00  
## emergency\_department\_indicatorY 6.417e+02  
## total\_costs 1.192e+00  
## ccscancer 4.014e+04  
## ccscirculatory\_sysytem\_disease 1.262e+04  
## ccscongenital\_anomalies 1.412e+04  
## ccsdigestive\_disease 1.787e+04  
## ccsendocrine\_metabolic\_disease 2.766e+04  
## ccsgenitourinary\_disease 3.006e+04  
## ccsinfectious\_disease 1.066e+04  
## ccsinjury\_poisoning 2.328e+04  
## ccsmental\_disorder -2.339e+04  
## ccsmusculoskeletal\_disease 5.193e+04  
## ccsnervous\_system\_disease 2.234e+04  
## ccsother 3.757e+04  
## ccsperinatal\_condition 3.470e+04  
## ccspregnancy\_childbirth\_complication 2.386e+04  
## ccsrespiratory\_disease 3.314e+04  
## ccsskin\_disease 3.678e+04  
## Std. Error  
## (Intercept) 1.643e+04  
## hospital\_countyManhattan 1.259e+03  
## hospital\_countyQueens 1.575e+03  
## hospital\_countyRichmond 1.074e+03  
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus 1.055e+03  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division 1.306e+03  
## facility\_nameBrookdale.Hospital.Medical.Center 1.313e+03  
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus 1.121e+03  
## facility\_nameConey.Island.Hospital 1.677e+03  
## facility\_nameElmhurst.Hospital.Center 1.418e+03  
## facility\_nameFlushing.Hospital.Medical.Center 1.434e+03  
## facility\_nameForest.Hills.Hospital 1.479e+03  
## facility\_nameHarlem.Hospital.Center 1.438e+03  
## facility\_nameJacobi.Medical.Center 1.103e+03  
## facility\_nameJamaica.Hospital.Medical.Center 1.433e+03  
## facility\_nameKings.County.Hospital.Center 1.128e+03  
## facility\_nameLenox.Hill.Hospital 1.052e+03  
## facility\_nameLincoln.Medical...Mental.Health.Center 1.120e+03  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 1.361e+03  
## facility\_nameLutheran.Medical.Center 1.028e+03  
## facility\_nameMaimonides.Medical.Center 9.667e+02  
## facility\_nameMetropolitan.Hospital.Center 1.285e+03  
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div 1.025e+03  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div 2.171e+03  
## facility\_nameMontefiore.Medical.Center...North.Division 1.169e+03  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital 1.487e+03  
## facility\_nameMount.Sinai.Beth.Israel 2.501e+03  
## facility\_nameMount.Sinai.Hospital 9.982e+02  
## facility\_nameMount.Sinai.Roosevelt 2.183e+03  
## facility\_nameMount.Sinai.West 9.975e+03  
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens 1.401e+03  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens 2.124e+03  
## facility\_nameNew.York.Methodist.Hospital 1.005e+03  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital 1.150e+03  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center 1.023e+03  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center 1.012e+03  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital 1.139e+03  
## facility\_nameNew.York.Presbyterian.Queens 2.812e+04  
## facility\_nameNewYork.Presbyterian.Queens 1.154e+04  
## facility\_nameNorth.Central.Bronx.Hospital 1.410e+03  
## facility\_nameNYU.Hospitals.Center 1.039e+03  
## facility\_nameQueens.Hospital.Center 1.504e+03  
## facility\_nameRichmond.University.Medical.Center 8.584e+02  
## facility\_nameSBH.Health.System 3.505e+03  
## facility\_nameSt.Barnabas.Hospital 1.439e+03  
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore NA  
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division 1.017e+03  
## facility\_nameStaten.Island.University.Hosp.North NA  
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH 1.588e+03  
## facility\_nameUniversity.Hospital.of.Brooklyn 1.233e+03  
## facility\_nameWoodhull.Medical...Mental.Health.Center 1.146e+03  
## facility\_nameWyckoff.Heights.Medical.Center NA  
## genderM 1.877e+02  
## raceOther.Race 2.845e+02  
## raceUnknown 8.501e+03  
## raceWhite 3.074e+02  
## ethnicitySpanish.Hispanic 2.789e+02  
## ethnicityUnknown 8.511e+02  
## length\_of\_stay 3.188e+01  
## type\_of\_admissionEmergency 2.317e+03  
## type\_of\_admissionNewborn 2.093e+03  
## type\_of\_admissionUrgent 2.456e+03  
## apr\_severity\_of\_illness\_code 1.736e+02  
## apr\_risk\_of\_mortalityMajor 1.867e+03  
## apr\_risk\_of\_mortalityMinor 1.685e+03  
## apr\_risk\_of\_mortalityModerate 1.761e+03  
## apr\_medical\_surgical\_descriptionSurgical 1.528e+03  
## payment\_typology\_1Federal.State.Local.VA 2.162e+03  
## payment\_typology\_1Managed.Care..Unspecified 3.745e+03  
## payment\_typology\_1Medicaid 3.233e+02  
## payment\_typology\_1Medicare 2.153e+03  
## payment\_typology\_1Private.Health.Insurance 3.221e+02  
## payment\_typology\_1Self.Pay 7.215e+02  
## birth\_weight 1.823e-01  
## emergency\_department\_indicatorY 1.208e+03  
## total\_costs 1.117e-02  
## ccscancer 1.905e+04  
## ccscirculatory\_sysytem\_disease 2.297e+04  
## ccscongenital\_anomalies 1.636e+04  
## ccsdigestive\_disease 1.675e+04  
## ccsendocrine\_metabolic\_disease 1.779e+04  
## ccsgenitourinary\_disease 1.749e+04  
## ccsinfectious\_disease 1.649e+04  
## ccsinjury\_poisoning 1.703e+04  
## ccsmental\_disorder 2.055e+04  
## ccsmusculoskeletal\_disease 3.248e+04  
## ccsnervous\_system\_disease 1.702e+04  
## ccsother 1.644e+04  
## ccsperinatal\_condition 1.626e+04  
## ccspregnancy\_childbirth\_complication 2.149e+04  
## ccsrespiratory\_disease 1.639e+04  
## ccsskin\_disease 1.781e+04  
## t value  
## (Intercept) -0.478  
## hospital\_countyManhattan -0.870  
## hospital\_countyQueens 2.458  
## hospital\_countyRichmond 2.481  
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus 13.700  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division -7.499  
## facility\_nameBrookdale.Hospital.Medical.Center 0.896  
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus 1.336  
## facility\_nameConey.Island.Hospital 3.403  
## facility\_nameElmhurst.Hospital.Center -0.235  
## facility\_nameFlushing.Hospital.Medical.Center 1.562  
## facility\_nameForest.Hills.Hospital 5.221  
## facility\_nameHarlem.Hospital.Center -1.769  
## facility\_nameJacobi.Medical.Center -1.475  
## facility\_nameJamaica.Hospital.Medical.Center -1.525  
## facility\_nameKings.County.Hospital.Center 0.741  
## facility\_nameLenox.Hill.Hospital 19.304  
## facility\_nameLincoln.Medical...Mental.Health.Center -0.521  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 14.114  
## facility\_nameLutheran.Medical.Center 11.376  
## facility\_nameMaimonides.Medical.Center 14.854  
## facility\_nameMetropolitan.Hospital.Center 1.443  
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div 22.477  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div 17.280  
## facility\_nameMontefiore.Medical.Center...North.Division 18.070  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital 13.465  
## facility\_nameMount.Sinai.Beth.Israel 5.609  
## facility\_nameMount.Sinai.Hospital 9.256  
## facility\_nameMount.Sinai.Roosevelt 2.212  
## facility\_nameMount.Sinai.West -1.563  
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens 6.483  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens 4.541  
## facility\_nameNew.York.Methodist.Hospital 9.578  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital 12.946  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center 23.281  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center 18.138  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital 9.828  
## facility\_nameNew.York.Presbyterian.Queens 0.341  
## facility\_nameNewYork.Presbyterian.Queens 0.613  
## facility\_nameNorth.Central.Bronx.Hospital 1.314  
## facility\_nameNYU.Hospitals.Center 14.236  
## facility\_nameQueens.Hospital.Center -1.169  
## facility\_nameRichmond.University.Medical.Center 6.857  
## facility\_nameSBH.Health.System -0.972  
## facility\_nameSt.Barnabas.Hospital -0.632  
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore NA  
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division 7.994  
## facility\_nameStaten.Island.University.Hosp.North NA  
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH -2.753  
## facility\_nameUniversity.Hospital.of.Brooklyn -8.996  
## facility\_nameWoodhull.Medical...Mental.Health.Center 0.156  
## facility\_nameWyckoff.Heights.Medical.Center NA  
## genderM -1.920  
## raceOther.Race 2.585  
## raceUnknown 0.376  
## raceWhite 0.501  
## ethnicitySpanish.Hispanic 1.070  
## ethnicityUnknown 1.875  
## length\_of\_stay 179.336  
## type\_of\_admissionEmergency 2.131  
## type\_of\_admissionNewborn 2.795  
## type\_of\_admissionUrgent 12.799  
## apr\_severity\_of\_illness\_code -1.037  
## apr\_risk\_of\_mortalityMajor -16.208  
## apr\_risk\_of\_mortalityMinor -35.036  
## apr\_risk\_of\_mortalityModerate -35.204  
## apr\_medical\_surgical\_descriptionSurgical 46.433  
## payment\_typology\_1Federal.State.Local.VA -0.323  
## payment\_typology\_1Managed.Care..Unspecified -2.879  
## payment\_typology\_1Medicaid 1.972  
## payment\_typology\_1Medicare -0.695  
## payment\_typology\_1Private.Health.Insurance -0.386  
## payment\_typology\_1Self.Pay 3.506  
## birth\_weight 11.405  
## emergency\_department\_indicatorY 0.531  
## total\_costs 106.676  
## ccscancer 2.107  
## ccscirculatory\_sysytem\_disease 0.549  
## ccscongenital\_anomalies 0.863  
## ccsdigestive\_disease 1.067  
## ccsendocrine\_metabolic\_disease 1.555  
## ccsgenitourinary\_disease 1.719  
## ccsinfectious\_disease 0.647  
## ccsinjury\_poisoning 1.368  
## ccsmental\_disorder -1.138  
## ccsmusculoskeletal\_disease 1.599  
## ccsnervous\_system\_disease 1.313  
## ccsother 2.285  
## ccsperinatal\_condition 2.134  
## ccspregnancy\_childbirth\_complication 1.111  
## ccsrespiratory\_disease 2.022  
## ccsskin\_disease 2.065  
## Pr(>|t|)  
## (Intercept) 0.632492  
## hospital\_countyManhattan 0.384211  
## hospital\_countyQueens 0.013983  
## hospital\_countyRichmond 0.013119  
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus < 2e-16  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division 6.49e-14  
## facility\_nameBrookdale.Hospital.Medical.Center 0.370195  
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus 0.181646  
## facility\_nameConey.Island.Hospital 0.000666  
## facility\_nameElmhurst.Hospital.Center 0.814054  
## facility\_nameFlushing.Hospital.Medical.Center 0.118235  
## facility\_nameForest.Hills.Hospital 1.78e-07  
## facility\_nameHarlem.Hospital.Center 0.076853  
## facility\_nameJacobi.Medical.Center 0.140213  
## facility\_nameJamaica.Hospital.Medical.Center 0.127350  
## facility\_nameKings.County.Hospital.Center 0.458735  
## facility\_nameLenox.Hill.Hospital < 2e-16  
## facility\_nameLincoln.Medical...Mental.Health.Center 0.602027  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division < 2e-16  
## facility\_nameLutheran.Medical.Center < 2e-16  
## facility\_nameMaimonides.Medical.Center < 2e-16  
## facility\_nameMetropolitan.Hospital.Center 0.148982  
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div < 2e-16  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div < 2e-16  
## facility\_nameMontefiore.Medical.Center...North.Division < 2e-16  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital < 2e-16  
## facility\_nameMount.Sinai.Beth.Israel 2.04e-08  
## facility\_nameMount.Sinai.Hospital < 2e-16  
## facility\_nameMount.Sinai.Roosevelt 0.026938  
## facility\_nameMount.Sinai.West 0.118057  
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens 9.02e-11  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens 5.60e-06  
## facility\_nameNew.York.Methodist.Hospital < 2e-16  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital < 2e-16  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center < 2e-16  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center < 2e-16  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital < 2e-16  
## facility\_nameNew.York.Presbyterian.Queens 0.732917  
## facility\_nameNewYork.Presbyterian.Queens 0.539760  
## facility\_nameNorth.Central.Bronx.Hospital 0.188796  
## facility\_nameNYU.Hospitals.Center < 2e-16  
## facility\_nameQueens.Hospital.Center 0.242399  
## facility\_nameRichmond.University.Medical.Center 7.09e-12  
## facility\_nameSBH.Health.System 0.331274  
## facility\_nameSt.Barnabas.Hospital 0.527379  
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore NA  
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division 1.33e-15  
## facility\_nameStaten.Island.University.Hosp.North NA  
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH 0.005909  
## facility\_nameUniversity.Hospital.of.Brooklyn < 2e-16  
## facility\_nameWoodhull.Medical...Mental.Health.Center 0.875880  
## facility\_nameWyckoff.Heights.Medical.Center NA  
## genderM 0.054875  
## raceOther.Race 0.009734  
## raceUnknown 0.706867  
## raceWhite 0.616094  
## ethnicitySpanish.Hispanic 0.284585  
## ethnicityUnknown 0.060789  
## length\_of\_stay < 2e-16  
## type\_of\_admissionEmergency 0.033067  
## type\_of\_admissionNewborn 0.005198  
## type\_of\_admissionUrgent < 2e-16  
## apr\_severity\_of\_illness\_code 0.299915  
## apr\_risk\_of\_mortalityMajor < 2e-16  
## apr\_risk\_of\_mortalityMinor < 2e-16  
## apr\_risk\_of\_mortalityModerate < 2e-16  
## apr\_medical\_surgical\_descriptionSurgical < 2e-16  
## payment\_typology\_1Federal.State.Local.VA 0.746475  
## payment\_typology\_1Managed.Care..Unspecified 0.003986  
## payment\_typology\_1Medicaid 0.048613  
## payment\_typology\_1Medicare 0.486763  
## payment\_typology\_1Private.Health.Insurance 0.699206  
## payment\_typology\_1Self.Pay 0.000455  
## birth\_weight < 2e-16  
## emergency\_department\_indicatorY 0.595184  
## total\_costs < 2e-16  
## ccscancer 0.035127  
## ccscirculatory\_sysytem\_disease 0.582876  
## ccscongenital\_anomalies 0.388113  
## ccsdigestive\_disease 0.285902  
## ccsendocrine\_metabolic\_disease 0.120001  
## ccsgenitourinary\_disease 0.085683  
## ccsinfectious\_disease 0.517909  
## ccsinjury\_poisoning 0.171438  
## ccsmental\_disorder 0.255210  
## ccsmusculoskeletal\_disease 0.109852  
## ccsnervous\_system\_disease 0.189247  
## ccsother 0.022342  
## ccsperinatal\_condition 0.032846  
## ccspregnancy\_childbirth\_complication 0.266780  
## ccsrespiratory\_disease 0.043211  
## ccsskin\_disease 0.038921  
##   
## (Intercept)   
## hospital\_countyManhattan   
## hospital\_countyQueens \*   
## hospital\_countyRichmond \*   
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus \*\*\*  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division \*\*\*  
## facility\_nameBrookdale.Hospital.Medical.Center   
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus   
## facility\_nameConey.Island.Hospital \*\*\*  
## facility\_nameElmhurst.Hospital.Center   
## facility\_nameFlushing.Hospital.Medical.Center   
## facility\_nameForest.Hills.Hospital \*\*\*  
## facility\_nameHarlem.Hospital.Center .   
## facility\_nameJacobi.Medical.Center   
## facility\_nameJamaica.Hospital.Medical.Center   
## facility\_nameKings.County.Hospital.Center   
## facility\_nameLenox.Hill.Hospital \*\*\*  
## facility\_nameLincoln.Medical...Mental.Health.Center   
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division \*\*\*  
## facility\_nameLutheran.Medical.Center \*\*\*  
## facility\_nameMaimonides.Medical.Center \*\*\*  
## facility\_nameMetropolitan.Hospital.Center   
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div \*\*\*  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div \*\*\*  
## facility\_nameMontefiore.Medical.Center...North.Division \*\*\*  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital \*\*\*  
## facility\_nameMount.Sinai.Beth.Israel \*\*\*  
## facility\_nameMount.Sinai.Hospital \*\*\*  
## facility\_nameMount.Sinai.Roosevelt \*   
## facility\_nameMount.Sinai.West   
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens \*\*\*  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens \*\*\*  
## facility\_nameNew.York.Methodist.Hospital \*\*\*  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital \*\*\*  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center \*\*\*  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center \*\*\*  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital \*\*\*  
## facility\_nameNew.York.Presbyterian.Queens   
## facility\_nameNewYork.Presbyterian.Queens   
## facility\_nameNorth.Central.Bronx.Hospital   
## facility\_nameNYU.Hospitals.Center \*\*\*  
## facility\_nameQueens.Hospital.Center   
## facility\_nameRichmond.University.Medical.Center \*\*\*  
## facility\_nameSBH.Health.System   
## facility\_nameSt.Barnabas.Hospital   
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore   
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division \*\*\*  
## facility\_nameStaten.Island.University.Hosp.North   
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH \*\*   
## facility\_nameUniversity.Hospital.of.Brooklyn \*\*\*  
## facility\_nameWoodhull.Medical...Mental.Health.Center   
## facility\_nameWyckoff.Heights.Medical.Center   
## genderM .   
## raceOther.Race \*\*   
## raceUnknown   
## raceWhite   
## ethnicitySpanish.Hispanic   
## ethnicityUnknown .   
## length\_of\_stay \*\*\*  
## type\_of\_admissionEmergency \*   
## type\_of\_admissionNewborn \*\*   
## type\_of\_admissionUrgent \*\*\*  
## apr\_severity\_of\_illness\_code   
## apr\_risk\_of\_mortalityMajor \*\*\*  
## apr\_risk\_of\_mortalityMinor \*\*\*  
## apr\_risk\_of\_mortalityModerate \*\*\*  
## apr\_medical\_surgical\_descriptionSurgical \*\*\*  
## payment\_typology\_1Federal.State.Local.VA   
## payment\_typology\_1Managed.Care..Unspecified \*\*   
## payment\_typology\_1Medicaid \*   
## payment\_typology\_1Medicare   
## payment\_typology\_1Private.Health.Insurance   
## payment\_typology\_1Self.Pay \*\*\*  
## birth\_weight \*\*\*  
## emergency\_department\_indicatorY   
## total\_costs \*\*\*  
## ccscancer \*   
## ccscirculatory\_sysytem\_disease   
## ccscongenital\_anomalies   
## ccsdigestive\_disease   
## ccsendocrine\_metabolic\_disease   
## ccsgenitourinary\_disease .   
## ccsinfectious\_disease   
## ccsinjury\_poisoning   
## ccsmental\_disorder   
## ccsmusculoskeletal\_disease   
## ccsnervous\_system\_disease   
## ccsother \*   
## ccsperinatal\_condition \*   
## ccspregnancy\_childbirth\_complication   
## ccsrespiratory\_disease \*   
## ccsskin\_disease \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 28080 on 90947 degrees of freedom  
## Multiple R-squared: 0.8573, Adjusted R-squared: 0.8572   
## F-statistic: 6210 on 88 and 90947 DF, p-value: < 2.2e-16

#test error   
pred\_lm <- predict(fit\_lm, discharge\_data2[-trRows, -1])

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient  
## fit may be misleading

mse(y2, pred\_lm)

## [1] 937604771

rmse(y2, pred\_lm)

## [1] 30620.33

We found the test erorr (MSE) to be 937604771 with a lambda value of 28.03162.

# RIDGE REGRESSION

Fiting a ridge regression model on the training data, with λ chosen by cross-validation. Report the test error.

set.seed(100)  
ridge.fit <- train(x, y,  
 method = "glmnet",  
 tuneGrid = expand.grid(alpha = 0,   
 lambda = exp(seq(-50, 50, length=100))),   
 preProc = c("center", "scale"),  
 trControl = ctrl1)

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: ccsmusculoskeletal\_disease

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut  
## = 19, uniqueCut = 10, : These variables have zero variances:  
## facility\_nameNew.York.Presbyterian.Queens

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

ridge.fit

## glmnet   
##   
## 91036 samples  
## 92 predictor  
##   
## Pre-processing: centered (92), scaled (92)   
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 81932, 81933, 81933, 81933, 81932, 81933, ...   
## Resampling results across tuning parameters:  
##   
## lambda RMSE Rsquared MAE   
## 1.928750e-22 28517.52 0.8548152 8931.058  
## 5.296112e-22 28517.52 0.8548152 8931.058  
## 1.454248e-21 28517.52 0.8548152 8931.058  
## 3.993188e-21 28517.52 0.8548152 8931.058  
## 1.096481e-20 28517.52 0.8548152 8931.058  
## 3.010803e-20 28517.52 0.8548152 8931.058  
## 8.267300e-20 28517.52 0.8548152 8931.058  
## 2.270100e-19 28517.52 0.8548152 8931.058  
## 6.233418e-19 28517.52 0.8548152 8931.058  
## 1.711621e-18 28517.52 0.8548152 8931.058  
## 4.699903e-18 28517.52 0.8548152 8931.058  
## 1.290536e-17 28517.52 0.8548152 8931.058  
## 3.543655e-17 28517.52 0.8548152 8931.058  
## 9.730446e-17 28517.52 0.8548152 8931.058  
## 2.671862e-16 28517.52 0.8548152 8931.058  
## 7.336608e-16 28517.52 0.8548152 8931.058  
## 2.014543e-15 28517.52 0.8548152 8931.058  
## 5.531691e-15 28517.52 0.8548152 8931.058  
## 1.518935e-14 28517.52 0.8548152 8931.058  
## 4.170811e-14 28517.52 0.8548152 8931.058  
## 1.145254e-13 28517.52 0.8548152 8931.058  
## 3.144728e-13 28517.52 0.8548152 8931.058  
## 8.635041e-13 28517.52 0.8548152 8931.058  
## 2.371077e-12 28517.52 0.8548152 8931.058  
## 6.510689e-12 28517.52 0.8548152 8931.058  
## 1.787756e-11 28517.52 0.8548152 8931.058  
## 4.908961e-11 28517.52 0.8548152 8931.058  
## 1.347941e-10 28517.52 0.8548152 8931.058  
## 3.701282e-10 28517.52 0.8548152 8931.058  
## 1.016327e-09 28517.52 0.8548152 8931.058  
## 2.790710e-09 28517.52 0.8548152 8931.058  
## 7.662951e-09 28517.52 0.8548152 8931.058  
## 2.104153e-08 28517.52 0.8548152 8931.058  
## 5.777749e-08 28517.52 0.8548152 8931.058  
## 1.586499e-07 28517.52 0.8548152 8931.058  
## 4.356335e-07 28517.52 0.8548152 8931.058  
## 1.196196e-06 28517.52 0.8548152 8931.058  
## 3.284610e-06 28517.52 0.8548152 8931.058  
## 9.019140e-06 28517.52 0.8548152 8931.058  
## 2.476546e-05 28517.52 0.8548152 8931.058  
## 6.800294e-05 28517.52 0.8548152 8931.058  
## 1.867278e-04 28517.52 0.8548152 8931.058  
## 5.127318e-04 28517.52 0.8548152 8931.058  
## 1.407899e-03 28517.52 0.8548152 8931.058  
## 3.865920e-03 28517.52 0.8548152 8931.058  
## 1.061535e-02 28517.52 0.8548152 8931.058  
## 2.914845e-02 28517.52 0.8548152 8931.058  
## 8.003810e-02 28517.52 0.8548152 8931.058  
## 2.197749e-01 28517.52 0.8548152 8931.058  
## 6.034751e-01 28517.52 0.8548152 8931.058  
## 1.657069e+00 28517.52 0.8548152 8931.058  
## 4.550110e+00 28517.52 0.8548152 8931.058  
## 1.249405e+01 28517.52 0.8548152 8931.058  
## 3.430714e+01 28517.52 0.8548152 8931.058  
## 9.420324e+01 28517.52 0.8548152 8931.058  
## 2.586706e+02 28517.52 0.8548152 8931.058  
## 7.102781e+02 28517.52 0.8548152 8931.058  
## 1.950337e+03 28517.52 0.8548152 8931.058  
## 5.355389e+03 28517.52 0.8548152 8931.058  
## 1.470525e+04 29261.90 0.8510164 8774.302  
## 4.037882e+04 32270.89 0.8368981 9083.627  
## 1.108753e+05 39473.04 0.8066683 10351.515  
## 3.044501e+05 50558.71 0.7706085 13473.051  
## 8.359831e+05 61540.03 0.7442838 17638.888  
## 2.295508e+06 68605.90 0.7303976 20520.063  
## 6.303185e+06 71961.77 0.7243796 21935.011  
## 1.730778e+07 73324.29 0.7220232 22518.938  
## 4.752506e+07 73841.55 0.7211408 22742.225  
## 1.304980e+08 74142.98 NaN 22872.730  
## 3.583317e+08 74142.98 NaN 22872.730  
## 9.839353e+08 74142.98 NaN 22872.730  
## 2.701767e+09 74142.98 NaN 22872.730  
## 7.418723e+09 74142.98 NaN 22872.730  
## 2.037091e+10 74142.98 NaN 22872.730  
## 5.593604e+10 74142.98 NaN 22872.730  
## 1.535936e+11 74142.98 NaN 22872.730  
## 4.217492e+11 74142.98 NaN 22872.730  
## 1.158072e+12 74142.98 NaN 22872.730  
## 3.179925e+12 74142.98 NaN 22872.730  
## 8.731688e+12 74142.98 NaN 22872.730  
## 2.397615e+13 74142.98 NaN 22872.730  
## 6.583560e+13 74142.98 NaN 22872.730  
## 1.807765e+14 74142.98 NaN 22872.730  
## 4.963904e+14 74142.98 NaN 22872.730  
## 1.363028e+15 74142.98 NaN 22872.730  
## 3.742708e+15 74142.98 NaN 22872.730  
## 1.027702e+16 74142.98 NaN 22872.730  
## 2.821945e+16 74142.98 NaN 22872.730  
## 7.748718e+16 74142.98 NaN 22872.730  
## 2.127704e+17 74142.98 NaN 22872.730  
## 5.842416e+17 74142.98 NaN 22872.730  
## 1.604256e+18 74142.98 NaN 22872.730  
## 4.405092e+18 74142.98 NaN 22872.730  
## 1.209585e+19 74142.98 NaN 22872.730  
## 3.321373e+19 74142.98 NaN 22872.730  
## 9.120086e+19 74142.98 NaN 22872.730  
## 2.504265e+20 74142.98 NaN 22872.730  
## 6.876406e+20 74142.98 NaN 22872.730  
## 1.888177e+21 74142.98 NaN 22872.730  
## 5.184706e+21 74142.98 NaN 22872.730  
##   
## Tuning parameter 'alpha' was held constant at a value of 0  
## RMSE was used to select the optimal model using the smallest value.  
## The final values used for the model were alpha = 0 and lambda = 5355.389.

#lambda value   
ridge.fit$bestTune   
  
#model coefficients   
coef(ridge.fit$finalModel,ridge.fit$bestTune$lambda)

## 93 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 19730.125339  
## hospital\_countyKings -401.124965  
## hospital\_countyManhattan 14.316032  
## hospital\_countyQueens 581.278826  
## hospital\_countyRichmond -372.020868  
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus 733.725487  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division -2080.176381  
## facility\_nameBrookdale.Hospital.Medical.Center -654.546375  
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus -735.433220  
## facility\_nameConey.Island.Hospital -140.546209  
## facility\_nameElmhurst.Hospital.Center -996.484138  
## facility\_nameFlushing.Hospital.Medical.Center -545.512933  
## facility\_nameForest.Hills.Hospital 143.690693  
## facility\_nameHarlem.Hospital.Center -948.864755  
## facility\_nameJacobi.Medical.Center -1362.565113  
## facility\_nameJamaica.Hospital.Medical.Center -984.549466  
## facility\_nameKings.County.Hospital.Center -867.973533  
## facility\_nameLenox.Hill.Hospital 1639.869532  
## facility\_nameLincoln.Medical...Mental.Health.Center -1059.966460  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 2754.386266  
## facility\_nameLutheran.Medical.Center 589.615420  
## facility\_nameMaimonides.Medical.Center 1579.636701  
## facility\_nameMetropolitan.Hospital.Center -729.943336  
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div 2468.095667  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div 1376.594982  
## facility\_nameMontefiore.Medical.Center...North.Division 1404.718677  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital 800.201974  
## facility\_nameMount.Sinai.Beth.Israel 154.559294  
## facility\_nameMount.Sinai.Hospital -344.117085  
## facility\_nameMount.Sinai.Roosevelt -250.730767  
## facility\_nameMount.Sinai.West -193.883334  
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens 460.632957  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens 152.760090  
## facility\_nameNew.York.Methodist.Hospital 272.018795  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital 665.075833  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center 2510.947143  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center 1586.670338  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital 150.320525  
## facility\_nameNew.York.Presbyterian.Queens 9.411302  
## facility\_nameNewYork.Presbyterian.Queens 56.917918  
## facility\_nameNorth.Central.Bronx.Hospital -559.019941  
## facility\_nameNYU.Hospitals.Center 630.786746  
## facility\_nameQueens.Hospital.Center -885.154739  
## facility\_nameRichmond.University.Medical.Center 191.577118  
## facility\_nameSBH.Health.System -235.823892  
## facility\_nameSt.Barnabas.Hospital -639.418840  
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore -366.435397  
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division -545.254261  
## facility\_nameStaten.Island.University.Hosp.North -662.537196  
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH -797.180318  
## facility\_nameUniversity.Hospital.of.Brooklyn -1982.245196  
## facility\_nameWoodhull.Medical...Mental.Health.Center -870.397360  
## facility\_nameWyckoff.Heights.Medical.Center -632.406044  
## genderM -1.718848  
## raceOther.Race 329.219469  
## raceUnknown 20.114862  
## raceWhite 223.887375  
## ethnicitySpanish.Hispanic 30.870556  
## ethnicityUnknown -99.484525  
## length\_of\_stay 34912.197953  
## type\_of\_admissionEmergency -253.098642  
## type\_of\_admissionNewborn 231.600854  
## type\_of\_admissionUrgent 1813.481118  
## apr\_severity\_of\_illness\_code 1073.877856  
## apr\_risk\_of\_mortalityMajor 1027.961143  
## apr\_risk\_of\_mortalityMinor -5236.365124  
## apr\_risk\_of\_mortalityModerate -3152.331587  
## apr\_medical\_surgical\_descriptionSurgical 5343.062880  
## payment\_typology\_1Federal.State.Local.VA -49.638935  
## payment\_typology\_1Managed.Care..Unspecified -226.020868  
## payment\_typology\_1Medicaid -48.988506  
## payment\_typology\_1Medicare -73.974618  
## payment\_typology\_1Private.Health.Insurance -73.161229  
## payment\_typology\_1Self.Pay 62.420121  
## birth\_weight -509.087805  
## emergency\_department\_indicatorY 83.994292  
## total\_costs 27082.687743  
## ccscancer 76.265313  
## ccscirculatory\_sysytem\_disease -107.139854  
## ccscongenital\_anomalies -724.251919  
## ccsdigestive\_disease -228.653419  
## ccsendocrine\_metabolic\_disease -17.115423  
## ccsgenitourinary\_disease 6.234718  
## ccsinfectious\_disease -407.916662  
## ccsinjury\_poisoning -124.041390  
## ccsmental\_disorder -277.830011  
## ccsmusculoskeletal\_disease 68.442824  
## ccsnervous\_system\_disease -110.213547  
## ccsother 286.026870  
## ccsperinatal\_condition 583.960114  
## ccspregnancy\_childbirth\_complication -12.658190  
## ccsrespiratory\_disease 182.569953  
## ccsskin\_disease 107.487776

#test error   
pred\_ridge <- predict(ridge.fit, x)  
mse(y2, pred\_ridge)

## [1] 9790669585

rmse(y2, pred\_ridge)

## [1] 98947.81

We found the test erorr (MSE) to be 9790669585 with a lambda value of 5355.389.

# LASSO MODEL

Fiting a lasso model on the training data, with λ chosen by cross-validation. Report the test error, along with the number of non-zero coefficient estimates.

set.seed(100)  
lasso.fit <- train(x, y,  
 method = "glmnet",  
 tuneGrid = expand.grid(alpha = 1,   
 lambda = exp(seq(-10, 10, length=100))),  
 preProc = c("center", "scale"),  
 trControl = ctrl1)

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: ccsmusculoskeletal\_disease

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut  
## = 19, uniqueCut = 10, : These variables have zero variances:  
## facility\_nameNew.York.Presbyterian.Queens

lasso.fit

## glmnet   
##   
## 91036 samples  
## 92 predictor  
##   
## Pre-processing: centered (92), scaled (92)   
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 81932, 81933, 81933, 81933, 81932, 81933, ...   
## Resampling results across tuning parameters:  
##   
## lambda RMSE Rsquared MAE   
## 4.539993e-05 28180.84 0.8568294 9399.915  
## 5.556374e-05 28180.84 0.8568294 9399.915  
## 6.800294e-05 28180.84 0.8568294 9399.915  
## 8.322695e-05 28180.84 0.8568294 9399.915  
## 1.018592e-04 28180.84 0.8568294 9399.915  
## 1.246627e-04 28180.84 0.8568294 9399.915  
## 1.525713e-04 28180.84 0.8568294 9399.915  
## 1.867278e-04 28180.84 0.8568294 9399.915  
## 2.285311e-04 28180.84 0.8568294 9399.915  
## 2.796929e-04 28180.84 0.8568294 9399.915  
## 3.423086e-04 28180.84 0.8568294 9399.915  
## 4.189421e-04 28180.84 0.8568294 9399.915  
## 5.127318e-04 28180.84 0.8568294 9399.915  
## 6.275185e-04 28180.84 0.8568294 9399.915  
## 7.680028e-04 28180.84 0.8568294 9399.915  
## 9.399377e-04 28180.84 0.8568294 9399.915  
## 1.150364e-03 28180.84 0.8568294 9399.915  
## 1.407899e-03 28180.84 0.8568294 9399.915  
## 1.723090e-03 28180.84 0.8568294 9399.915  
## 2.108842e-03 28180.84 0.8568294 9399.915  
## 2.580955e-03 28180.84 0.8568294 9399.915  
## 3.158760e-03 28180.84 0.8568294 9399.915  
## 3.865920e-03 28180.84 0.8568294 9399.915  
## 4.731394e-03 28180.84 0.8568294 9399.915  
## 5.790624e-03 28180.84 0.8568294 9399.915  
## 7.086987e-03 28180.84 0.8568294 9399.915  
## 8.673571e-03 28180.84 0.8568294 9399.915  
## 1.061535e-02 28180.84 0.8568294 9399.915  
## 1.299183e-02 28180.84 0.8568294 9399.915  
## 1.590035e-02 28180.84 0.8568294 9399.915  
## 1.946001e-02 28180.84 0.8568294 9399.915  
## 2.381657e-02 28180.84 0.8568294 9399.915  
## 2.914845e-02 28180.84 0.8568294 9399.915  
## 3.567399e-02 28180.84 0.8568294 9399.915  
## 4.366043e-02 28180.84 0.8568294 9399.915  
## 5.343481e-02 28180.84 0.8568294 9399.915  
## 6.539740e-02 28180.84 0.8568294 9399.915  
## 8.003810e-02 28180.84 0.8568294 9399.915  
## 9.795645e-02 28180.84 0.8568294 9399.915  
## 1.198862e-01 28180.84 0.8568294 9399.915  
## 1.467255e-01 28180.84 0.8568294 9399.915  
## 1.795733e-01 28180.84 0.8568294 9399.915  
## 2.197749e-01 28180.84 0.8568294 9399.915  
## 2.689765e-01 28180.84 0.8568294 9399.915  
## 3.291930e-01 28180.84 0.8568294 9399.915  
## 4.028903e-01 28180.84 0.8568294 9399.915  
## 4.930865e-01 28180.84 0.8568294 9399.915  
## 6.034751e-01 28180.84 0.8568294 9399.915  
## 7.385767e-01 28180.84 0.8568294 9399.915  
## 9.039239e-01 28180.84 0.8568294 9399.915  
## 1.106288e+00 28180.84 0.8568294 9399.915  
## 1.353955e+00 28180.84 0.8568294 9399.915  
## 1.657069e+00 28180.84 0.8568294 9399.915  
## 2.028042e+00 28180.84 0.8568294 9399.915  
## 2.482065e+00 28180.84 0.8568294 9399.915  
## 3.037732e+00 28180.84 0.8568294 9399.915  
## 3.717797e+00 28180.84 0.8568294 9399.915  
## 4.550110e+00 28180.84 0.8568294 9399.915  
## 5.568756e+00 28180.84 0.8568294 9399.915  
## 6.815449e+00 28180.84 0.8568294 9399.915  
## 8.341242e+00 28180.84 0.8568294 9399.915  
## 1.020862e+01 28180.84 0.8568294 9399.915  
## 1.249405e+01 28180.84 0.8568294 9399.915  
## 1.529113e+01 28180.84 0.8568294 9399.915  
## 1.871439e+01 28180.84 0.8568294 9399.915  
## 2.290404e+01 28180.84 0.8568294 9399.915  
## 2.803162e+01 28180.84 0.8568294 9399.915  
## 3.430714e+01 28180.88 0.8568286 9396.599  
## 4.198757e+01 28182.54 0.8568143 9377.129  
## 5.138745e+01 28184.20 0.8568005 9351.376  
## 6.289170e+01 28186.43 0.8567813 9319.573  
## 7.697143e+01 28189.93 0.8567500 9280.133  
## 9.420324e+01 28195.58 0.8566978 9233.274  
## 1.152928e+02 28204.21 0.8566167 9178.637  
## 1.411037e+02 28216.10 0.8565049 9115.597  
## 1.726929e+02 28233.23 0.8563418 9041.879  
## 2.113542e+02 28255.57 0.8561268 8958.110  
## 2.586706e+02 28277.52 0.8559215 8863.992  
## 3.165799e+02 28307.90 0.8556366 8759.435  
## 3.874535e+02 28348.81 0.8552522 8635.987  
## 4.741938e+02 28407.55 0.8546913 8507.700  
## 5.803529e+02 28499.87 0.8537979 8384.984  
## 7.102781e+02 28627.82 0.8525438 8271.624  
## 8.692900e+02 28780.48 0.8510418 8163.239  
## 1.063900e+03 28918.49 0.8496954 8037.850  
## 1.302079e+03 29065.20 0.8482861 7943.197  
## 1.593578e+03 29239.59 0.8466256 7899.393  
## 1.950337e+03 29408.53 0.8451000 7824.036  
## 2.386965e+03 29623.12 0.8431770 7869.759  
## 2.921341e+03 29814.85 0.8416657 7999.469  
## 3.575349e+03 29994.54 0.8405469 8095.567  
## 4.375773e+03 30186.35 0.8396859 8112.916  
## 5.355389e+03 30442.71 0.8386834 8120.993  
## 6.554314e+03 30815.57 0.8371866 8136.782  
## 8.021647e+03 31353.47 0.8349369 8205.437  
## 9.817475e+03 31919.18 0.8343169 8295.444  
## 1.201534e+04 32688.93 0.8342619 8454.966  
## 1.470525e+04 33811.25 0.8341770 8722.504  
## 1.799735e+04 35428.69 0.8340378 9242.582  
## 2.202647e+04 37725.97 0.8337939 10041.835  
##   
## 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 = 28.03162.

#lamda value   
lasso.fit$bestTune  
  
#model coefficients   
coef(lasso.fit$finalModel,lasso.fit$bestTune$lambda)

## 93 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 19730.125339  
## hospital\_countyKings .   
## hospital\_countyManhattan .   
## hospital\_countyQueens .   
## hospital\_countyRichmond .   
## facility\_nameBeth.Israel.Medical.Center.Petrie.Campus 1253.296728  
## facility\_nameBronx.Lebanon.Hospital.Center...Concourse.Division -1927.048841  
## facility\_nameBrookdale.Hospital.Medical.Center -410.688177  
## facility\_nameBrooklyn.Hospital.Center...Downtown.Campus -539.361660  
## facility\_nameConey.Island.Hospital .   
## facility\_nameElmhurst.Hospital.Center -297.762345  
## facility\_nameFlushing.Hospital.Medical.Center 47.869940  
## facility\_nameForest.Hills.Hospital 734.885566  
## facility\_nameHarlem.Hospital.Center -740.452319  
## facility\_nameJacobi.Medical.Center -983.014566  
## facility\_nameJamaica.Hospital.Medical.Center -487.323881  
## facility\_nameKings.County.Hospital.Center -642.017675  
## facility\_nameLenox.Hill.Hospital 2347.193864  
## facility\_nameLincoln.Medical...Mental.Health.Center -774.653451  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 3959.737628  
## facility\_nameLutheran.Medical.Center 1040.284929  
## facility\_nameMaimonides.Medical.Center 2204.539822  
## facility\_nameMetropolitan.Hospital.Center -439.919462  
## facility\_nameMontefiore.Med.Center...Jack.D.Weiler.Hosp.of.A.Einstein.College.Div 3087.361968  
## facility\_nameMontefiore.Medical.Center...Henry...Lucy.Moses.Div 1533.327873  
## facility\_nameMontefiore.Medical.Center...North.Division 1838.022457  
## facility\_nameMontefiore.Medical.Center.Wakefield.Hospital 1082.467888  
## facility\_nameMount.Sinai.Beth.Israel 250.822670  
## facility\_nameMount.Sinai.Hospital 458.649521  
## facility\_nameMount.Sinai.Roosevelt -66.561253  
## facility\_nameMount.Sinai.West -175.963487  
## facility\_nameNew.York.Hospital.Medical.Center...of.Queens 1236.791038  
## facility\_nameNew.York.Hospital.Medical.Center.of.Queens 399.744436  
## facility\_nameNew.York.Methodist.Hospital 743.899247  
## facility\_nameNew.York.Presbyterian.Hospital...Allen.Hospital 1032.453456  
## facility\_nameNew.York.Presbyterian.Hospital...Columbia.Presbyterian.Center 3256.294508  
## facility\_nameNew.York.Presbyterian.Hospital...New.York.Weill.Cornell.Center 2372.593259  
## facility\_nameNew.York.Presbyterian.Lower.Manhattan.Hospital 546.092537  
## facility\_nameNew.York.Presbyterian.Queens .   
## facility\_nameNewYork.Presbyterian.Queens 8.151311  
## facility\_nameNorth.Central.Bronx.Hospital -279.255296  
## facility\_nameNYU.Hospitals.Center 1454.699943  
## facility\_nameQueens.Hospital.Center -389.396505  
## facility\_nameRichmond.University.Medical.Center 371.985616  
## facility\_nameSBH.Health.System -191.129113  
## facility\_nameSt.Barnabas.Hospital -502.407235  
## facility\_nameSt.Johns.Episcopal.Hospital.So.Shore -57.691242  
## facility\_nameSt.Lukes.Roosevelt.Hospital.Center...Roosevelt.Hospital.Division 187.495235  
## facility\_nameStaten.Island.University.Hosp.North -450.805985  
## facility\_nameSUNY.Downstate.Medical.Center.at.LICH -688.120908  
## facility\_nameUniversity.Hospital.of.Brooklyn -1868.694266  
## facility\_nameWoodhull.Medical...Mental.Health.Center -658.778545  
## facility\_nameWyckoff.Heights.Medical.Center -537.034992  
## genderM -134.088218  
## raceOther.Race 294.472697  
## raceUnknown .   
## raceWhite 61.921620  
## ethnicitySpanish.Hispanic 6.477588  
## ethnicityUnknown 73.245662  
## length\_of\_stay 41874.636145  
## type\_of\_admissionEmergency .   
## type\_of\_admissionNewborn 179.913215  
## type\_of\_admissionUrgent 1911.021823  
## apr\_severity\_of\_illness\_code -3.692325  
## apr\_risk\_of\_mortalityMajor -2373.660323  
## apr\_risk\_of\_mortalityMinor -9126.047204  
## apr\_risk\_of\_mortalityModerate -7322.214494  
## apr\_medical\_surgical\_descriptionSurgical 5348.276370  
## payment\_typology\_1Federal.State.Local.VA -7.311927  
## payment\_typology\_1Managed.Care..Unspecified -233.257406  
## payment\_typology\_1Medicaid 105.428357  
## payment\_typology\_1Medicare -33.747856  
## payment\_typology\_1Private.Health.Insurance -31.172982  
## payment\_typology\_1Self.Pay 184.192214  
## birth\_weight 1123.571828  
## emergency\_department\_indicatorY 52.423955  
## total\_costs 24814.138366  
## ccscancer 29.995993  
## ccscirculatory\_sysytem\_disease -62.229681  
## ccscongenital\_anomalies -909.892358  
## ccsdigestive\_disease -264.471243  
## ccsendocrine\_metabolic\_disease -7.301272  
## ccsgenitourinary\_disease .   
## ccsinfectious\_disease -604.070166  
## ccsinjury\_poisoning -98.040208  
## ccsmental\_disorder -361.338041  
## ccsmusculoskeletal\_disease 34.939721  
## ccsnervous\_system\_disease -119.708990  
## ccsother 218.214467  
## ccsperinatal\_condition 406.451419  
## ccspregnancy\_childbirth\_complication -9.067202  
## ccsrespiratory\_disease 90.515491  
## ccsskin\_disease 48.524308

#test error   
pred\_lasso <- predict(lasso.fit, newdata = x2)  
mse(y2, pred\_lasso)

## [1] 937116653

rmse(y2, pred\_lasso)

## [1] 30612.36

We found the test erorr (MSE) to be 937116653 with a lambda value of 28.03162.

# PCR MODEL

Fitting a principle component regression model on the training data, with M chosen by cross-validation.

set.seed(100)  
pcr.fit <-train(x, y,  
 method = "pcr",  
 tuneGrid =data.frame(ncomp = 1:92),   
 trControl = ctrl1,  
 preProc =c("center", "scale"))

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: ccsmusculoskeletal\_disease

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut  
## = 19, uniqueCut = 10, : These variables have zero variances:  
## facility\_nameNew.York.Presbyterian.Queens

pcr.fit

## Principal Component Analysis   
##   
## 91036 samples  
## 92 predictor  
##   
## Pre-processing: centered (92), scaled (92)   
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 81932, 81933, 81933, 81933, 81932, 81933, ...   
## Resampling results across tuning parameters:  
##   
## ncomp RMSE Rsquared MAE   
## 1 6.095158e+04 0.3260462 1.718126e+04  
## 2 5.431094e+04 0.4647305 2.556397e+04  
## 3 4.427509e+04 0.6447749 1.650835e+04  
## 4 4.422823e+04 0.6455109 1.655123e+04  
## 5 4.421291e+04 0.6457462 1.640458e+04  
## 6 4.415976e+04 0.6465798 1.619183e+04  
## 7 4.413171e+04 0.6470296 1.616590e+04  
## 8 4.330294e+04 0.6604009 1.660777e+04  
## 9 4.309022e+04 0.6637221 1.672886e+04  
## 10 4.305787e+04 0.6642161 1.685328e+04  
## 11 4.293798e+04 0.6660658 1.748823e+04  
## 12 4.271250e+04 0.6695164 1.800796e+04  
## 13 3.975373e+04 0.7132381 1.806576e+04  
## 14 3.970237e+04 0.7139577 1.800994e+04  
## 15 3.916038e+04 0.7215704 1.716327e+04  
## 16 3.915452e+04 0.7216549 1.716026e+04  
## 17 3.915103e+04 0.7217101 1.711614e+04  
## 18 3.915333e+04 0.7216863 1.708905e+04  
## 19 3.915740e+04 0.7216364 1.708915e+04  
## 20 3.906941e+04 0.7228829 1.699409e+04  
## 21 3.898997e+04 0.7240505 1.674684e+04  
## 22 3.898536e+04 0.7241114 1.670462e+04  
## 23 3.897689e+04 0.7242399 1.663296e+04  
## 24 3.894187e+04 0.7247529 1.651064e+04  
## 25 3.889678e+04 0.7253547 1.649983e+04  
## 26 3.888213e+04 0.7255244 1.646569e+04  
## 27 3.882963e+04 0.7262787 1.639411e+04  
## 28 3.878487e+04 0.7269397 1.635929e+04  
## 29 3.877435e+04 0.7270884 1.633722e+04  
## 30 3.871588e+04 0.7278979 1.627342e+04  
## 31 3.870614e+04 0.7280348 1.625892e+04  
## 32 3.869395e+04 0.7282030 1.625618e+04  
## 33 3.867536e+04 0.7284678 1.625494e+04  
## 34 3.866840e+04 0.7285718 1.625347e+04  
## 35 3.866257e+04 0.7286577 1.625147e+04  
## 36 3.865402e+04 0.7287767 1.626248e+04  
## 37 3.863901e+04 0.7289944 1.627351e+04  
## 38 3.863495e+04 0.7290487 1.626760e+04  
## 39 3.863292e+04 0.7291179 1.626990e+04  
## 40 3.860390e+04 0.7295206 1.629274e+04  
## 41 3.857182e+04 0.7299785 1.631054e+04  
## 42 3.856385e+04 0.7300995 1.632346e+04  
## 43 3.853652e+04 0.7304639 1.632239e+04  
## 44 3.853217e+04 0.7305267 1.632789e+04  
## 45 3.853272e+04 0.7305188 1.633049e+04  
## 46 3.852815e+04 0.7305847 1.632453e+04  
## 47 3.852849e+04 0.7305853 1.631559e+04  
## 48 3.852420e+04 0.7306540 1.630240e+04  
## 49 3.850975e+04 0.7308541 1.628237e+04  
## 50 3.850098e+04 0.7309789 1.626616e+04  
## 51 3.850911e+04 0.7308661 1.624474e+04  
## 52 3.850913e+04 0.7308654 1.624486e+04  
## 53 3.851270e+04 0.7308132 1.624389e+04  
## 54 3.851134e+04 0.7308374 1.624317e+04  
## 55 3.848917e+04 0.7311835 1.623221e+04  
## 56 3.846841e+04 0.7314780 1.623198e+04  
## 57 3.845664e+04 0.7316310 1.623700e+04  
## 58 3.845131e+04 0.7317078 1.622923e+04  
## 59 3.843864e+04 0.7318992 1.622510e+04  
## 60 3.843459e+04 0.7319443 1.622391e+04  
## 61 3.830779e+04 0.7336342 1.610522e+04  
## 62 3.827997e+04 0.7340375 1.607777e+04  
## 63 3.825540e+04 0.7343400 1.599482e+04  
## 64 3.820822e+04 0.7349760 1.595260e+04  
## 65 3.817187e+04 0.7354897 1.591057e+04  
## 66 3.814377e+04 0.7358114 1.590084e+04  
## 67 3.804295e+04 0.7372287 1.585740e+04  
## 68 3.800288e+04 0.7377297 1.581199e+04  
## 69 3.797749e+04 0.7380833 1.581913e+04  
## 70 3.777078e+04 0.7404980 1.622695e+04  
## 71 3.707245e+04 0.7504823 1.753329e+04  
## 72 3.671118e+04 0.7549202 1.756730e+04  
## 73 3.606853e+04 0.7640209 1.758294e+04  
## 74 3.509466e+04 0.7756369 1.717807e+04  
## 75 3.284542e+04 0.8026516 1.516267e+04  
## 76 2.891091e+04 0.8494121 9.652062e+03  
## 77 2.880699e+04 0.8505486 9.390927e+03  
## 78 2.874134e+04 0.8511500 9.391043e+03  
## 79 2.855866e+04 0.8531654 9.343040e+03  
## 80 2.855401e+04 0.8532185 9.340143e+03  
## 81 2.855180e+04 0.8532359 9.343469e+03  
## 82 2.852913e+04 0.8534676 9.334426e+03  
## 83 2.844743e+04 0.8541889 9.349442e+03  
## 84 2.833575e+04 0.8552646 9.467717e+03  
## 85 2.820369e+04 0.8565860 9.516266e+03  
## 86 2.819868e+04 0.8566254 9.523623e+03  
## 87 2.817368e+04 0.8568859 9.541537e+03  
## 88 1.289959e+12 0.7697503 1.352022e+10  
## 89 1.236455e+12 0.6827486 1.295944e+10  
## 90 4.330026e+12 0.6827153 4.538354e+10  
## 91 4.456623e+12 0.6826888 4.671041e+10  
## 92 5.172463e+14 0.6826146 5.421028e+12  
##   
## RMSE was used to select the optimal model using the smallest value.  
## The final value used for the model was ncomp = 87.

#M value  
pcr.fit$bestTune  
  
#test error   
pred\_pcr <- predict(pcr.fit, x2, ncomp = 87)  
mse(y2, pred\_pcr)

## [1] 937635353

rmse(y2, pred\_pcr)

## [1] 30620.83

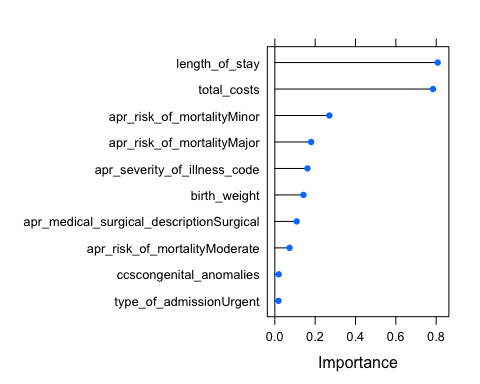
summary(pcr.fit)

## Data: X dimension: 91036 92   
## Y dimension: 91036 1  
## Fit method: svdpc  
## Number of components considered: 87  
## TRAINING: % variance explained  
## 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps 7 comps  
## X 4.665 8.267 11.64 14.34 16.67 18.88 20.83  
## .outcome 32.534 46.443 64.47 64.54 64.57 64.65 64.70  
## 8 comps 9 comps 10 comps 11 comps 12 comps 13 comps  
## X 22.57 23.98 25.35 26.66 27.97 29.24  
## .outcome 66.04 66.37 66.42 66.61 66.96 71.44  
## 14 comps 15 comps 16 comps 17 comps 18 comps 19 comps  
## X 30.47 31.68 32.88 34.05 35.21 36.37  
## .outcome 71.51 72.29 72.29 72.30 72.31 72.31  
## 20 comps 21 comps 22 comps 23 comps 24 comps 25 comps  
## X 37.53 38.67 39.81 40.95 42.09 43.22  
## .outcome 72.45 72.59 72.59 72.60 72.64 72.69  
## 26 comps 27 comps 28 comps 29 comps 30 comps 31 comps  
## X 44.35 45.47 46.60 47.72 48.84 49.95  
## .outcome 72.70 72.81 72.85 72.85 72.96 72.97  
## 32 comps 33 comps 34 comps 35 comps 36 comps 37 comps  
## X 51.07 52.18 53.29 54.40 55.51 56.62  
## .outcome 72.98 73.01 73.01 73.02 73.03 73.06  
## 38 comps 39 comps 40 comps 41 comps 42 comps 43 comps  
## X 57.72 58.83 59.93 61.03 62.13 63.23  
## .outcome 73.06 73.06 73.15 73.17 73.18 73.21  
## 44 comps 45 comps 46 comps 47 comps 48 comps 49 comps  
## X 64.32 65.42 66.51 67.60 68.69 69.79  
## .outcome 73.23 73.23 73.23 73.23 73.24 73.26  
## 50 comps 51 comps 52 comps 53 comps 54 comps 55 comps  
## X 70.87 71.96 73.05 74.14 75.22 76.30  
## .outcome 73.28 73.29 73.29 73.29 73.29 73.33  
## 56 comps 57 comps 58 comps 59 comps 60 comps 61 comps  
## X 77.39 78.47 79.55 80.63 81.7 82.77  
## .outcome 73.35 73.35 73.39 73.39 73.4 73.53  
## 62 comps 63 comps 64 comps 65 comps 66 comps 67 comps  
## X 83.84 84.91 85.96 87.01 88.05 89.08  
## .outcome 73.64 73.65 73.70 73.80 73.80 73.97  
## 68 comps 69 comps 70 comps 71 comps 72 comps 73 comps  
## X 90.1 91.08 92.05 92.97 93.88 94.73  
## .outcome 74.0 74.06 74.08 75.34 75.51 76.88  
## 74 comps 75 comps 76 comps 77 comps 78 comps 79 comps  
## X 95.57 96.29 96.98 97.64 98.29 98.80  
## .outcome 77.24 79.75 84.97 85.05 85.05 85.29  
## 80 comps 81 comps 82 comps 83 comps 84 comps 85 comps  
## X 99.08 99.32 99.56 99.78 99.88 99.94  
## .outcome 85.30 85.30 85.33 85.33 85.48 85.70  
## 86 comps 87 comps  
## X 99.97 100.00  
## .outcome 85.70 85.73

var\_imp = varImp(pcr.fit, scale=FALSE)  
var\_imp

## loess r-squared variable importance  
##   
## only 20 most important variables shown (out of 92)

#Figure 4: Importance of each variable  
plot(var\_imp, top = 10)

 We found the test erorr to be 937635353 with an M value of 87 selected by cross-validation.

# GAM MODEL

Fiting a generalized additive model (GAM) using all the predictors.

set.seed(100)   
  
gam.fit <- train(x, y,  
 method = "gam",  
 tuneGrid = data.frame(method = "GCV.Cp", select = c(TRUE,FALSE)),  
 trControl = ctrl1)  
  
gam.fit$bestTune  
gam.fit$finalModel

##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## .outcome ~ hospital\_countyKings + hospital\_countyManhattan +   
## hospital\_countyQueens + facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division +   
## facility\_nameMaimonides.Medical.Center + facility\_nameMount.Sinai.Hospital +   
## genderM + raceOther.Race + raceWhite + ethnicitySpanish.Hispanic +   
## payment\_typology\_1Medicaid + payment\_typology\_1Private.Health.Insurance +   
## apr\_severity\_of\_illness\_code + s(birth\_weight) + s(length\_of\_stay) +   
## s(total\_costs)  
##   
## Estimated degrees of freedom:  
## 8.90 8.97 8.96 total = 40.82   
##   
## GCV score: 820127821

pred\_gam <- predict(gam.fit, x2)  
mse(y2, pred\_gam)

## [1] 969194746

rmse(y2, pred\_gam)

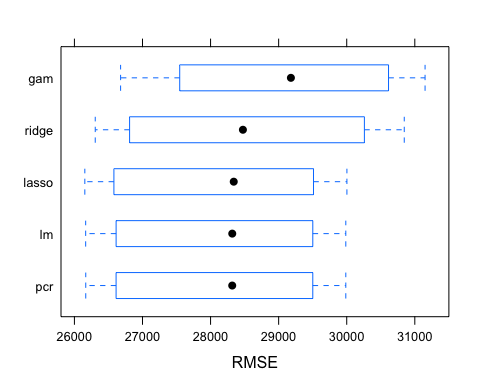
## [1] 31131.89

summary(gam.fit)

##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## .outcome ~ hospital\_countyKings + hospital\_countyManhattan +   
## hospital\_countyQueens + facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division +   
## facility\_nameMaimonides.Medical.Center + facility\_nameMount.Sinai.Hospital +   
## genderM + raceOther.Race + raceWhite + ethnicitySpanish.Hispanic +   
## payment\_typology\_1Medicaid + payment\_typology\_1Private.Health.Insurance +   
## apr\_severity\_of\_illness\_code + s(birth\_weight) + s(length\_of\_stay) +   
## s(total\_costs)  
##   
## Parametric coefficients:  
## Estimate  
## (Intercept) 12214.95  
## hospital\_countyKings -3179.65  
## hospital\_countyManhattan 4082.29  
## hospital\_countyQueens -425.77  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 16773.17  
## facility\_nameMaimonides.Medical.Center 9901.53  
## facility\_nameMount.Sinai.Hospital -5848.08  
## genderM 70.95  
## raceOther.Race 3145.21  
## raceWhite 2503.48  
## ethnicitySpanish.Hispanic -812.46  
## payment\_typology\_1Medicaid -1661.56  
## payment\_typology\_1Private.Health.Insurance -100.30  
## apr\_severity\_of\_illness\_code 3272.90  
## Std. Error  
## (Intercept) 458.98  
## hospital\_countyKings 322.24  
## hospital\_countyManhattan 297.23  
## hospital\_countyQueens 338.33  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 482.32  
## facility\_nameMaimonides.Medical.Center 435.95  
## facility\_nameMount.Sinai.Hospital 432.61  
## genderM 191.74  
## raceOther.Race 267.76  
## raceWhite 293.54  
## ethnicitySpanish.Hispanic 266.04  
## payment\_typology\_1Medicaid 289.68  
## payment\_typology\_1Private.Health.Insurance 308.59  
## apr\_severity\_of\_illness\_code 176.75  
## t value  
## (Intercept) 26.613  
## hospital\_countyKings -9.867  
## hospital\_countyManhattan 13.734  
## hospital\_countyQueens -1.258  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division 34.776  
## facility\_nameMaimonides.Medical.Center 22.712  
## facility\_nameMount.Sinai.Hospital -13.518  
## genderM 0.370  
## raceOther.Race 11.746  
## raceWhite 8.529  
## ethnicitySpanish.Hispanic -3.054  
## payment\_typology\_1Medicaid -5.736  
## payment\_typology\_1Private.Health.Insurance -0.325  
## apr\_severity\_of\_illness\_code 18.517  
## Pr(>|t|)  
## (Intercept) < 2e-16  
## hospital\_countyKings < 2e-16  
## hospital\_countyManhattan < 2e-16  
## hospital\_countyQueens 0.20824  
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division < 2e-16  
## facility\_nameMaimonides.Medical.Center < 2e-16  
## facility\_nameMount.Sinai.Hospital < 2e-16  
## genderM 0.71138  
## raceOther.Race < 2e-16  
## raceWhite < 2e-16  
## ethnicitySpanish.Hispanic 0.00226  
## payment\_typology\_1Medicaid 9.73e-09  
## payment\_typology\_1Private.Health.Insurance 0.74517  
## apr\_severity\_of\_illness\_code < 2e-16  
##   
## (Intercept) \*\*\*  
## hospital\_countyKings \*\*\*  
## hospital\_countyManhattan \*\*\*  
## hospital\_countyQueens   
## facility\_nameLong.Island.Jewish.Schneiders.Children.s.Hospital.Division \*\*\*  
## facility\_nameMaimonides.Medical.Center \*\*\*  
## facility\_nameMount.Sinai.Hospital \*\*\*  
## genderM   
## raceOther.Race \*\*\*  
## raceWhite \*\*\*  
## ethnicitySpanish.Hispanic \*\*   
## payment\_typology\_1Medicaid \*\*\*  
## payment\_typology\_1Private.Health.Insurance   
## apr\_severity\_of\_illness\_code \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Approximate significance of smooth terms:  
## edf Ref.df F p-value   
## s(birth\_weight) 8.895 8.993 107.9 <2e-16 \*\*\*  
## s(length\_of\_stay) 8.966 8.999 3143.9 <2e-16 \*\*\*  
## s(total\_costs) 8.958 8.999 2110.4 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## R-sq.(adj) = 0.852 Deviance explained = 85.2%  
## GCV = 8.2013e+08 Scale est. = 8.1976e+08 n = 91036

# Choosing best model to predict total healthcare charges

resamp <- resamples(list(lm = fit\_lm,   
 lasso = lasso.fit,   
 pcr = pcr.fit,   
 gam = gam.fit,  
 ridge = ridge.fit))  
  
#Figure 3: Boxplot of all 3 models  
bwplot(resamp, metric = "RMSE")



summary(resamp)

##   
## Call:  
## summary.resamples(object = resamp)  
##   
## Models: lm, lasso, pcr, gam, ridge   
## Number of resamples: 10   
##   
## MAE   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## lm 9203.205 9392.192 9569.649 9541.894 9721.208 9814.831 0  
## lasso 9042.496 9249.446 9425.081 9399.915 9593.689 9674.869 0  
## pcr 9203.646 9392.051 9566.931 9541.537 9721.594 9815.186 0  
## gam 8140.379 8344.797 8477.576 8485.779 8696.123 8825.095 0  
## ridge 8601.740 8712.862 8985.624 8931.058 9146.153 9224.492 0  
##   
## RMSE   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## lm 26165.24 26844.17 28318.51 28174.24 29429.14 29986.34 0  
## lasso 26152.81 26816.58 28340.35 28180.84 29441.95 30002.69 0  
## pcr 26166.42 26842.52 28318.00 28173.68 29428.94 29986.34 0  
## gam 26678.34 27548.32 29180.06 29076.52 30604.48 31150.71 0  
## ridge 26305.89 26931.26 28475.02 28517.52 30140.80 30845.07 0  
##   
## Rsquared   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## lm 0.8378441 0.8507103 0.8551594 0.8568802 0.8664776 0.8734834 0  
## lasso 0.8379075 0.8508073 0.8548708 0.8568294 0.8664416 0.8738037 0  
## pcr 0.8379046 0.8507118 0.8551640 0.8568859 0.8664763 0.8734683 0  
## gam 0.8280900 0.8378257 0.8409091 0.8475277 0.8575302 0.8755479 0  
## ridge 0.8341629 0.8488328 0.8536992 0.8548152 0.8631452 0.8731384 0

# Using final PCR model to predict test data

#Using PCR model to predict total charges of test dataset  
predictions <- pcr.fit %>% predict(x2)  
predictions

## [1] 463312.72333 -7570.31599 -2393.73226 753541.03700  
## [5] -1770.21342 -556.19516 -4833.81766 -1714.70853  
## [9] -8344.03682 25730.71087 34396.58579 16200.85973  
## [13] 584.18873 -7429.87220 -7587.95560 -1876.16538  
## [17] 33131.94499 -8781.16275 -10233.82119 -501.92483  
## [21] -6253.82193 -3913.07012 -1255.18887 -7534.12508  
## [25] -2350.54150 -1317.05285 -7511.35872 -9017.65945  
## [29] -7832.47408 34129.54259 -1299.24063 -6681.37518  
## [33] -2557.68974 210341.92435 25047.49941 -5460.24238  
## [37] -9269.47001 -7590.77165 -7559.60764 -2504.15541  
## [41] -7831.27342 -7994.33853 -8982.60159 -403.50299  
## [45] 10472.27951 -7147.79431 -3175.73796 589735.73273  
## [49] 37979.12648 -7097.73104 403.34184 -578.99670  
## [53] 3065.47224 23509.48165 -6691.14485 -9215.99412  
## [57] -7612.46253 -8011.69512 -8004.70567 -1388.08692  
## [61] 310.43764 -7522.50330 -7611.28275 -7653.79784  
## [65] 8232.01894 -8953.48194 -694.22538 -6781.79544  
## [69] -9689.67887 -5119.07855 -2751.40168 -8055.53517  
## [73] 272667.96675 -10.58417 -7588.23293 531071.98314  
## [77] -2944.71188 -7436.57213 -7686.80777 32465.34764  
## [81] -6907.03952 -8948.31938 3343.75432 12176.29776  
## [85] -7956.64676 -6167.32099 -585.39748 -8642.37733  
## [89] -7939.35038 -517.32182 -8853.79079 -8671.67146  
## [93] -6702.76663 -7869.66146 -9905.49324 345.66404  
## [97] 40636.47274 2323.05349 -10357.66071 -10851.20436  
## [101] -7839.62165 -6420.58736 -738.77153 -9336.84665  
## [105] -7795.38786 1008.48345 161.31895 -1306.52003  
## [109] -19766.84940 -7814.59273 -2340.18658 148317.95378  
## [113] -8069.47707 126.85315 -2029.44785 -8810.51120  
## [117] -1859.52101 -126.91970 1052.74133 -1234.68198  
## [121] -3287.07435 429.22249 -669.34813 -7820.99350  
## [125] -7764.22100 6826.96227 -7168.12622 4907.31084  
## [129] -3989.06126 -6371.33723 -7825.52314 1554.93313  
## [133] -1067.89458 -7050.61092 -8241.89373 -6753.51994  
## [137] -7698.42408 118420.25979 -6965.50761 -5479.08652  
## [141] -5895.53563 -9074.10567 -5737.36446 -7834.23674  
## [145] -7304.38856 -7488.10073 -9306.06081 -7916.53882  
## [149] -7981.66052 -8847.58842 -15260.14474 85.15110  
## [153] -9282.71324 -4312.28964 230200.10812 -7380.58372  
## [157] -13039.50124 657.06469 -7743.61897 -12572.75797  
## [161] -3696.29171 -2238.73561 1849.17888 322666.07527  
## [165] -4903.22844 -6934.73369 14152.61943 -8328.71872  
## [169] -8089.80898 -9063.16302 9935.68931 -9079.61392  
## [173] -4781.35102 22367.85960 -5742.59739 -7405.31428  
## [177] 21855.25923 2825.37920 48534.38119 -7613.15389  
## [181] -167.91510 -9307.72700 -6756.87448 7326.56041  
## [185] -2615.40253 11880.06353 -8882.35487 -7736.74369  
## [189] -6352.81704 -1722.44958 -9103.45412 -3652.64790  
## [193] -2058.45207 -8312.08313 -6934.73369 -9477.30962  
## [197] -8272.09568 -7141.39353 -903.03885 58126.26429  
## [201] -6232.49571 -6287.08741 -3843.22786 -8181.77136  
## [205] -8268.30015 -6726.89408 -6949.61885 -71.68766  
## [209] -7085.18038 -6669.50115 -2427.28498 589.33907  
## [213] -7137.34038 -5642.75249 -8394.83198 -7861.63746  
## [217] -9066.85140 -3194.31678 -3493.00772 -11622.31592  
## [221] 142030.34612 -7258.45053 7094.91602 1595.90377  
## [225] 418035.10762 -7902.80468 -5926.59680 14228.56206  
## [229] -7571.20230 58402.55539 -6447.50680 -4183.83550  
## [233] -8633.38707 1363.15695 -8402.56075 -2809.41148  
## [237] -4936.38676 -7610.30865 450277.81051 8882.25776  
## [241] 1298.55964 486.18209 770.44931 -6373.98646  
## [245] -3516.98905 -5941.67037 -5091.22668 385167.02836  
## [249] -7146.57765 -96.07396 -3868.31646 1511.77068  
## [253] -12916.69999 -7305.71826 1274.58262 50835.73667  
## [257] -6515.33004 3263.15762 -5744.09397 -6716.76888  
## [261] 2423.58865 1640.52304 -5474.95221 -6991.98513  
## [265] 55117.95240 17358.43196 323.99616 -8791.65671  
## [269] -6527.83444 591.95815 -6498.05698 -4779.12693  
## [273] 2157.18199 1565.84565 4053.52066 -5696.04698  
## [277] -5972.47212 -6320.74558 2141.36402 3419.41560  
## [281] -5910.96335 356542.05674 15991.70491 9365.79220  
## [285] -6769.14128 -887.23384 -7093.92852 25713.78803  
## [289] 3805.03806 751.29717 -5621.35767 2355.43704  
## [293] 936.19148 -6818.71497 -3778.96439 1285.81969  
## [297] -3427.49420 -11249.36044 3768.59952 -4087.45157  
## [301] 3204.86353 -3628.93304 1090702.91364 1825.04490  
## [305] -3501.51694 -7523.99988 -5456.40993 -5536.25435  
## [309] 16121.66292 -3468.13816 95143.85317 515.76338  
## [313] 1572.74655 -3386.81122 29879.23926 1458.23635  
## [317] -3848.39443 2151.75951 1962.58561 3926.45267  
## [321] 4121.31337 -5413.51805 40470.61905 -5536.25435  
## [325] -6815.21861 8542.49131 -5987.35653 -7205.04575  
## [329] 3091.19491 -4643.76226 1077.98008 -6900.48105  
## [333] 2962.44255 2404.10500 1264.80590 -630.79877  
## [337] -12298.21795 -5518.59958 -6271.74121 -4735.50774  
## [341] 19606.82763 -4280.06235 3323.22040 33619.86502  
## [345] -5752.34397 28315.73320 8434.71352 -5737.69319  
## [349] -6783.29202 -5920.38977 8024.70626 -6326.64257  
## [353] -12560.95892 -10534.01449 -5737.69319 -5081.48912  
## [357] 253.02240 -5418.73906 65859.03953 -2774.62692  
## [361] 12888.60140 3291.13984 288.65740 -4878.16607  
## [365] -6895.26004 -6460.61594 -7969.02755 -6104.87182  
## [369] -3169.97423 722.67161 24496.11070 -1937.87604  
## [373] 76334.84270 7059.66674 -5940.31180 -3776.65028  
## [377] 131800.31116 683.52067 26778.93920 -4766.63346  
## [381] -10963.42331 -262.34797 -500.90307 -6253.07549  
## [385] -5178.05167 3976.08769 764.65479 -4901.28111  
## [389] -5452.42060 4240.45126 -1814.56128 -6361.21203  
## [393] -6776.89125 -6203.22014 -6118.80296 -6839.04689  
## [397] 46607.67724 1998.82126 -3670.75676 150782.78723  
## [401] 53003.65316 -12998.76885 -6331.86358 -3524.45328  
## [405] -6118.80296 2148.72194 -5120.57513 40753.93031  
## [409] 2819.78422 -1855.10162 -6777.05862 5983.47635  
## [413] 3175.44472 -4001.33268 -4154.91480 -6098.47105  
## [417] 124.27004 -3868.31646 2380.74744 -5379.40689  
## [421] -7151.81058 930.07557 -5880.18816 4323.68166  
## [425] -5481.35299 -6020.94359 -5421.71704 -6215.52805  
## [429] -6575.45241 -5890.63144 66933.43473 -4199.74587  
## [433] -7587.80979 -4760.65077 -5446.78354 -5249.32749  
## [437] 237356.26268 1940.88233 -3848.39443 -4491.83529  
## [441] 194.69810 -5738.87296 -5446.78354 -4997.83883  
## [445] -13928.56946 -5506.90591 -6266.52020 -6931.00927  
## [449] -5856.06198 -5096.28560 -4214.36324 -7390.82030  
## [453] -5518.59958 -6985.91063 5570.52496 68732.35824  
## [457] -6188.05003 -1143.44522 -6042.25789 1786.98694  
## [461] 7314.23459 -6038.75856 1725.62635 -4411.79089  
## [465] -7444.14352 -731.27429 -5060.45276 -6747.54280  
## [469] 388.32281 -5335.26912 -6655.33472 -5916.18435  
## [473] 1084.38085 172096.17909 8075.37605 -6498.94299  
## [477] -4748.35300 -5318.62475 -3523.03952 15670.93895  
## [481] 10119.42598 668.70163 -8344.95324 1382.12578  
## [485] -6250.09750 -5834.41828 2196.26539 -32.57742  
## [489] 9008.13963 1809.93461 -6375.28314 -3372.34444  
## [493] -5245.79830 -6250.09750 -6576.30591 1084.38085  
## [497] -3938.71875 -13137.59862 -7187.51684 10159.73253  
## [501] -5363.83769 45.18279 -13308.06756 -5623.20354  
## [505] -6962.63668 -4795.22022 74.53124 1137.91518  
## [509] -4436.93393 -5198.26094 -5122.20850 -246.08926  
## [513] 16737.52327 -5971.25309 -5536.25435 1757.58005  
## [517] -6949.38385 -6362.39180 -5774.29591 -7163.22202  
## [521] -6479.58082 1786.28981 -6250.09750 -5083.81032  
## [525] -5506.90591 48316.34875 7594.70085 -1677.72154  
## [529] -8930.13325 339323.87069 -6826.74912 1761.84175  
## [533] 1125.15267 -4942.93746 -6470.44151 -6337.19351  
## [537] -4467.70785 -13288.04531 38748.10990 38566.84714  
## [541] -4170.68466 -5329.26824 -6711.54787 -3703.13730  
## [545] -5481.35299 -6569.05164 78101.98850 -4808.47538  
## [549] -14379.60255 -6124.02397 45118.97928 1015.05373  
## [553] -5579.70131 -5946.71257 -6504.45775 -3169.97423  
## [557] 67436.20124 -5744.09397 531.89422 -5891.81121  
## [561] 7974.47166 -5056.94447 2051.03970 -6118.80296  
## [565] -4259.80979 8528.32493 9760.87341 -6368.46630  
## [569] -6124.02397 1994.82655 -5917.57075 -4441.78322  
## [573] -5089.03132 -6290.18438 -4234.31532 -5732.47219  
## [577] -6355.99102 -6283.36299 -7182.12846 -5542.32885  
## [581] 1143.03264 -4608.00873 545.11183 -7071.01394  
## [585] 10239.78694 8158.73735 -4497.05630 -5537.10785  
## [589] 56365.84094 101536.54926 -6528.58519 -6295.86865  
## [593] -5916.18435 -6432.73450 37775.82557 912.84654  
## [597] 1328.52576 18709.00903 -5323.19374 337.34258  
## [601] 13580.17810 -6099.65082 21777.51009 15629.91340  
## [605] -6687.42043 -6988.77601 -6070.30237 255.48769  
## [609] 52495.40278 52714.83244 109607.53535 84909.49913  
## [613] 16112.53860 33489.00797 24532.45231 14015.16959  
## [617] 19851.88653 15223.20469 14634.41712 20001.28136  
## [621] 20707.00752 443686.26386 20686.11657 58446.11514  
## [625] 17537.77090 20747.17467 14556.14621 41754.01930  
## [629] 14327.53203 15660.26466 32430.58167 20125.11866  
## [633] 20484.97867 22077.08710 14573.65995 22709.60838  
## [637] 14444.83607 14151.16476 13404.28310 23037.25725  
## [641] 10874.63791 27907.19460 14953.95199 15568.41652  
## [645] 15351.63317 16346.19831 15165.05770 14483.37140  
## [649] 15884.89667 14265.20673 24285.26456 15168.55364  
## [653] 20550.28932 20814.37988 13979.96953 28143.08833  
## [657] 21389.92063 15449.32614 57279.08651 289742.48719  
## [661] 61969.22032 13290.67820 16310.87486 16735.76091  
## [665] 20934.92059 25835.97382 13331.69555 15830.43602  
## [669] 15475.99966 17298.04861 566381.28146 17384.85650  
## [673] 12682.04943 21134.77373 715487.10661 14550.69467  
## [677] 14880.21149 14633.07851 21950.47676 13789.70816  
## [681] 13728.08963 19670.90848 27146.56797 14667.29649  
## [685] 14102.96023 23611.69866 15293.12179 14238.84620  
## [689] 15106.85447 16761.33766 15667.21440 28157.21665  
## [693] 14977.38080 21239.73593 14491.75309 13092.04469  
## [697] 24601.74545 15244.58544 14426.60134 15626.62984  
## [701] 224827.18536 16977.75975 21808.21312 20729.22318  
## [705] 18928.99194 12949.35394 14628.01634 46857.00915  
## [709] 14692.12885 17845.17006 14957.21809 15569.42402  
## [713] 15232.23951 87051.44349 16503.01904 22858.69340  
## [717] 15580.73693 22119.07042 23133.77310 18389.06881  
## [721] 16086.77623 15356.35044 22273.72400 15847.23249  
## [725] 23578.83160 15768.27065 14374.10187 19484.53450  
## [729] 15843.17133 22659.41138 14780.54595 15284.65779  
## [733] 20625.71753 16961.83449 20884.42747 14527.37947  
## [737] 13765.94909 15259.17755 25693.65704 25485.81743  
## [741] 15961.26948 15491.27992 14595.92705 16018.50051  
## [745] 38683.03174 13509.60889 13667.22577 18390.03721  
## [749] 26014.20680 15997.28346 28429.22191 15410.70683  
## [753] 13522.59224 13752.63312 14582.81891 20124.71959  
## [757] 15176.70699 14809.38068 16331.35102 1053856.01185  
## [761] 34942.40679 14008.34435 14008.38011 20221.27372  
## [765] 109363.94406 20356.61967 16736.36548 196193.95312  
## [769] 22069.31519 124153.42891 14557.80033 15514.84319  
## [773] 21707.70786 14777.88745 14033.92436 53395.98582  
## [777] 25626.31690 56964.68639 15021.98949 22099.86358  
## [781] 14578.29586 13706.91431 15626.57773 22958.26895  
## [785] 17517.08462 20437.23532 39561.72222 14427.12248  
## [789] 15094.35007 22686.24530 15034.44222 14597.85871  
## [793] 17594.91016 129127.53841 46942.76545 123974.85126  
## [797] 15885.55195 21705.21163 21273.36128 18214.67414  
## [801] 20706.59760 16089.60391 13333.02306 40822.62753  
## [805] 16473.56490 217817.57224 225586.08254 13950.17688  
## [809] 21580.63392 15363.01890 15161.21086 29111.61413  
## [813] 15580.28269 56037.53917 15121.28056 46989.22521  
## [817] 16965.57933 16038.32283 15691.78173 12202.56208  
## [821] 19246.66063 14619.24210 25893.28840 23150.94954  
## [825] 17975.84161 20061.66121 27905.76376 15923.92381  
## [829] 33830.85298 13510.30241 23799.96783 55722.51803  
## [833] 45110.72880 22896.04864 21470.51980 15260.34961  
## [837] 25527.14856 60002.39245 27673.30841 21998.76582  
## [841] 15411.41917 15100.97060 30964.05888 21095.03138  
## [845] 16311.77403 15346.60606 15216.11223 16284.68257  
## [849] 21056.65783 22276.07316 13940.85017 15819.16687  
## [853] 25560.81916 22787.60313 22671.14264 14592.61723  
## [857] 13998.42301 22567.38405 14010.89828 13553.43984  
## [861] 13921.49213 14798.82358 16290.73323 20492.98117  
## [865] 16242.81338 19363.42446 15427.29445 13416.68279  
## [869] 14420.17673 99618.90535 20561.35149 26363.35058  
## [873] 21009.91565 14206.12573 11819.72485 24081.21477  
## [877] 14836.22516 15631.79145 15391.86555 15875.05401  
## [881] 21903.91187 15967.89790 8518.64361 16782.34318  
## [885] 16306.22937 26131.90803 18800.87857 27553.31446  
## [889] 16349.17487 15546.23998 338376.65659 17480.01362  
## [893] 147986.10390 26341.13111 14022.96233 20864.71381  
## [897] 14897.15809 20964.30343 15312.30744 14234.79623  
## [901] 20078.25617 21908.46535 13934.37467 45800.32491  
## [905] 18245.22986 15043.69556 16757.89137 15948.47539  
## [909] 14401.00332 15992.56931 13593.27243 14365.82034  
## [913] 15050.09634 15267.66855 20296.09634 152463.74569  
## [917] 15722.11577 21796.87067 15593.66076 17179.34288  
## [921] 15759.72175 87617.49990 61323.67313 15355.69996  
## [925] 40402.76462 12520.51126 15671.34733 15148.39734  
## [929] 14314.38196 54315.32720 16470.46884 20996.55406  
## [933] 12889.88904 14124.29158 36779.53021 15060.25040  
## [937] 17773.12418 19491.83677 14760.97297 27863.86664  
## [941] 15331.90948 23524.76147 363726.54792 116403.11755  
## [945] 174343.99983 15992.56931 15163.48964 15864.68395  
## [949] 19242.97030 78460.91368 14227.34421 21306.09496  
## [953] 14069.37568 21913.41571 15787.89841 14196.21455  
## [957] 1011501.82036 21707.74440 23161.44159 12864.74952  
## [961] 19914.84010 14080.09398 23727.78347 13664.26583  
## [965] 12987.49387 61620.41039 131054.37232 14412.79982  
## [969] 20131.00483 15669.56470 27896.95480 14332.13120  
## [973] 14440.87828 27506.82035 14099.60214 12562.26066  
## [977] 14103.73270 22059.05476 16532.59712 14902.72521  
## [981] 15103.85600 16764.90439 56660.88303 20433.96539  
## [985] 15542.13034 20354.37023 15784.72970 15587.65059  
## [989] 15844.13948 15933.54595 25774.76612 14900.58130  
## [993] 16258.85056 15257.93595 23307.09665 14717.36453  
## [997] 38699.49223 25026.21051 21368.01933 19270.35307  
## [ reached getOption("max.print") -- omitted 29344 entries ]

# Conclusion

Of all models, the PCR model best predicted total healthcare charges of patients at time of discharge. The PCR model had very similar RMSE values to the linear model and lasso regression model, and thus, they have similar prediction performance. By listing the most important variables in predicting charges, we gained insight into the most important predictors for healthcare charges, particularly gaining knowledge that length of stay and total costs have much greater importance in predicting the charges compared to other variables. Birthweight is the top 6th highest importance, which is expected, as it often recorded in literature. It was of valuable insight to learn that the APR risk of mortality (minor, major, and moderate) has strong importance in determining total charges as well as congenital anomalies and Urgent admission types.