DengAI Predicting Disease Spread

Radhika Sood, Rupanjali Chattopadhyay, Rekha Raj, Achal Khullar, Ajinkya Dalvi, Aman Rastogi, Prathiba Swamykannu, Umesh Singh

April 7, 2020

Summary

Dengue is a mosquito-borne disease. It occurs mainly in the tropical and subtropical parts of the world. Since, it is transmitted by mosquitoes, the transmission of the disease is related to the climatic conditions and environmental variables such as precipitation and temperature. The disease is prevalent in Southeast Asia and Pacific Islands and epidemics of this disease are expected based on differences in climatic condtions. Nearly half a million cases of the dengue fever every year are reported in the Latin America, as reported by DataDriven.org.

Data Source The dataset was collected and publicaly shared by *"DrivenData.org"*.The link to original dataset can be found here.The environmental data has been collected by the U.S. Federal Government agencies - Centers for Disease Control (CDC) and Prevention to the National Oceanic and Atmospheric Administration (NOAA).

Goal The goal of this project is to build supervised learning model to predict the number of dengue fever cases each week in the cities of *San Juan, Puerto Rico and Iquitos, Peru* based on the features provided in the test data set, from 2008 (week 18) till 2013 (week 13) for San Juan , and from 2010 (week 26) till 2013

(week 26) for Iquitos. The champion model will be used for predicting the total cases per week from the features in the test set.

Algorithms used We used several algorithms in supervised learning models including Decision (Regression) Tree, Random Forest and Extreme Gradient Boosting, Partial Least Squares, GLMNET for building the prediction model on a training set and compare their performance. Finally the champion model was chosen for predicting on a future dataset.

Loading Libraries

```
# install.packages("RCurl")
# install.packages("e1071")
# install.packages("caret")
# install.packages("doSNOW")
# install.packages("ipred")
# install.packages("xgboost")
# install.packages("dplyr")
# install.packages("tidyr")
# install.packages("naniar")
# install.packages("corrplot")
# install.packages("gbm")
# install.packages("mda")
# install.packages("psych")
# install.packages("kknn")
# install.packages("pls")
# install.packages("pamr")
# install.packages("mda")
# install.packages("rattle")
# install.packages("vtreat")
library(RCurl)
```

```
## Loading required package: bitops
```

```
library(e1071)
library(caret)
```

```
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
library(doSNOW)
```

```
## Loading required package: foreach

## Warning: package 'foreach' was built under R version 3.5.3

## Loading required package: iterators

## Warning: package 'iterators' was built under R version 3.5.3

## Loading required package: snow
```

```
library(ipred)
library(xgboost)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
## The following object is masked from 'package:xgboost':
##
##     slice
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
##
## Attaching package: 'tidyr'
## The following object is masked from 'package:RCurl':
##
##     complete
```

```r
library(naniar)
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
library(psych)
```

```
##
## Attaching package: 'psych'
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```r
library(grid)
library(ggplot2)
library(kknn)
```

```
##
## Attaching package: 'kknn'
## The following object is masked from 'package:caret':
##
##     contr.dummy
```

```r
library(pls)
```

```
##
## Attaching package: 'pls'
## The following object is masked from 'package:corrplot':
##
##     corrplot
## The following object is masked from 'package:caret':
##
```

```
##     R2
## The following object is masked from 'package:stats':
##
##     loadings
```

```r
library(pamr)
```

```
## Loading required package: cluster
## Loading required package: survival
##
## Attaching package: 'survival'
## The following object is masked from 'package:caret':
##
##     cluster
## The following object is masked from 'package:rpart':
##
##     solder
```

```r
library(mda)
```

```
## Loading required package: class
## Loaded mda 0.4-10
```

```r
library(rattle)
```

```
## Rattle: A free graphical interface for data science with R.
## Version 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
##
## Attaching package: 'rattle'
## The following object is masked from 'package:xgboost':
##
##     xgboost
```

```r
library(vtreat)
library(glmnet)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following object is masked from 'package:tidyr':
##
##     expand
## Loaded glmnet 2.0-16
```

Data Preparation Steps

The dataset contains both train and test data. We will split train data and use one part (i.e., the major part of the split) to train the predictive model and use the other smaller part to test the performance of the predictive model/regressor. The new test dataset will be used for validation.

Importing Datasets Into the R-Console

##Importing dengue_features_train and label dataset using "getURL" method from the RCurl package. This dataset contains information about the various features that can affect the incidence of the cases of dengue (mosquito-borne disease) per week.

Importing the training data features and labels and then merging them by their composite keys (i.e., a combination of 'city', 'year' and 'week of year')

```
trfeat <- getURL("https://s3.amazonaws.com/drivendata/data/44/public/dengue_features_train.csv")
trfeat <-read.csv(text = trfeat)
trfeat <- trfeat[, -c(4)]
trlabel <- getURL("https://s3.amazonaws.com/drivendata/data/44/public/dengue_labels_train.csv")
trlabel <- read.csv(text = trlabel)
trmerge <- merge(trfeat, trlabel, by=c("city", "year", "weekofyear"))
names(trmerge)
```

```
##  [1] "city"
##  [2] "year"
##  [3] "weekofyear"
##  [4] "ndvi_ne"
##  [5] "ndvi_nw"
##  [6] "ndvi_se"
##  [7] "ndvi_sw"
##  [8] "precipitation_amt_mm"
##  [9] "reanalysis_air_temp_k"
## [10] "reanalysis_avg_temp_k"
## [11] "reanalysis_dew_point_temp_k"
## [12] "reanalysis_max_air_temp_k"
## [13] "reanalysis_min_air_temp_k"
## [14] "reanalysis_precip_amt_kg_per_m2"
## [15] "reanalysis_relative_humidity_percent"
## [16] "reanalysis_sat_precip_amt_mm"
## [17] "reanalysis_specific_humidity_g_per_kg"
## [18] "reanalysis_tdtr_k"
## [19] "station_avg_temp_c"
## [20] "station_diur_temp_rng_c"
## [21] "station_max_temp_c"
## [22] "station_min_temp_c"
## [23] "station_precip_mm"
## [24] "total_cases"
```

```
dengue_train <- trmerge[,c(-2)]
names(dengue_train)
```

```
##  [1] "city"
##  [2] "weekofyear"
##  [3] "ndvi_ne"
##  [4] "ndvi_nw"
##  [5] "ndvi_se"
##  [6] "ndvi_sw"
##  [7] "precipitation_amt_mm"
##  [8] "reanalysis_air_temp_k"
```

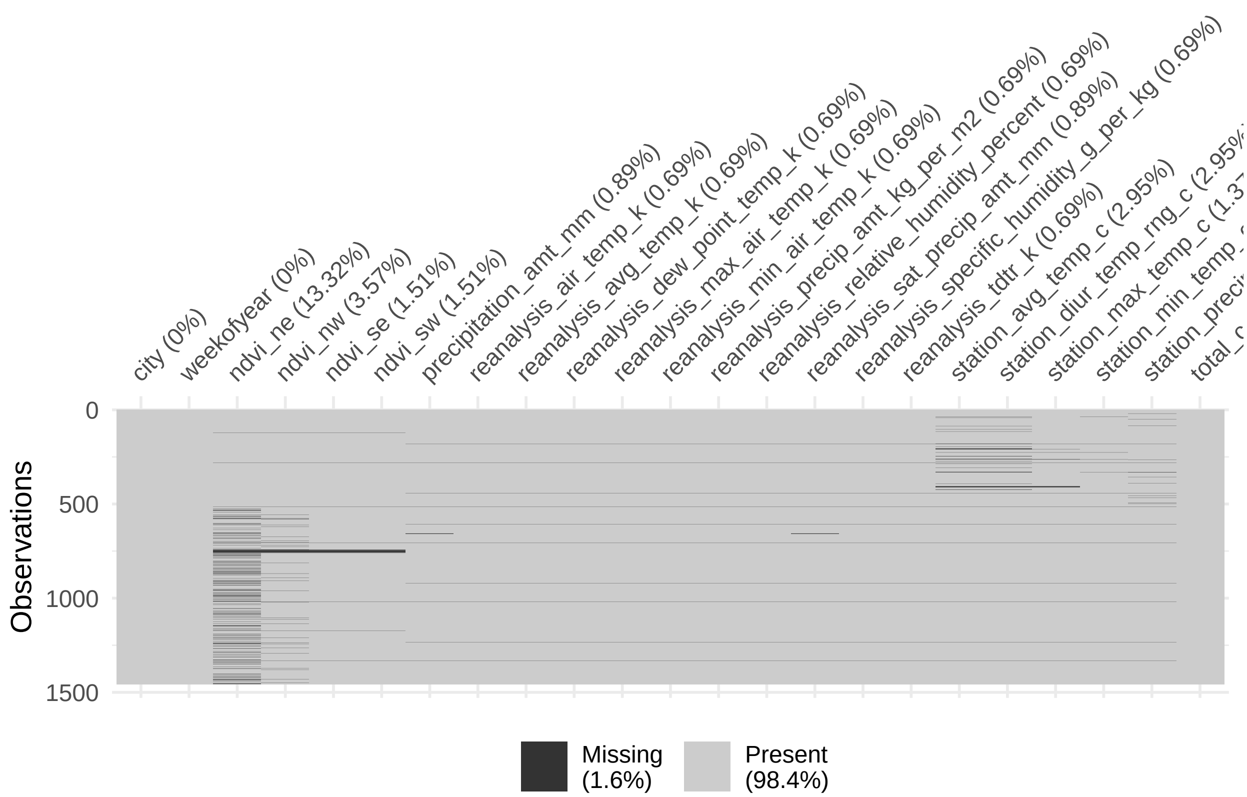
```
##  [9] "reanalysis_avg_temp_k"
## [10] "reanalysis_dew_point_temp_k"
## [11] "reanalysis_max_air_temp_k"
## [12] "reanalysis_min_air_temp_k"
## [13] "reanalysis_precip_amt_kg_per_m2"
## [14] "reanalysis_relative_humidity_percent"
## [15] "reanalysis_sat_precip_amt_mm"
## [16] "reanalysis_specific_humidity_g_per_kg"
## [17] "reanalysis_tdtr_k"
## [18] "station_avg_temp_c"
## [19] "station_diur_temp_rng_c"
## [20] "station_max_temp_c"
## [21] "station_min_temp_c"
## [22] "station_precip_mm"
## [23] "total_cases"
```

```r
dim(dengue_train)
```

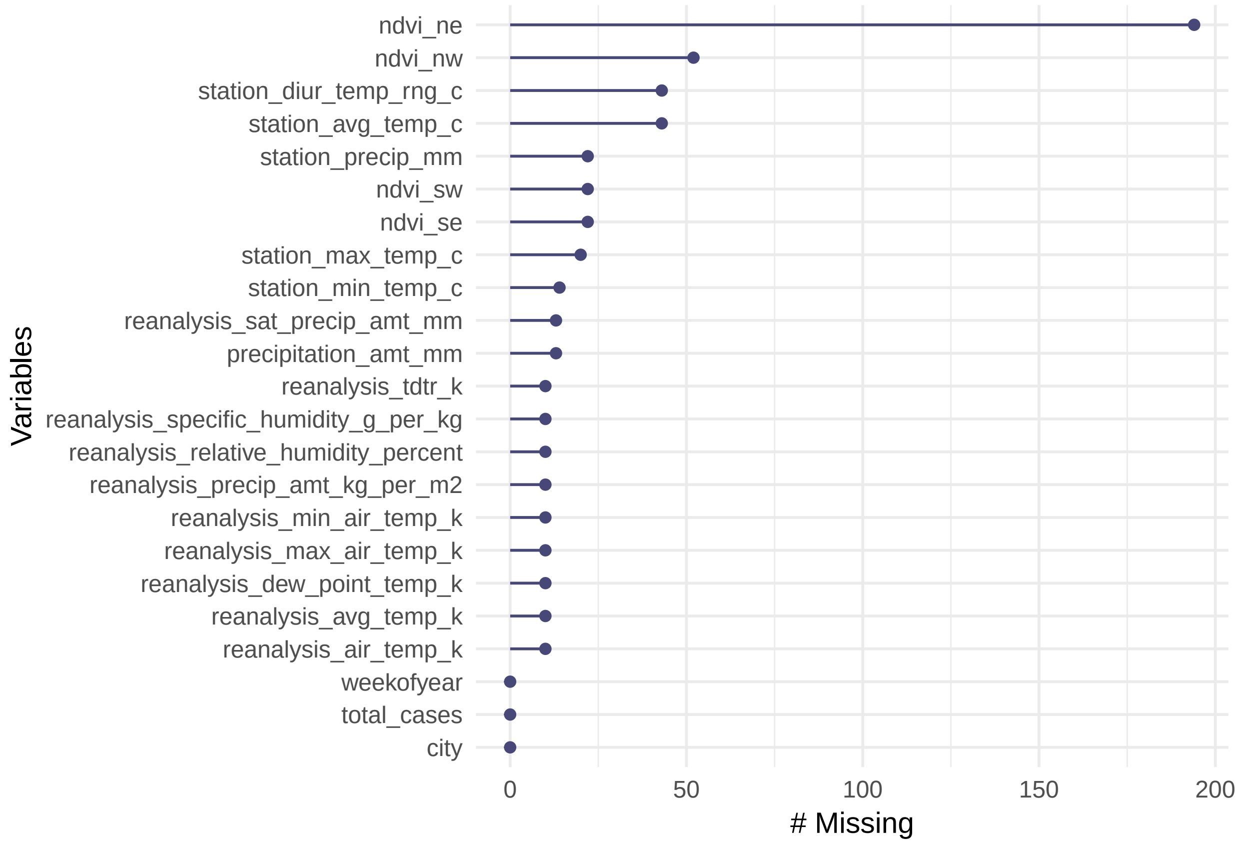
```
## [1] 1456   23
```

Checking and visualizing missing values in the merged training data

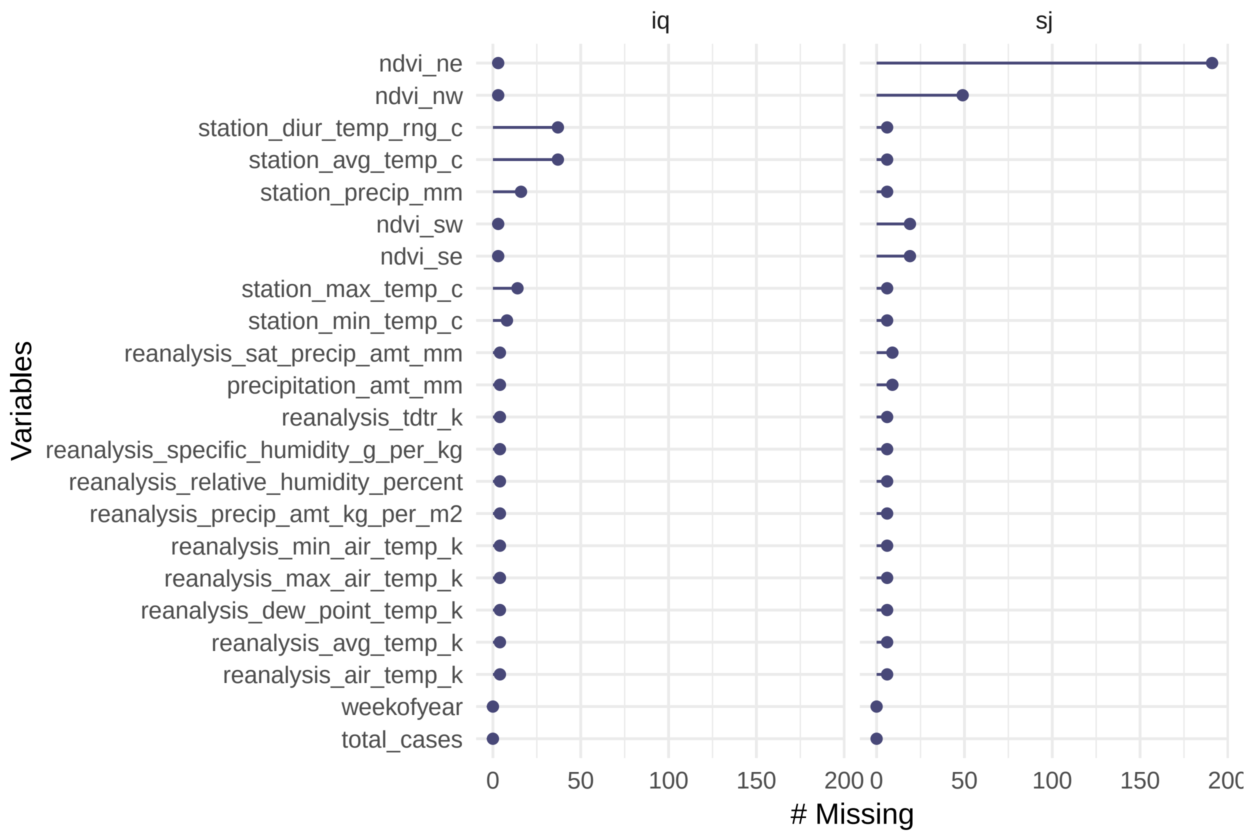
```r
# Visualizing missing values for the training data
vis_miss(dengue_train)
```



```r
gg_miss_var(dengue_train) + theme_minimal()
```

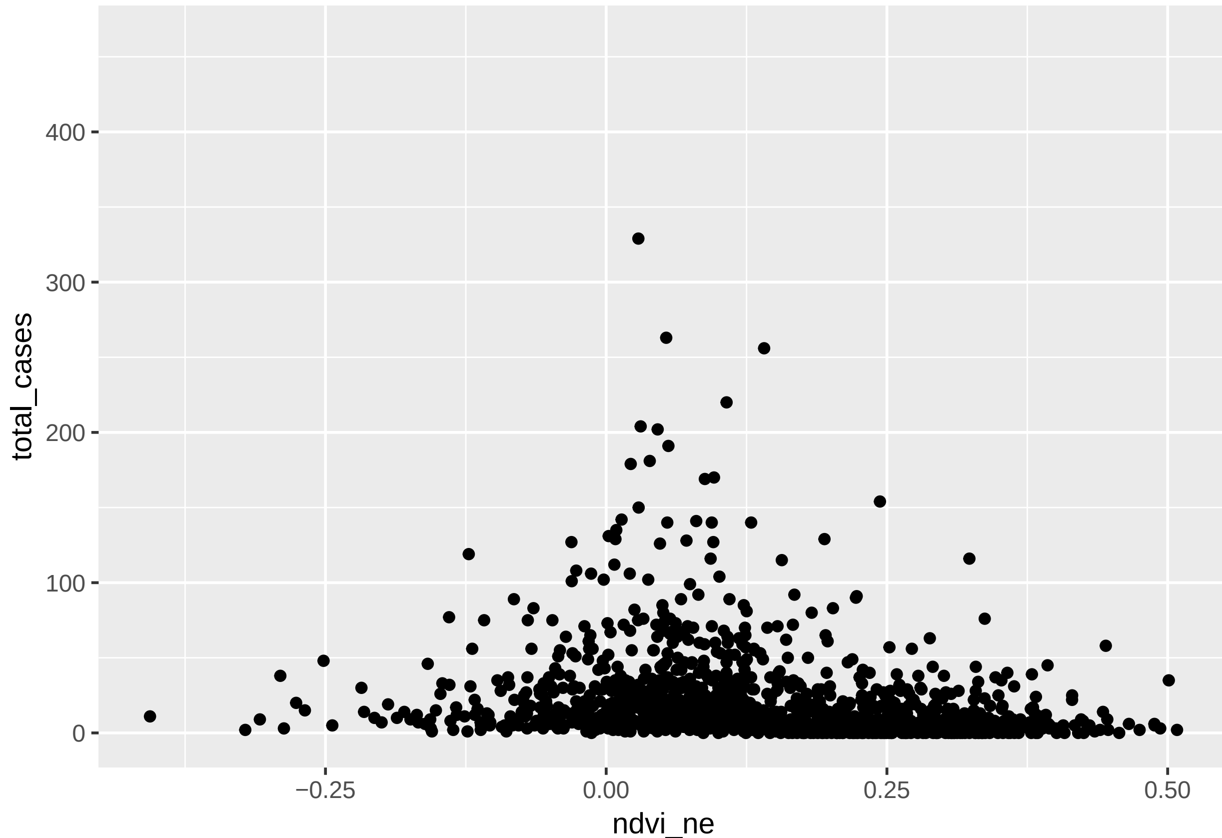



```r
gg_miss_var(dengue_train, facet = city) + theme_minimal()
```

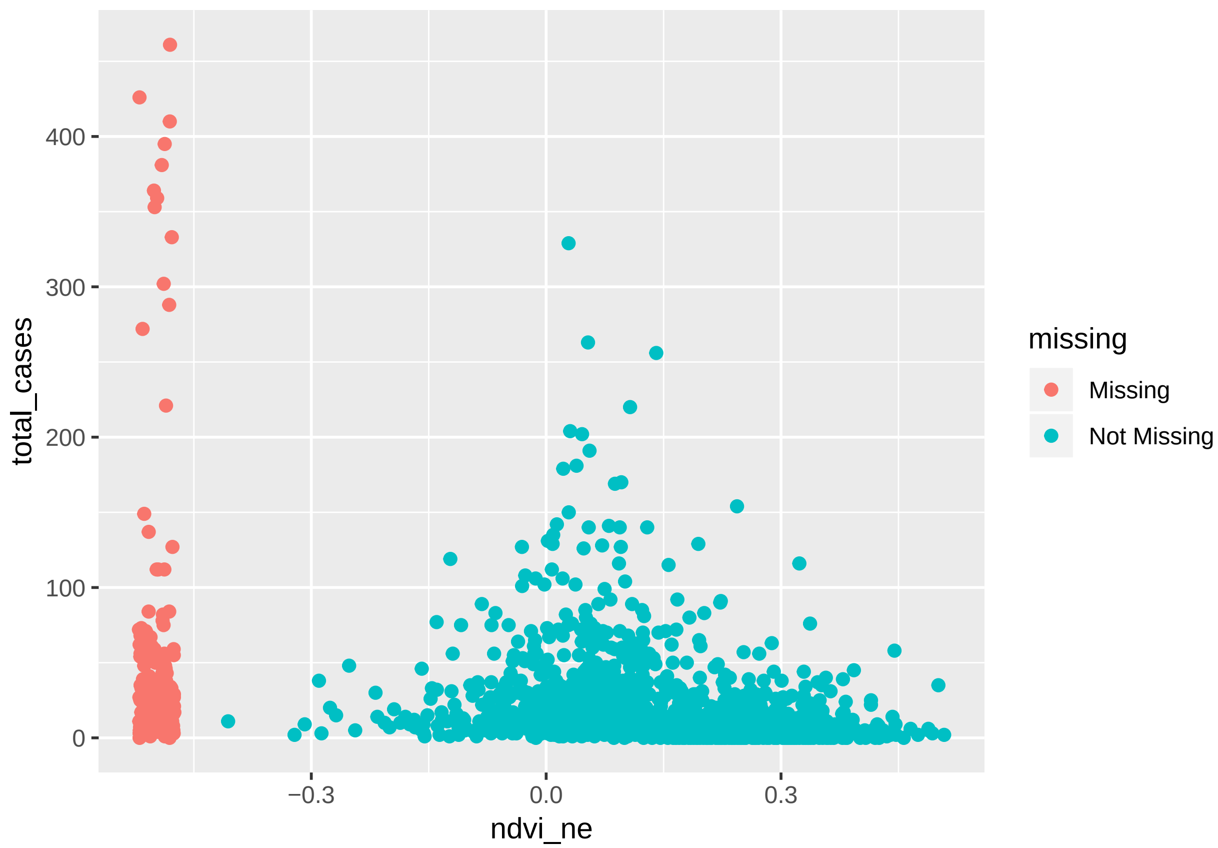



```r
ggplot(dengue_train, aes(x=ndvi_ne, y = total_cases)) + geom_point()
```

```
## Warning: Removed 194 rows containing missing values (geom_point).
```




```r
ggplot(dengue_train, aes(x=ndvi_ne, y = total_cases)) + geom_miss_point()
```



Conclusion: Most of the missing values can be classified as 'Missing Not At Random'.

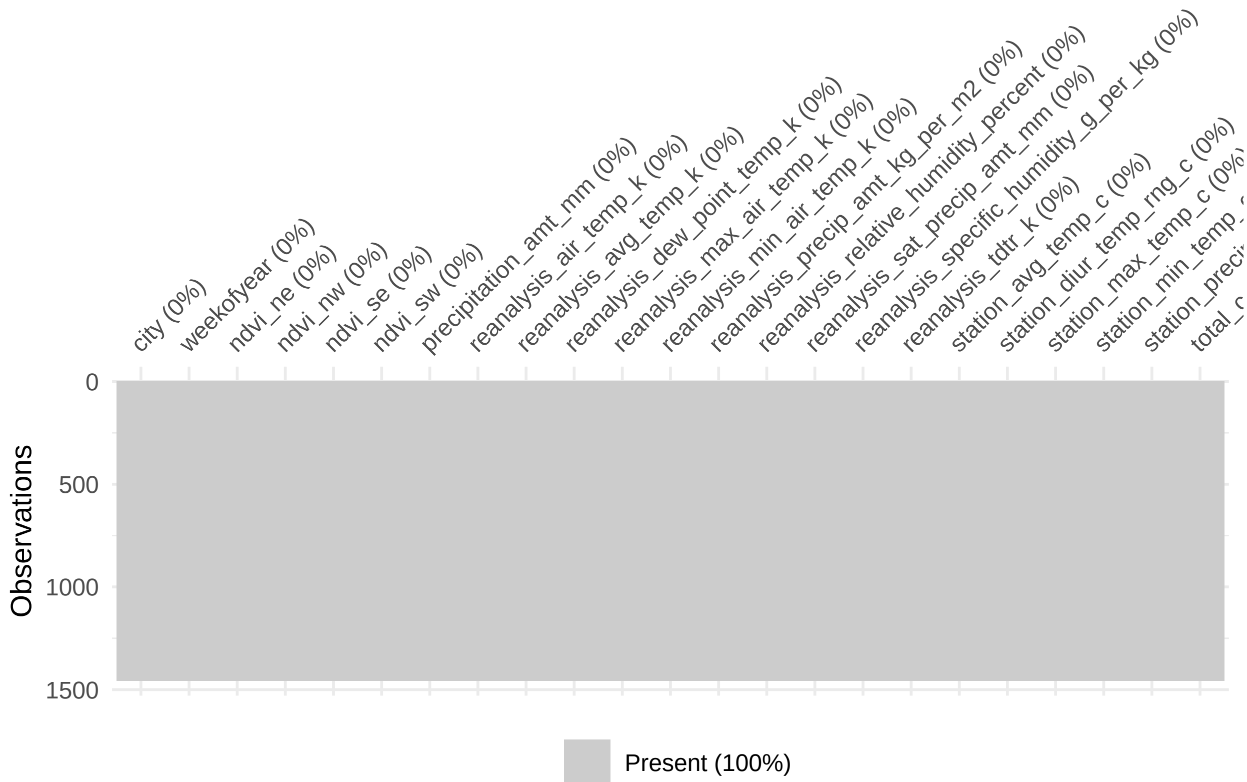
Imputation of missing values

```
pre.process <- preProcess(dengue_train, method = "bagImpute")
imputed.data <- predict(pre.process, dengue_train)
dengue_train$ndvi_ne <- imputed.data[,3]
dengue_train$ndvi_nw <- imputed.data[,4]
dengue_train$ndvi_se <- imputed.data[,5]
dengue_train$ndvi_sw <- imputed.data[,6]
dengue_train$precipitation_amt_mm <- imputed.data[,7]
dengue_train$reanalysis_air_temp_k <- imputed.data[, 8]
dengue_train$reanalysis_avg_temp_k <- imputed.data[,9]
dengue_train$reanalysis_dew_point_temp_k <- imputed.data[,10]
dengue_train$reanalysis_max_air_temp_k <- imputed.data[,11]
dengue_train$reanalysis_min_air_temp_k <- imputed.data[,12]
dengue_train$reanalysis_precip_amt_kg_per_m2 <- imputed.data[,13]
dengue_train$reanalysis_relative_humidity_percent <- imputed.data[,14]
dengue_train$reanalysis_sat_precip_amt_mm <- imputed.data[,15]
dengue_train$reanalysis_specific_humidity_g_per_kg <- imputed.data[,16]
dengue_train$reanalysis_tdtr_k <- imputed.data[,17]
dengue_train$station_avg_temp_c <- imputed.data[,18]
dengue_train$station_diur_temp_rng_c <- imputed.data[,19]
dengue_train$station_max_temp_c <- imputed.data[,20]
dengue_train$station_min_temp_c <- imputed.data[,21]
```

```r
dengue_train$station_precip_mm <- imputed.data[,22]

anyNA(dengue_train)
```

```
## [1] FALSE
```

```r
vis_miss(dengue_train)
```



Conclusion: All of the missing values were imputed.

Randomize the training data

```r
random_index <- sample(1:nrow(dengue_train), nrow(dengue_train))
random_train <- dengue_train[random_index, ]
names(random_train)
```

```
##  [1] "city"
##  [2] "weekofyear"
##  [3] "ndvi_ne"
##  [4] "ndvi_nw"
##  [5] "ndvi_se"
##  [6] "ndvi_sw"
##  [7] "precipitation_amt_mm"
##  [8] "reanalysis_air_temp_k"
##  [9] "reanalysis_avg_temp_k"
```

```
## [10] "reanalysis_dew_point_temp_k"
## [11] "reanalysis_max_air_temp_k"
## [12] "reanalysis_min_air_temp_k"
## [13] "reanalysis_precip_amt_kg_per_m2"
## [14] "reanalysis_relative_humidity_percent"
## [15] "reanalysis_sat_precip_amt_mm"
## [16] "reanalysis_specific_humidity_g_per_kg"
## [17] "reanalysis_tdtr_k"
## [18] "station_avg_temp_c"
## [19] "station_diur_temp_rng_c"
## [20] "station_max_temp_c"
## [21] "station_min_temp_c"
## [22] "station_precip_mm"
## [23] "total_cases"
```

```
dim(random_train)
```

```
## [1] 1456   23
```

```
anyNA(random_train)
```

```
## [1] FALSE
```

Defining the tuning grid

```
grid <- expand.grid(eta = c(0.05, 0.5),
                         nrounds = c(70, 90),
                         max_depth = 1:6,
                         min_child_weight = c(1.0, 4),
                         colsample_bytree = c(0.5, 1),
                         gamma = c(0, 0.1),
                         subsample = c(0.8, 1))
```

Defining trainControl for the ML Algorithms

```
train.control <- trainControl(method = "repeatedcv",
                              number = 10,
                              repeats = 5,
                              search = "grid")
```

#Applying ML Algorithms For Training the Prediction Model

K-Nearest Neighbor (knn) Algorithm to Train The Prediction Model

```
set.seed(45220)
model_kknn <- caret::train(total_cases ~ .,
                           data = random_train[,c(2,3:6,12,16,20,23)],
                           type="prob",
                           method = "kknn",
                           tuneLength = 10,
                           preProcess = NULL,
```

```
                           trControl = train.control)
model_kknn
```

```
## k-Nearest Neighbors
##
## 1456 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1310, 1309, 1310, 1311, 1311, 1311, ...
## Resampling results across tuning parameters:
##
##   kmax  RMSE      Rsquared   MAE
##    5    32.08063  0.4449575  17.70653
##    7    31.72765  0.4490605  17.61308
##    9    31.84266  0.4471819  17.61459
##   11    31.95553  0.4452667  17.63847
##   13    31.97997  0.4447213  17.64092
##   15    32.01345  0.4444422  17.64742
##   17    32.02947  0.4439074  17.64962
##   19    32.04736  0.4435655  17.65310
##   21    32.07460  0.4430591  17.66140
##   23    32.07171  0.4433770  17.65868
##
## Tuning parameter 'distance' was held constant at a value of 2
##
## Tuning parameter 'kernel' was held constant at a value of optimal
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were kmax = 7, distance = 2 and
##  kernel = optimal.
```

GLMNET Algorithm to Train The Prediction Model: generalized linear model via penalized maximum likelihood; the regulaization path is computed for elasticnet penalty at a grid of values for the regularization parameter lambada

```
set.seed(45220)
model_glmnet <- caret::train(total_cases ~ .,
                             data = random_train[,c(2,3:6,12,16,20,23)],
                             method = "glmnet",
                             preProcess = NULL,
                             trControl = train.control)
model_glmnet
```

```
## glmnet
##
## 1456 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1310, 1309, 1310, 1311, 1311, 1311, ...
```

```
## Resampling results across tuning parameters:
##
##   alpha  lambda      RMSE      Rsquared   MAE
##   0.10   0.02843102  39.26132  0.1500108  20.93842
##   0.10   0.28431018  39.25446  0.1501951  20.90350
##   0.10   2.84310180  39.20494  0.1517795  20.59967
##   0.55   0.02843102  39.26191  0.1500124  20.94252
##   0.55   0.28431018  39.23127  0.1510429  20.81516
##   0.55   2.84310180  39.22430  0.1519902  20.40782
##   1.00   0.02843102  39.25970  0.1500870  20.93313
##   1.00   0.28431018  39.21615  0.1515980  20.74542
##   1.00   2.84310180  39.31499  0.1506882  20.32348
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0.1 and lambda = 2.843102.
```

Random Forest Algorithm to Train The Prediction Model

```
x <- random_train[,2:22]

metric <- "MAE"
mtry <- sqrt(ncol(x))
model_rf <- caret::train(total_cases ~ .,
                         data = random_train[,c(2,3:6,12,16,20,23)],
                         method = "rf",
                         preProcess = NULL,
                         metric = metric,
                         tuneGrid = expand.grid(.mtry = mtry),
                         trControl = train.control)
model_rf
```

```
## Random Forest
##
## 1456 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1310, 1310, 1309, 1311, 1311, 1310, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   30.52333  0.4839504  16.83161
##
## Tuning parameter 'mtry' was held constant at a value of 4.582576
```

Regression Tree Algorithm to Train The Prediction Model

```
set.seed(123)
model_rpart <- caret::train(total_cases ~ ., data = random_train,
                            method = "rpart",
```

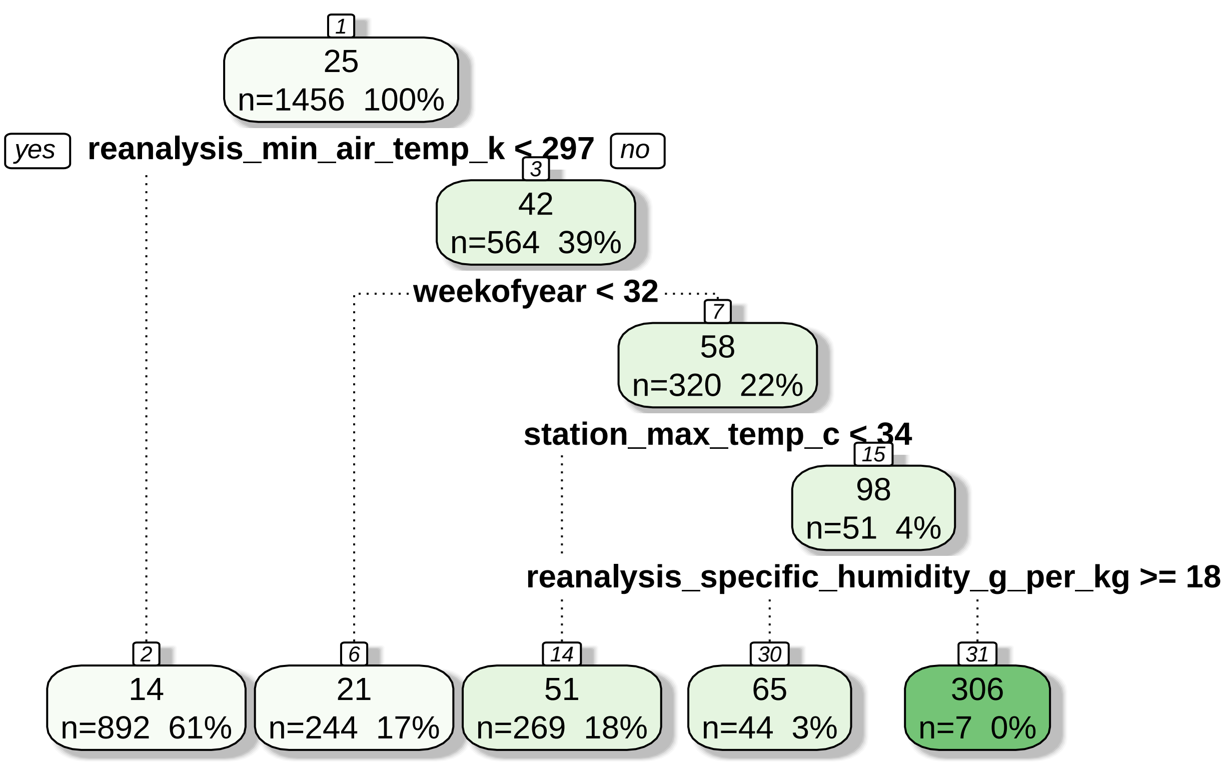
```r
                     preProcess = NULL,
                     trControl = train.control)
```

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

```r
model_rpart
```

```
## CART
##
## 1456 samples
##   22 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1311, 1310, 1312, 1310, 1310, 1309, ...
## Resampling results across tuning parameters:
##
##   cp          RMSE      Rsquared    MAE
##   0.02426438  38.40981  0.22829163  19.70701
##   0.07763399  40.63197  0.13067041  20.48623
##   0.10154375  41.88394  0.07935057  22.35448
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was cp = 0.02426438.
```

```r
fancyRpartPlot(model_rpart$finalModel)
```



Rattle 2020−Apr−07 10:24:31 Jesus

Partial Least Squares (PLS) to Train The Prediction Model

```
set.seed(27)
model_pls <- caret::train(total_cases ~ .,
                          data = random_train[,c(2,3:6,12,16,20,23)],
                          method = "pls",
                          preProcess = NULL,
                          trControl = train.control)
model_pls
```

```
## Partial Least Squares
##
## 1456 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1312, 1310, 1312, 1310, 1308, 1309, ...
## Resampling results across tuning parameters:
##
##   ncomp  RMSE      Rsquared    MAE
##   1      41.20706  0.05916277  22.39654
##   2      39.43403  0.14559185  20.74194
##   3      39.43646  0.14565291  20.85513
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was ncomp = 2.
```

Extreme Gradient Boosting

```
cl <- makeCluster(3, type = "SOCK")
registerDoSNOW(cl)

model_xgb <- caret::train(total_cases ~ .,
                          data = random_train[,c(2,3:6,12,16,20,23)],
                          method = "xgbTree",
                          tuneGrid = grid,
                          trControl = train.control)


model_xgb
```

```
## eXtreme Gradient Boosting
##
## 1456 samples
##    8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1312, 1310, 1311, 1311, 1310, 1311, ...
## Resampling results across tuning parameters:
##
```

```
##   eta   max_depth  gamma  colsample_bytree  min_child_weight  subsample
##   0.05  1          0.0    0.5               1                 0.8
##   0.05  1          0.0    0.5               1                 0.8
##   0.05  1          0.0    0.5               1                 1.0
##   0.05  1          0.0    0.5               1                 1.0
##   0.05  1          0.0    0.5               4                 0.8
##   0.05  1          0.0    0.5               4                 0.8
##   0.05  1          0.0    0.5               4                 1.0
##   0.05  1          0.0    0.5               4                 1.0
##   0.05  1          0.0    1.0               1                 0.8
##   0.05  1          0.0    1.0               1                 0.8
##   0.05  1          0.0    1.0               1                 1.0
##   0.05  1          0.0    1.0               1                 1.0
##   0.05  1          0.0    1.0               4                 0.8
##   0.05  1          0.0    1.0               4                 0.8
##   0.05  1          0.0    1.0               4                 1.0
##   0.05  1          0.0    1.0               4                 1.0
##   0.05  1          0.1    0.5               1                 0.8
##   0.05  1          0.1    0.5               1                 0.8
##   0.05  1          0.1    0.5               1                 1.0
##   0.05  1          0.1    0.5               1                 1.0
##   0.05  1          0.1    0.5               4                 0.8
##   0.05  1          0.1    0.5               4                 0.8
##   0.05  1          0.1    0.5               4                 1.0
##   0.05  1          0.1    0.5               4                 1.0
##   0.05  1          0.1    1.0               1                 0.8
##   0.05  1          0.1    1.0               1                 0.8
##   0.05  1          0.1    1.0               1                 1.0
##   0.05  1          0.1    1.0               1                 1.0
##   0.05  1          0.1    1.0               4                 0.8
##   0.05  1          0.1    1.0               4                 0.8
##   0.05  1          0.1    1.0               4                 1.0
##   0.05  1          0.1    1.0               4                 1.0
##   0.05  2          0.0    0.5               1                 0.8
##   0.05  2          0.0    0.5               1                 0.8
##   0.05  2          0.0    0.5               1                 1.0
##   0.05  2          0.0    0.5               1                 1.0
##   0.05  2          0.0    0.5               4                 0.8
##   0.05  2          0.0    0.5               4                 0.8
##   0.05  2          0.0    0.5               4                 1.0
##   0.05  2          0.0    0.5               4                 1.0
##   0.05  2          0.0    1.0               1                 0.8
##   0.05  2          0.0    1.0               1                 0.8
##   0.05  2          0.0    1.0               1                 1.0
##   0.05  2          0.0    1.0               1                 1.0
##   0.05  2          0.0    1.0               4                 0.8
##   0.05  2          0.0    1.0               4                 0.8
##   0.05  2          0.0    1.0               4                 1.0
##   0.05  2          0.0    1.0               4                 1.0
##   0.05  2          0.1    0.5               1                 0.8
##   0.05  2          0.1    0.5               1                 0.8
##   0.05  2          0.1    0.5               1                 1.0
##   0.05  2          0.1    0.5               1                 1.0
##   0.05  2          0.1    0.5               4                 0.8
```

```
##   0.05  2         0.1    0.5          4          0.8
##   0.05  2         0.1    0.5          4          1.0
##   0.05  2         0.1    0.5          4          1.0
##   0.05  2         0.1    1.0          1          0.8
##   0.05  2         0.1    1.0          1          0.8
##   0.05  2         0.1    1.0          1          1.0
##   0.05  2         0.1    1.0          1          1.0
##   0.05  2         0.1    1.0          4          0.8
##   0.05  2         0.1    1.0          4          0.8
##   0.05  2         0.1    1.0          4          1.0
##   0.05  2         0.1    1.0          4          1.0
##   0.05  3         0.0    0.5          1          0.8
##   0.05  3         0.0    0.5          1          0.8
##   0.05  3         0.0    0.5          1          1.0
##   0.05  3         0.0    0.5          1          1.0
##   0.05  3         0.0    0.5          4          0.8
##   0.05  3         0.0    0.5          4          0.8
##   0.05  3         0.0    0.5          4          1.0
##   0.05  3         0.0    0.5          4          1.0
##   0.05  3         0.0    1.0          1          0.8
##   0.05  3         0.0    1.0          1          0.8
##   0.05  3         0.0    1.0          1          1.0
##   0.05  3         0.0    1.0          1          1.0
##   0.05  3         0.0    1.0          4          0.8
##   0.05  3         0.0    1.0          4          0.8
##   0.05  3         0.0    1.0          4          1.0
##   0.05  3         0.0    1.0          4          1.0
##   0.05  3         0.1    0.5          1          0.8
##   0.05  3         0.1    0.5          1          0.8
##   0.05  3         0.1    0.5          1          1.0
##   0.05  3         0.1    0.5          1          1.0
##   0.05  3         0.1    0.5          4          0.8
##   0.05  3         0.1    0.5          4          0.8
##   0.05  3         0.1    0.5          4          1.0
##   0.05  3         0.1    0.5          4          1.0
##   0.05  3         0.1    1.0          1          0.8
##   0.05  3         0.1    1.0          1          0.8
##   0.05  3         0.1    1.0          1          1.0
##   0.05  3         0.1    1.0          1          1.0
##   0.05  3         0.1    1.0          4          0.8
##   0.05  3         0.1    1.0          4          0.8
##   0.05  3         0.1    1.0          4          1.0
##   0.05  3         0.1    1.0          4          1.0
##   0.05  4         0.0    0.5          1          0.8
##   0.05  4         0.0    0.5          1          0.8
##   0.05  4         0.0    0.5          1          1.0
##   0.05  4         0.0    0.5          1          1.0
##   0.05  4         0.0    0.5          4          0.8
##   0.05  4         0.0    0.5          4          0.8
##   0.05  4         0.0    0.5          4          1.0
##   0.05  4         0.0    0.5          4          1.0
##   0.05  4         0.0    1.0          1          0.8
##   0.05  4         0.0    1.0          1          0.8
##   0.05  4         0.0    1.0          1          1.0
```

```
##   0.05  4        0.0   1.0            1              1.0
##   0.05  4        0.0   1.0            4              0.8
##   0.05  4        0.0   1.0            4              0.8
##   0.05  4        0.0   1.0            4              1.0
##   0.05  4        0.0   1.0            4              1.0
##   0.05  4        0.1   0.5            1              0.8
##   0.05  4        0.1   0.5            1              0.8
##   0.05  4        0.1   0.5            1              1.0
##   0.05  4        0.1   0.5            1              1.0
##   0.05  4        0.1   0.5            4              0.8
##   0.05  4        0.1   0.5            4              0.8
##   0.05  4        0.1   0.5            4              1.0
##   0.05  4        0.1   0.5            4              1.0
##   0.05  4        0.1   1.0            1              0.8
##   0.05  4        0.1   1.0            1              0.8
##   0.05  4        0.1   1.0            1              1.0
##   0.05  4        0.1   1.0            1              1.0
##   0.05  4        0.1   1.0            4              0.8
##   0.05  4        0.1   1.0            4              0.8
##   0.05  4        0.1   1.0            4              1.0
##   0.05  4        0.1   1.0            4              1.0
##   0.05  5        0.0   0.5            1              0.8
##   0.05  5        0.0   0.5            1              0.8
##   0.05  5        0.0   0.5            1              1.0
##   0.05  5        0.0   0.5            1              1.0
##   0.05  5        0.0   0.5            4              0.8
##   0.05  5        0.0   0.5            4              0.8
##   0.05  5        0.0   0.5            4              1.0
##   0.05  5        0.0   0.5            4              1.0
##   0.05  5        0.0   1.0            1              0.8
##   0.05  5        0.0   1.0            1              0.8
##   0.05  5        0.0   1.0            1              1.0
##   0.05  5        0.0   1.0            1              1.0
##   0.05  5        0.0   1.0            4              0.8
##   0.05  5        0.0   1.0            4              0.8
##   0.05  5        0.0   1.0            4              1.0
##   0.05  5        0.0   1.0            4              1.0
##   0.05  5        0.1   0.5            1              0.8
##   0.05  5        0.1   0.5            1              0.8
##   0.05  5        0.1   0.5            1              1.0
##   0.05  5        0.1   0.5            1              1.0
##   0.05  5        0.1   0.5            4              0.8
##   0.05  5        0.1   0.5            4              0.8
##   0.05  5        0.1   0.5            4              1.0
##   0.05  5        0.1   0.5            4              1.0
##   0.05  5        0.1   1.0            1              0.8
##   0.05  5        0.1   1.0            1              0.8
##   0.05  5        0.1   1.0            1              1.0
##   0.05  5        0.1   1.0            1              1.0
##   0.05  5        0.1   1.0            4              0.8
##   0.05  5        0.1   1.0            4              0.8
##   0.05  5        0.1   1.0            4              1.0
##   0.05  5        0.1   1.0            4              1.0
##   0.05  6        0.0   0.5            1              0.8
```

```
##   0.05  6          0.0   0.5              1                 0.8
##   0.05  6          0.0   0.5              1                 1.0
##   0.05  6          0.0   0.5              1                 1.0
##   0.05  6          0.0   0.5              4                 0.8
##   0.05  6          0.0   0.5              4                 0.8
##   0.05  6          0.0   0.5              4                 1.0
##   0.05  6          0.0   0.5              4                 1.0
##   0.05  6          0.0   1.0              1                 0.8
##   0.05  6          0.0   1.0              1                 0.8
##   0.05  6          0.0   1.0              1                 1.0
##   0.05  6          0.0   1.0              1                 1.0
##   0.05  6          0.0   1.0              4                 0.8
##   0.05  6          0.0   1.0              4                 0.8
##   0.05  6          0.0   1.0              4                 1.0
##   0.05  6          0.0   1.0              4                 1.0
##   0.05  6          0.1   0.5              1                 0.8
##   0.05  6          0.1   0.5              1                 0.8
##   0.05  6          0.1   0.5              1                 1.0
##   0.05  6          0.1   0.5              1                 1.0
##   0.05  6          0.1   0.5              4                 0.8
##   0.05  6          0.1   0.5              4                 0.8
##   0.05  6          0.1   0.5              4                 1.0
##   0.05  6          0.1   0.5              4                 1.0
##   0.05  6          0.1   1.0              1                 0.8
##   0.05  6          0.1   1.0              1                 0.8
##   0.05  6          0.1   1.0              1                 1.0
##   0.05  6          0.1   1.0              1                 1.0
##   0.05  6          0.1   1.0              4                 0.8
##   0.05  6          0.1   1.0              4                 0.8
##   0.05  6          0.1   1.0              4                 1.0
##   0.05  6          0.1   1.0              4                 1.0
##   0.50  1          0.0   0.5              1                 0.8
##   0.50  1          0.0   0.5              1                 0.8
##   0.50  1          0.0   0.5              1                 1.0
##   0.50  1          0.0   0.5              1                 1.0
##   0.50  1          0.0   0.5              4                 0.8
##   0.50  1          0.0   0.5              4                 0.8
##   0.50  1          0.0   0.5              4                 1.0
##   0.50  1          0.0   0.5              4                 1.0
##   0.50  1          0.0   1.0              1                 0.8
##   0.50  1          0.0   1.0              1                 0.8
##   0.50  1          0.0   1.0              1                 1.0
##   0.50  1          0.0   1.0              1                 1.0
##   0.50  1          0.0   1.0              4                 0.8
##   0.50  1          0.0   1.0              4                 0.8
##   0.50  1          0.0   1.0              4                 1.0
##   0.50  1          0.0   1.0              4                 1.0
##   0.50  1          0.1   0.5              1                 0.8
##   0.50  1          0.1   0.5              1                 0.8
##   0.50  1          0.1   0.5              1                 1.0
##   0.50  1          0.1   0.5              1                 1.0
##   0.50  1          0.1   0.5              4                 0.8
##   0.50  1          0.1   0.5              4                 0.8
##   0.50  1          0.1   0.5              4                 1.0
```

```
## 0.50 1 0.1 0.5 4 1.0
## 0.50 1 0.1 1.0 1 0.8
## 0.50 1 0.1 1.0 1 0.8
## 0.50 1 0.1 1.0 1 1.0
## 0.50 1 0.1 1.0 1 1.0
## 0.50 1 0.1 1.0 4 0.8
## 0.50 1 0.1 1.0 4 0.8
## 0.50 1 0.1 1.0 4 1.0
## 0.50 1 0.1 1.0 4 1.0
## 0.50 2 0.0 0.5 1 0.8
## 0.50 2 0.0 0.5 1 0.8
## 0.50 2 0.0 0.5 1 1.0
## 0.50 2 0.0 0.5 1 1.0
## 0.50 2 0.0 0.5 4 0.8
## 0.50 2 0.0 0.5 4 0.8
## 0.50 2 0.0 0.5 4 1.0
## 0.50 2 0.0 0.5 4 1.0
## 0.50 2 0.0 1.0 1 0.8
## 0.50 2 0.0 1.0 1 0.8
## 0.50 2 0.0 1.0 1 1.0
## 0.50 2 0.0 1.0 1 1.0
## 0.50 2 0.0 1.0 4 0.8
## 0.50 2 0.0 1.0 4 0.8
## 0.50 2 0.0 1.0 4 1.0
## 0.50 2 0.0 1.0 4 1.0
## 0.50 2 0.1 0.5 1 0.8
## 0.50 2 0.1 0.5 1 0.8
## 0.50 2 0.1 0.5 1 1.0
## 0.50 2 0.1 0.5 1 1.0
## 0.50 2 0.1 0.5 4 0.8
## 0.50 2 0.1 0.5 4 0.8
## 0.50 2 0.1 0.5 4 1.0
## 0.50 2 0.1 0.5 4 1.0
## 0.50 2 0.1 1.0 1 0.8
## 0.50 2 0.1 1.0 1 0.8
## 0.50 2 0.1 1.0 1 1.0
## 0.50 2 0.1 1.0 1 1.0
## 0.50 2 0.1 1.0 4 0.8
## 0.50 2 0.1 1.0 4 0.8
## 0.50 2 0.1 1.0 4 1.0
## 0.50 2 0.1 1.0 4 1.0
## 0.50 3 0.0 0.5 1 0.8
## 0.50 3 0.0 0.5 1 0.8
## 0.50 3 0.0 0.5 1 1.0
## 0.50 3 0.0 0.5 1 1.0
## 0.50 3 0.0 0.5 4 0.8
## 0.50 3 0.0 0.5 4 0.8
## 0.50 3 0.0 0.5 4 1.0
## 0.50 3 0.0 0.5 4 1.0
## 0.50 3 0.0 1.0 1 0.8
## 0.50 3 0.0 1.0 1 0.8
## 0.50 3 0.0 1.0 1 1.0
## 0.50 3 0.0 1.0 1 1.0
## 0.50 3 0.0 1.0 4 0.8
```

```
##   0.50  3          0.0   1.0                4                 0.8
##   0.50  3          0.0   1.0                4                 1.0
##   0.50  3          0.0   1.0                4                 1.0
##   0.50  3          0.1   0.5                1                 0.8
##   0.50  3          0.1   0.5                1                 0.8
##   0.50  3          0.1   0.5                1                 1.0
##   0.50  3          0.1   0.5                1                 1.0
##   0.50  3          0.1   0.5                4                 0.8
##   0.50  3          0.1   0.5                4                 0.8
##   0.50  3          0.1   0.5                4                 1.0
##   0.50  3          0.1   0.5                4                 1.0
##   0.50  3          0.1   1.0                1                 0.8
##   0.50  3          0.1   1.0                1                 0.8
##   0.50  3          0.1   1.0                1                 1.0
##   0.50  3          0.1   1.0                1                 1.0
##   0.50  3          0.1   1.0                4                 0.8
##   0.50  3          0.1   1.0                4                 0.8
##   0.50  3          0.1   1.0                4                 1.0
##   0.50  3          0.1   1.0                4                 1.0
##   0.50  4          0.0   0.5                1                 0.8
##   0.50  4          0.0   0.5                1                 0.8
##   0.50  4          0.0   0.5                1                 1.0
##   0.50  4          0.0   0.5                1                 1.0
##   0.50  4          0.0   0.5                4                 0.8
##   0.50  4          0.0   0.5                4                 0.8
##   0.50  4          0.0   0.5                4                 1.0
##   0.50  4          0.0   0.5                4                 1.0
##   0.50  4          0.0   1.0                1                 0.8
##   0.50  4          0.0   1.0                1                 0.8
##   0.50  4          0.0   1.0                1                 1.0
##   0.50  4          0.0   1.0                1                 1.0
##   0.50  4          0.0   1.0                4                 0.8
##   0.50  4          0.0   1.0                4                 0.8
##   0.50  4          0.0   1.0                4                 1.0
##   0.50  4          0.0   1.0                4                 1.0
##   0.50  4          0.1   0.5                1                 0.8
##   0.50  4          0.1   0.5                1                 0.8
##   0.50  4          0.1   0.5                1                 1.0
##   0.50  4          0.1   0.5                1                 1.0
##   0.50  4          0.1   0.5                4                 0.8
##   0.50  4          0.1   0.5                4                 0.8
##   0.50  4          0.1   0.5                4                 1.0
##   0.50  4          0.1   0.5                4                 1.0
##   0.50  4          0.1   1.0                1                 0.8
##   0.50  4          0.1   1.0                1                 0.8
##   0.50  4          0.1   1.0                1                 1.0
##   0.50  4          0.1   1.0                1                 1.0
##   0.50  4          0.1   1.0                4                 0.8
##   0.50  4          0.1   1.0                4                 0.8
##   0.50  4          0.1   1.0                4                 1.0
##   0.50  4          0.1   1.0                4                 1.0
##   0.50  5          0.0   0.5                1                 0.8
##   0.50  5          0.0   0.5                1                 0.8
##   0.50  5          0.0   0.5                1                 1.0
```

```
## 0.50 5 0.0 0.5 1 1.0
## 0.50 5 0.0 0.5 4 0.8
## 0.50 5 0.0 0.5 4 0.8
## 0.50 5 0.0 0.5 4 1.0
## 0.50 5 0.0 0.5 4 1.0
## 0.50 5 0.0 1.0 1 0.8
## 0.50 5 0.0 1.0 1 0.8
## 0.50 5 0.0 1.0 1 1.0
## 0.50 5 0.0 1.0 1 1.0
## 0.50 5 0.0 1.0 4 0.8
## 0.50 5 0.0 1.0 4 0.8
## 0.50 5 0.0 1.0 4 1.0
## 0.50 5 0.0 1.0 4 1.0
## 0.50 5 0.1 0.5 1 0.8
## 0.50 5 0.1 0.5 1 0.8
## 0.50 5 0.1 0.5 1 1.0
## 0.50 5 0.1 0.5 1 1.0
## 0.50 5 0.1 0.5 4 0.8
## 0.50 5 0.1 0.5 4 0.8
## 0.50 5 0.1 0.5 4 1.0
## 0.50 5 0.1 0.5 4 1.0
## 0.50 5 0.1 1.0 1 0.8
## 0.50 5 0.1 1.0 1 0.8
## 0.50 5 0.1 1.0 1 1.0
## 0.50 5 0.1 1.0 1 1.0
## 0.50 5 0.1 1.0 4 0.8
## 0.50 5 0.1 1.0 4 0.8
## 0.50 5 0.1 1.0 4 1.0
## 0.50 5 0.1 1.0 4 1.0
## 0.50 6 0.0 0.5 1 0.8
## 0.50 6 0.0 0.5 1 0.8
## 0.50 6 0.0 0.5 1 1.0
## 0.50 6 0.0 0.5 1 1.0
## 0.50 6 0.0 0.5 4 0.8
## 0.50 6 0.0 0.5 4 0.8
## 0.50 6 0.0 0.5 4 1.0
## 0.50 6 0.0 0.5 4 1.0
## 0.50 6 0.0 1.0 1 0.8
## 0.50 6 0.0 1.0 1 0.8
## 0.50 6 0.0 1.0 1 1.0
## 0.50 6 0.0 1.0 1 1.0
## 0.50 6 0.0 1.0 4 0.8
## 0.50 6 0.0 1.0 4 0.8
## 0.50 6 0.0 1.0 4 1.0
## 0.50 6 0.0 1.0 4 1.0
## 0.50 6 0.1 0.5 1 0.8
## 0.50 6 0.1 0.5 1 0.8
## 0.50 6 0.1 0.5 1 1.0
## 0.50 6 0.1 0.5 1 1.0
## 0.50 6 0.1 0.5 4 0.8
## 0.50 6 0.1 0.5 4 0.8
## 0.50 6 0.1 0.5 4 1.0
## 0.50 6 0.1 0.5 4 1.0
## 0.50 6 0.1 1.0 1 0.8
```

```
##   0.50  6          0.1    1.0              1                 0.8
##   0.50  6          0.1    1.0              1                 1.0
##   0.50  6          0.1    1.0              1                 1.0
##   0.50  6          0.1    1.0              4                 0.8
##   0.50  6          0.1    1.0              4                 0.8
##   0.50  6          0.1    1.0              4                 1.0
##   0.50  6          0.1    1.0              4                 1.0
##   nrounds  RMSE      Rsquared   MAE
##   70       38.93380  0.1811865  19.65144
##   90       38.75062  0.1884744  19.71423
##   70       39.00722  0.1778792  19.68096
##   90       38.86030  0.1834527  19.79050
##   70       38.93968  0.1804497  19.64199
##   90       38.73566  0.1890442  19.70852
##   70       39.00119  0.1781800  19.68745
##   90       38.85714  0.1837445  19.79708
##   70       38.91931  0.1815206  19.62425
##   90       38.71840  0.1901745  19.68537
##   70       39.00397  0.1784542  19.68371
##   90       38.86922  0.1833651  19.80421
##   70       38.92391  0.1811875  19.64981
##   90       38.73350  0.1890720  19.71377
##   70       39.00397  0.1784542  19.68371
##   90       38.86922  0.1833651  19.80421
##   70       38.92781  0.1815651  19.65428
##   90       38.73998  0.1893634  19.70127
##   70       39.00390  0.1780704  19.68443
##   90       38.85916  0.1837981  19.79145
##   70       38.94398  0.1806536  19.64717
##   90       38.74769  0.1886212  19.71482
##   70       39.00367  0.1780555  19.68086
##   90       38.85777  0.1836602  19.78702
##   70       38.90915  0.1817254  19.63188
##   90       38.70962  0.1902086  19.68265
##   70       39.00397  0.1784542  19.68371
##   90       38.86922  0.1833651  19.80421
##   70       38.92868  0.1808875  19.61471
##   90       38.74498  0.1887385  19.68799
##   70       39.00397  0.1784542  19.68371
##   90       38.86922  0.1833651  19.80421
##   70       35.84972  0.3235034  17.96664
##   90       34.54454  0.3705581  17.88460
##   70       35.56104  0.3361276  17.96228
##   90       34.03300  0.3879934  17.86556
##   70       35.79299  0.3264867  17.95570
##   90       34.41189  0.3772544  17.87272
##   70       35.50927  0.3383005  17.96661
##   90       33.82476  0.3941956  17.81327
##   70       35.65256  0.3285653  17.95976
##   90       34.42660  0.3714123  17.91153
##   70       35.48176  0.3379096  17.97487
##   90       34.04353  0.3854159  17.86416
##   70       35.60412  0.3332009  17.87980
##   90       34.33363  0.3775531  17.85228
```

```
## 70 35.44613 0.3398192 17.95465
## 90 33.95001 0.3891928 17.83760
## 70 35.82011 0.3242790 17.92563
## 90 34.53060 0.3702405 17.87523
## 70 35.70765 0.3304450 17.98320
## 90 33.91057 0.3924805 17.82293
## 70 35.78347 0.3269908 17.94749
## 90 34.38717 0.3778985 17.81983
## 70 35.56691 0.3369841 17.99790
## 90 33.89499 0.3936127 17.87117
## 70 35.47190 0.3371538 17.85537
## 90 34.27929 0.3786043 17.81701
## 70 35.48176 0.3379096 17.97487
## 90 34.04353 0.3854159 17.86416
## 70 35.50708 0.3371729 17.86153
## 90 34.36183 0.3765412 17.83044
## 70 35.44613 0.3398192 17.95465
## 90 33.95001 0.3891928 17.83760
## 70 32.41080 0.4398004 17.17372
## 90 31.54636 0.4582108 17.13542
## 70 32.36896 0.4427821 17.13233
## 90 31.59248 0.4585362 17.13127
## 70 32.17453 0.4466124 17.08531
## 90 31.38715 0.4626057 17.08961
## 70 32.29724 0.4427478 17.12071
## 90 31.50800 0.4594526 17.11193
## 70 32.40625 0.4386393 17.01008
## 90 31.50130 0.4603295 16.95997
## 70 32.66315 0.4271446 17.14496
## 90 31.78915 0.4500060 17.09879
## 70 32.23860 0.4435239 17.00490
## 90 31.38326 0.4634397 16.99090
## 70 32.61626 0.4289618 17.17503
## 90 31.75532 0.4501522 17.14432
## 70 32.26378 0.4459287 17.08666
## 90 31.44187 0.4636288 17.07080
## 70 32.32827 0.4422645 17.13189
## 90 31.49046 0.4602367 17.09333
## 70 31.99076 0.4574056 17.04988
## 90 31.28202 0.4695641 17.08201
## 70 32.27846 0.4409741 17.11747
## 90 31.40688 0.4593071 17.06677
## 70 32.39534 0.4367507 17.04653
## 90 31.54369 0.4574322 17.01059
## 70 32.66315 0.4271446 17.14496
## 90 31.78915 0.4500060 17.09879
## 70 32.34883 0.4381627 17.03074
## 90 31.43616 0.4600881 16.98290
## 70 32.61626 0.4289618 17.17503
## 90 31.75532 0.4501522 17.14432
## 70 30.72661 0.4830078 16.72953
## 90 30.17882 0.4900561 16.74778
## 70 30.72395 0.4800030 16.68122
## 90 30.23124 0.4889268 16.70962
```

```
## 70      30.59136 0.4815418 16.68365
## 90      30.11921 0.4882307 16.70035
## 70      30.74853 0.4757912 16.69117
## 90      30.19466 0.4855680 16.71121
## 70      30.41586 0.4853927 16.44326
## 90      29.91159 0.4948687 16.47957
## 70      30.00040 0.4904151 16.37641
## 90      29.53965 0.5012919 16.42901
## 70      30.06253 0.4970370 16.38223
## 90      29.52312 0.5065568 16.41140
## 70      29.87647 0.4932038 16.40944
## 90      29.44787 0.5024174 16.46693
## 70      30.73638 0.4792939 16.72448
## 90      30.19514 0.4887178 16.70334
## 70      30.95246 0.4729681 16.76501
## 90      30.45559 0.4816629 16.78971
## 70      30.61908 0.4845244 16.66176
## 90      30.07100 0.4927244 16.66709
## 70      30.56755 0.4824449 16.68460
## 90      30.09377 0.4900970 16.72591
## 70      30.25158 0.4921180 16.36409
## 90      29.72739 0.5013216 16.40979
## 70      30.00040 0.4904151 16.37641
## 90      29.53965 0.5012905 16.42835
## 70      30.14351 0.4934133 16.39469
## 90      29.58956 0.5040379 16.40149
## 70      29.87647 0.4932038 16.40944
## 90      29.44787 0.5024174 16.46693
## 70      29.50097 0.5082819 16.28136
## 90      29.18651 0.5099628 16.37007
## 70      29.83411 0.4982878 16.39700
## 90      29.59209 0.5005620 16.52407
## 70      29.71906 0.5002562 16.38785
## 90      29.38988 0.5033017 16.46059
## 70      29.91350 0.4928540 16.49860
## 90      29.60380 0.4967459 16.56496
## 70      29.60024 0.5021617 16.17512
## 90      29.25608 0.5075869 16.26582
## 70      29.57206 0.4987872 16.31189
## 90      29.29491 0.5053331 16.42451
## 70      29.55402 0.5042854 16.19260
## 90      29.25599 0.5076697 16.30515
## 70      29.40492 0.5009533 16.32415
## 90      29.13154 0.5066029 16.40671
## 70      29.56013 0.5113736 16.30482
## 90      29.23734 0.5132021 16.37972
## 70      29.83760 0.4958452 16.40741
## 90      29.59310 0.4980113 16.52002
## 70      29.63124 0.5039334 16.33061
## 90      29.26756 0.5068694 16.37611
## 70      29.96231 0.4903694 16.50984
## 90      29.62852 0.4949486 16.58379
## 70      29.56852 0.5049576 16.15761
## 90      29.18899 0.5112107 16.25278
```

```
## 70   29.57213 0.4987838 16.31189
## 90   29.29553 0.5053027 16.42526
## 70   29.39959 0.5077494 16.11165
## 90   29.07132 0.5128366 16.20769
## 70   29.40492 0.5009533 16.32415
## 90   29.13154 0.5066029 16.40671
## 70   29.26479 0.5107976 16.23977
## 90   29.02907 0.5118618 16.35765
## 70   29.38017 0.5048296 16.27899
## 90   29.18387 0.5053661 16.40903
## 70   29.38038 0.5051692 16.31876
## 90   29.11215 0.5073081 16.41483
## 70   29.25268 0.5085446 16.36441
## 90   29.07403 0.5084985 16.48799
## 70   29.07501 0.5130483 15.94604
## 90   28.81851 0.5163021 16.05299
## 70   29.88874 0.4860657 16.27820
## 90   29.72878 0.4902940 16.43458
## 70   28.85962 0.5172747 15.98583
## 90   28.71478 0.5179434 16.12768
## 70   29.46742 0.4969129 16.23030
## 90   29.31771 0.4993093 16.37965
## 70   29.27494 0.5114814 16.28642
## 90   29.04310 0.5121681 16.41456
## 70   29.45984 0.5041248 16.26302
## 90   29.27336 0.5043921 16.41240
## 70   29.22788 0.5113015 16.28249
## 90   28.96773 0.5129881 16.37434
## 70   29.07563 0.5114832 16.28838
## 90   28.91115 0.5108919 16.40839
## 70   29.28461 0.5076716 16.00617
## 90   29.05540 0.5111310 16.13832
## 70   29.88383 0.4861456 16.28020
## 90   29.69717 0.4909166 16.43553
## 70   29.05052 0.5119041 16.01769
## 90   28.83842 0.5144064 16.15154
## 70   29.46739 0.4969126 16.23033
## 90   29.31742 0.4992894 16.37951
## 70   35.98060 0.3015864 20.13145
## 90   35.43582 0.3210418 20.11454
## 70   35.65601 0.3155451 19.76704
## 90   35.06844 0.3370865 19.68742
## 70   35.99480 0.3004137 20.09721
## 90   35.48595 0.3197571 20.07503
## 70   35.67726 0.3149085 19.69424
## 90   35.07329 0.3366886 19.65049
## 70   35.47556 0.3203613 19.77163
## 90   34.97233 0.3389534 19.83753
## 70   35.04275 0.3442549 19.29808
## 90   34.36665 0.3668574 19.19739
## 70   35.50774 0.3186878 19.77951
## 90   34.90026 0.3411037 19.73372
## 70   35.04124 0.3443104 19.29092
## 90   34.37638 0.3661398 19.20255
```

```
## 70       35.93019 0.3021627 20.02438
## 90       35.49257 0.3185020 20.09216
## 70       35.65846 0.3166420 19.66582
## 90       35.03845 0.3391553 19.67980
## 70       35.88299 0.3029580 20.04783
## 90       35.37043 0.3217758 20.01334
## 70       35.62401 0.3170001 19.65158
## 90       35.00853 0.3398895 19.61109
## 70       35.51376 0.3205009 19.86105
## 90       34.91719 0.3420988 19.85246
## 70       35.04275 0.3442549 19.29808
## 90       34.36665 0.3668574 19.19739
## 70       35.59372 0.3154529 19.85094
## 90       34.99581 0.3373584 19.85453
## 70       35.04124 0.3443104 19.29092
## 90       34.37638 0.3661398 19.20255
## 70       32.34011 0.4183742 19.14101
## 90       32.37309 0.4177610 19.30704
## 70       31.82108 0.4352718 18.60892
## 90       31.93202 0.4333120 18.81224
## 70       32.41569 0.4203684 19.24165
## 90       32.55981 0.4171633 19.50413
## 70       31.78268 0.4333030 19.00007
## 90       31.97464 0.4291061 19.23146
## 70       32.33535 0.4195302 19.19326
## 90       32.50347 0.4170781 19.45657
## 70       31.55345 0.4371784 18.60847
## 90       31.63112 0.4365241 18.76661
## 70       32.01270 0.4305010 19.02135
## 90       32.21501 0.4264963 19.35959
## 70       31.85675 0.4318786 18.80310
## 90       31.88410 0.4320160 19.04337
## 70       32.01470 0.4304105 19.06212
## 90       32.22609 0.4275873 19.31758
## 70       31.70401 0.4359145 18.82903
## 90       31.86282 0.4338760 19.04207
## 70       32.25203 0.4238945 19.30807
## 90       32.33842 0.4233952 19.54238
## 70       32.24746 0.4229531 18.89919
## 90       32.50817 0.4166323 19.27577
## 70       32.61527 0.4156039 19.15539
## 90       32.83104 0.4131757 19.49814
## 70       31.55345 0.4371784 18.60847
## 90       31.63112 0.4365241 18.76661
## 70       32.30332 0.4241118 19.15850
## 90       32.46076 0.4215501 19.38086
## 70       31.85675 0.4318786 18.80310
## 90       31.88410 0.4320160 19.04337
## 70       32.96675 0.4177799 19.85073
## 90       33.28630 0.4114856 20.18791
## 70       32.35368 0.4222981 19.26682
## 90       32.49920 0.4210122 19.48663
## 70       33.29089 0.4005706 19.95211
## 90       33.47952 0.3971331 20.26343
```

```
## 70        32.15653  0.4287342  19.38228
## 90        32.39967  0.4243681  19.70438
## 70        32.68858  0.4219741  19.43899
## 90        32.89982  0.4184413  19.76925
## 70        31.48422  0.4467396  18.70313
## 90        31.66028  0.4439748  18.92441
## 70        32.55196  0.4267835  19.49737
## 90        32.74326  0.4250784  19.82013
## 70        32.09785  0.4313415  19.12810
## 90        32.30688  0.4286341  19.43530
## 70        32.99750  0.4087248  19.60822
## 90        33.07423  0.4096525  19.82749
## 70        32.51953  0.4125217  19.22148
## 90        32.56703  0.4127274  19.44190
## 70        32.59423  0.4225366  19.78123
## 90        32.81494  0.4191601  20.13688
## 70        32.23464  0.4273166  19.20146
## 90        32.44600  0.4229090  19.47061
## 70        32.74845  0.4233864  19.44138
## 90        32.94088  0.4203954  19.73249
## 70        31.48422  0.4467395  18.70314
## 90        31.66028  0.4439747  18.92442
## 70        32.34641  0.4360200  19.52222
## 90        32.70338  0.4293619  19.89032
## 70        32.09785  0.4313415  19.12810
## 90        32.30688  0.4286341  19.43530
## 70        33.69960  0.4022402  20.11107
## 90        33.87817  0.3985232  20.34104
## 70        32.44009  0.4201828  19.14083
## 90        32.61572  0.4161615  19.38832
## 70        33.79062  0.4015656  20.77950
## 90        33.93451  0.4000871  20.97369
## 70        32.53174  0.4209192  19.41026
## 90        32.73461  0.4184506  19.64508
## 70        32.99633  0.4241486  19.68822
## 90        33.09239  0.4226283  19.90933
## 70        31.43337  0.4456181  18.74513
## 90        31.49702  0.4444286  18.87851
## 70        33.31093  0.4105693  20.13628
## 90        33.67669  0.4038387  20.45039
## 70        31.51826  0.4434107  18.86351
## 90        31.61327  0.4426255  19.10036
## 70        34.04311  0.3938530  20.39246
## 90        34.07959  0.3948376  20.55346
## 70        32.11633  0.4282034  19.11877
## 90        32.21371  0.4268985  19.28714
## 70        33.58896  0.4059543  20.39236
## 90        33.76634  0.4031223  20.62969
## 70        33.00855  0.4032486  19.86942
## 90        33.17189  0.4011758  20.08813
## 70        32.80520  0.4170382  19.64462
## 90        32.89685  0.4160966  19.79776
## 70        31.44111  0.4455216  18.75737
## 90        31.52020  0.4441008  18.89260
```

```
## 70 33.20114 0.4077427 20.02558
## 90 33.36473 0.4049219 20.25950
## 70 31.51826 0.4434107 18.86351
## 90 31.61468 0.4425241 19.10614
## 70 34.22135 0.3848097 20.51856
## 90 34.27109 0.3841094 20.61150
## 70 32.64281 0.4178022 19.36089
## 90 32.68431 0.4176388 19.42442
## 70 34.11638 0.3973618 20.72923
## 90 34.23368 0.3949044 20.86619
## 70 33.19060 0.4035324 19.72386
## 90 33.29929 0.4015923 19.88236
## 70 33.52457 0.4115801 19.84656
## 90 33.63427 0.4095034 19.96908
## 70 32.24681 0.4308471 18.65851
## 90 32.25383 0.4312087 18.70005
## 70 33.32195 0.4209169 19.87993
## 90 33.45185 0.4186099 20.04155
## 70 32.67800 0.4251655 19.35587
## 90 32.77037 0.4235386 19.48998
## 70 33.85171 0.3941979 20.00044
## 90 33.91974 0.3931417 20.12051
## 70 33.23194 0.4060054 19.42678
## 90 33.29580 0.4051601 19.52709
## 70 33.91710 0.3898881 20.66078
## 90 34.02794 0.3881594 20.83186
## 70 32.83932 0.4134830 19.56506
## 90 32.92899 0.4111087 19.71729
## 70 33.26958 0.4144730 19.61832
## 90 33.30261 0.4139914 19.69236
## 70 32.37112 0.4271569 18.75771
## 90 32.42407 0.4263558 18.83244
## 70 34.28660 0.3972902 20.39188
## 90 34.31807 0.3975530 20.49382
## 70 32.67408 0.4251443 19.35504
## 90 32.78314 0.4233457 19.50483
## 70 33.78737 0.3905845 19.75666
## 90 33.81378 0.3902440 19.79304
## 70 33.00400 0.4030792 19.13792
## 90 33.03172 0.4027419 19.17733
## 70 34.75917 0.3760577 21.29134
## 90 34.79627 0.3758256 21.35274
## 70 33.38699 0.4020020 19.94065
## 90 33.41120 0.4019735 20.02273
## 70 32.67476 0.4219345 19.13517
## 90 32.70360 0.4213693 19.15593
## 70 32.06768 0.4290151 18.37188
## 90 32.06913 0.4290303 18.39056
## 70 33.84588 0.3978720 19.84670
## 90 33.86619 0.3977249 19.89071
## 70 32.74757 0.4169792 19.12366
## 90 32.78798 0.4161733 19.18363
## 70 33.28438 0.3948133 19.77152
## 90 33.32312 0.3940366 19.82632
```

```
##   70       32.63134  0.4161330  19.02502
##   90       32.63669  0.4160452  19.04708
##   70       34.51287  0.3810938  21.00366
##   90       34.52759  0.3810824  21.06806
##   70       32.18748  0.4241761  19.24108
##   90       32.23077  0.4234210  19.31860
##   70       33.39993  0.4022029  19.31937
##   90       33.40500  0.4024852  19.34629
##   70       31.99655  0.4302456  18.34360
##   90       32.02546  0.4295806  18.37356
##   70       33.48658  0.4101938  19.84974
##   90       33.51046  0.4099644  19.90047
##   70       32.78186  0.4161460  19.14843
##   90       32.83146  0.4149973  19.20984
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were nrounds = 90, max_depth = 6,
##  eta = 0.05, gamma = 0, colsample_bytree = 1, min_child_weight = 4
##  and subsample = 0.8.
```

Comparing Prediction Models

```r
models <- list( xgb = model_xgb,
                rf = model_rf,
                glmnet = model_glmnet,
                kknn = model_kknn,
                pls = model_pls,
                tree = model_rpart
)
resample_results <- resamples(models)
resample_results
```

```
##
## Call:
## resamples.default(x = models)
##
## Models: xgb, rf, glmnet, kknn, pls, tree
## Number of resamples: 50
## Performance metrics: MAE, RMSE, Rsquared
## Time estimates for: everything, final model fit
```

```r
summary(resample_results)
```

```
##
## Call:
## summary.resamples(object = resample_results)
##
## Models: xgb, rf, glmnet, kknn, pls, tree
## Number of resamples: 50
##
## MAE
##            Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## xgb    12.45974 15.07062 15.79873 16.12768 17.35844 21.05400    0
## rf     12.90208 15.15153 16.89774 16.83161 18.13347 22.23094    0
```

```
## glmnet 15.54462 18.44451 21.09668 20.59967 22.18272 26.84386    0
## kknn   14.36115 15.85349 17.71759 17.61308 18.88676 21.77379    0
## pls    15.68592 18.86746 20.49156 20.74194 22.59253 27.48153    0
## tree   16.59666 18.13871 19.76543 19.70701 20.98316 25.09659    0
##
## RMSE
##            Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## xgb    20.36910 24.77598 28.11116 28.71478 31.22733 45.04036    0
## rf     20.45757 24.54976 30.26079 30.52333 35.55357 43.44913    0
## glmnet 19.79341 29.83750 41.17531 39.20494 46.35801 62.35546    0
## kknn   23.16103 27.69886 31.41306 31.72765 35.80104 45.65694    0
## pls    19.66215 31.15205 39.53229 39.43403 45.76994 64.13915    0
## tree   24.80703 32.61077 38.19259 38.40981 43.57984 56.62911    0
##
## Rsquared
##              Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## xgb    0.12043064 0.3544449 0.5336537 0.5179434 0.6995061 0.8208837    0
## rf     0.16720157 0.3758397 0.5001517 0.4839504 0.6040610 0.7114627    0
## glmnet 0.09559345 0.1254328 0.1376330 0.1517795 0.1649333 0.2775887    0
## kknn   0.18126067 0.3035724 0.4302695 0.4490605 0.5936667 0.7679223    0
## pls    0.08464352 0.1169886 0.1334244 0.1455918 0.1669473 0.2601101    0
## tree   0.05603703 0.1175996 0.1792506 0.2282916 0.3035135 0.5714219    0
```

Conclusion: The prediction model based on **extreme gradient boosting** algorithm is the champion model.

Prediction of Total Cases in Future Using XGB model

Importing the test data features on which the predictive model will be applied to predict total number of cases per week at a future date)

```r
testset <- getURL("https://s3.amazonaws.com/drivendata/data/44/public/dengue_features_test.csv")
testset <- read.csv(text=testset)
names(testset)
```

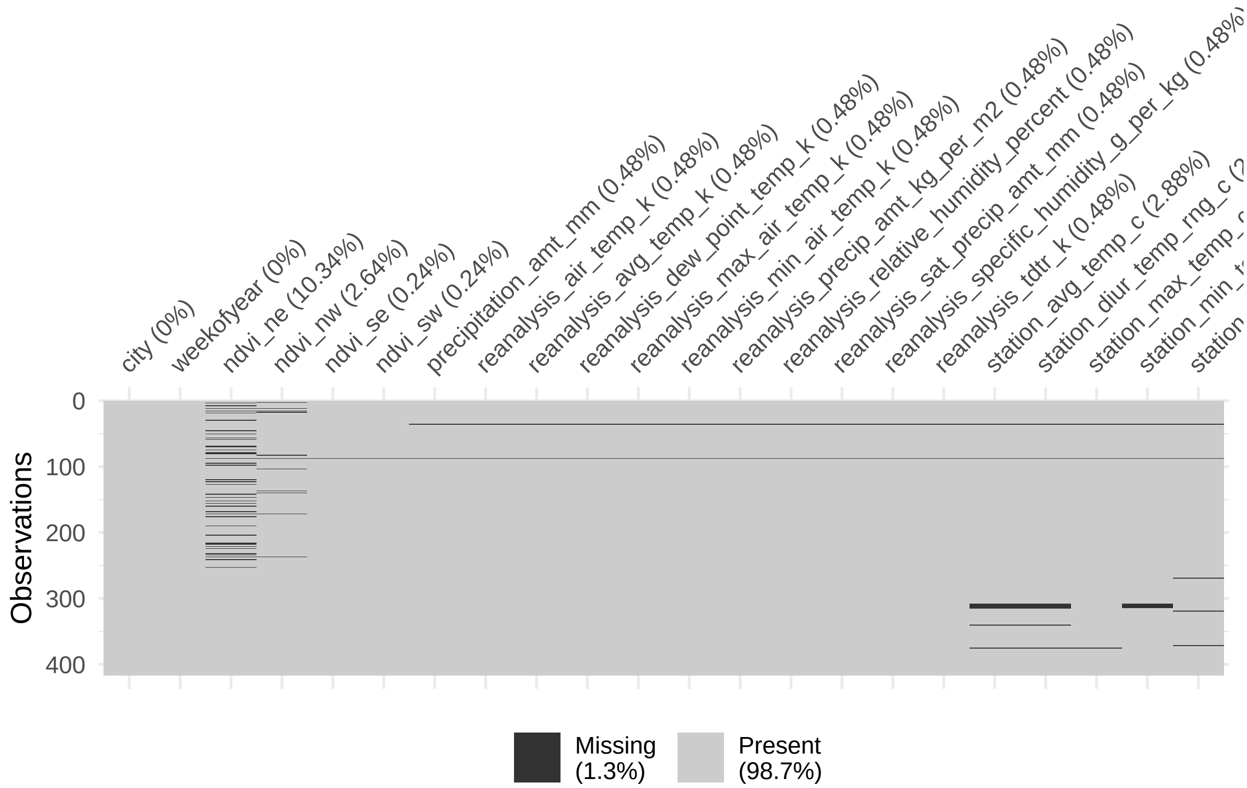
```
##  [1] "city"
##  [2] "year"
##  [3] "weekofyear"
##  [4] "week_start_date"
##  [5] "ndvi_ne"
##  [6] "ndvi_nw"
##  [7] "ndvi_se"
##  [8] "ndvi_sw"
##  [9] "precipitation_amt_mm"
## [10] "reanalysis_air_temp_k"
## [11] "reanalysis_avg_temp_k"
## [12] "reanalysis_dew_point_temp_k"
## [13] "reanalysis_max_air_temp_k"
## [14] "reanalysis_min_air_temp_k"
## [15] "reanalysis_precip_amt_kg_per_m2"
## [16] "reanalysis_relative_humidity_percent"
## [17] "reanalysis_sat_precip_amt_mm"
## [18] "reanalysis_specific_humidity_g_per_kg"
## [19] "reanalysis_tdtr_k"
```

```
## [20] "station_avg_temp_c"
## [21] "station_diur_temp_rng_c"
## [22] "station_max_temp_c"
## [23] "station_min_temp_c"
## [24] "station_precip_mm"
```

```r
dengue_test <- testset[, -c(2, 4)]
names(dengue_test)
```

```
##  [1] "city"
##  [2] "weekofyear"
##  [3] "ndvi_ne"
##  [4] "ndvi_nw"
##  [5] "ndvi_se"
##  [6] "ndvi_sw"
##  [7] "precipitation_amt_mm"
##  [8] "reanalysis_air_temp_k"
##  [9] "reanalysis_avg_temp_k"
## [10] "reanalysis_dew_point_temp_k"
## [11] "reanalysis_max_air_temp_k"
## [12] "reanalysis_min_air_temp_k"
## [13] "reanalysis_precip_amt_kg_per_m2"
## [14] "reanalysis_relative_humidity_percent"
## [15] "reanalysis_sat_precip_amt_mm"
## [16] "reanalysis_specific_humidity_g_per_kg"
## [17] "reanalysis_tdtr_k"
## [18] "station_avg_temp_c"
## [19] "station_diur_temp_rng_c"
## [20] "station_max_temp_c"
## [21] "station_min_temp_c"
## [22] "station_precip_mm"
```

```r
# Visualizing missing values for the test data
vis_miss(dengue_test)
```



Imputation of missing values inthe test data

```r
names(dengue_test)
```

```
##  [1] "city"
##  [2] "weekofyear"
##  [3] "ndvi_ne"
##  [4] "ndvi_nw"
##  [5] "ndvi_se"
##  [6] "ndvi_sw"
##  [7] "precipitation_amt_mm"
##  [8] "reanalysis_air_temp_k"
##  [9] "reanalysis_avg_temp_k"
## [10] "reanalysis_dew_point_temp_k"
## [11] "reanalysis_max_air_temp_k"
## [12] "reanalysis_min_air_temp_k"
## [13] "reanalysis_precip_amt_kg_per_m2"
## [14] "reanalysis_relative_humidity_percent"
## [15] "reanalysis_sat_precip_amt_mm"
## [16] "reanalysis_specific_humidity_g_per_kg"
## [17] "reanalysis_tdtr_k"
## [18] "station_avg_temp_c"
## [19] "station_diur_temp_rng_c"
## [20] "station_max_temp_c"
## [21] "station_min_temp_c"
```

```
## [22] "station_precip_mm"
```

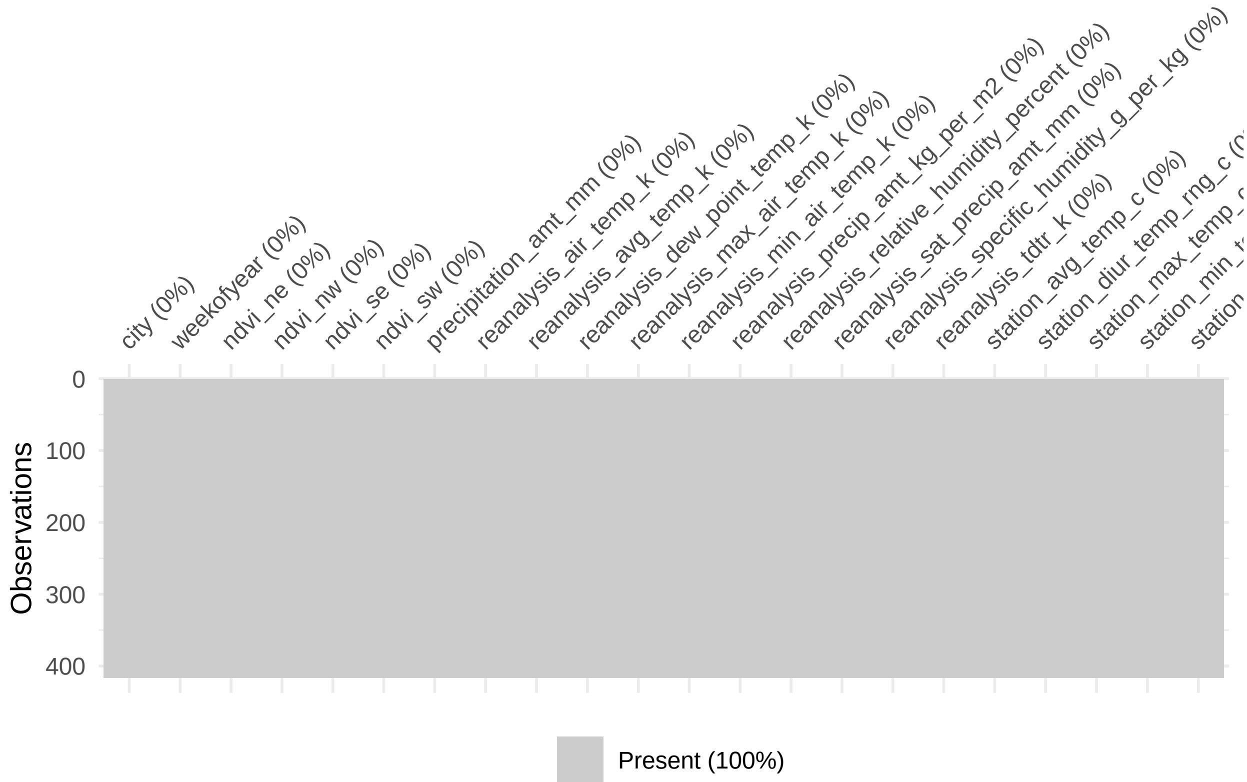
```r
pre.process <- preProcess(dengue_test, method = "bagImpute")
imputed.data <- predict(pre.process, dengue_test)
dengue_test$ndvi_ne <- imputed.data[,3]
dengue_test$ndvi_nw <- imputed.data[,4]
dengue_test$ndvi_se <- imputed.data[,5]
dengue_test$ndvi_sw <- imputed.data[,6]
dengue_test$precipitation_amt_mm <- imputed.data[,7]
dengue_test$reanalysis_air_temp_k <- imputed.data[, 8]
dengue_test$reanalysis_avg_temp_k <- imputed.data[,9]
dengue_test$reanalysis_dew_point_temp_k <- imputed.data[,10]
dengue_test$reanalysis_max_air_temp_k <- imputed.data[,11]
dengue_test$reanalysis_min_air_temp_k <- imputed.data[,12]
dengue_test$reanalysis_precip_amt_kg_per_m2 <- imputed.data[,13]
dengue_test$reanalysis_relative_humidity_percent <- imputed.data[,14]
dengue_test$reanalysis_sat_precip_amt_mm <- imputed.data[,15]
dengue_test$reanalysis_specific_humidity_g_per_kg <- imputed.data[,16]
dengue_test$reanalysis_tdtr_k <- imputed.data[,17]
dengue_test$station_avg_temp_c <- imputed.data[,18]
dengue_test$station_diur_temp_rng_c <- imputed.data[,19]
dengue_test$station_max_temp_c <- imputed.data[,20]
dengue_test$station_min_temp_c <- imputed.data[,21]
dengue_test$station_precip_mm <- imputed.data[,22]

dim(dengue_test)
```

```
## [1] 416  22
```

```r
anyNA(dengue_test)
```

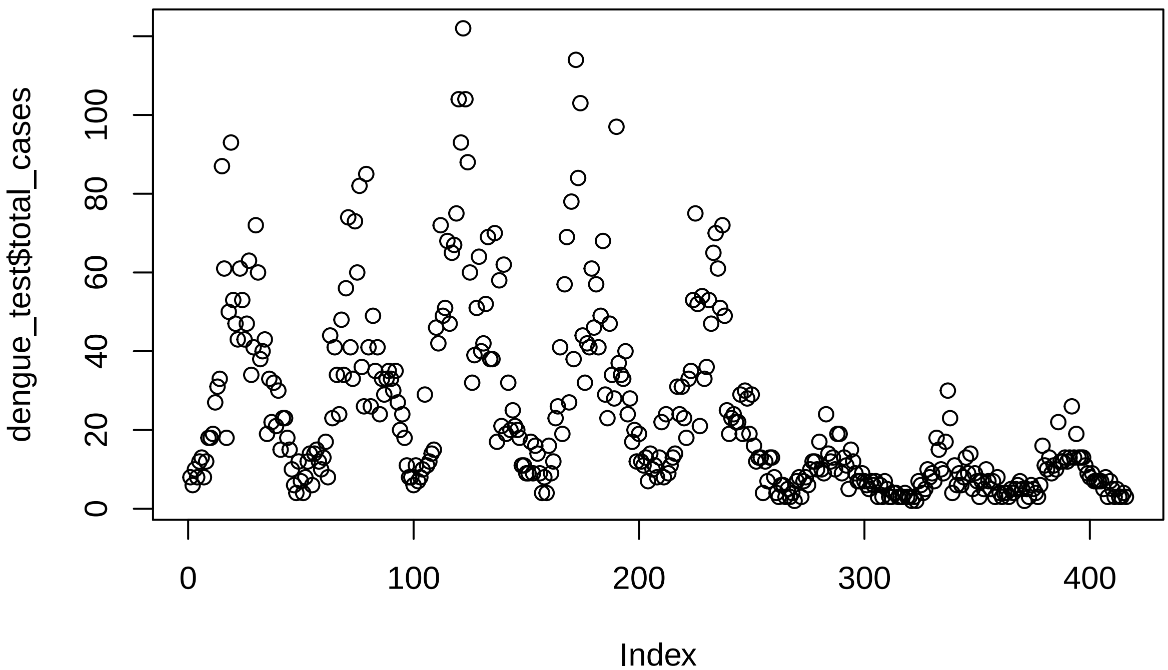
```
## [1] FALSE
```

```r
vis_miss(dengue_test)
```



Predicting total cases on test data

```r
# predict values for test data
pred <- predict(model_xgb, dengue_test)
dengue_test$total_cases <- round(pred, digits = 0)

# Visualizing the time-series total cases on the test data
plot(dengue_test$total_cases)
```




```r
# Summary of the predicted total cases
summary(dengue_test$total_cases)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    8.00   14.00   23.74   33.00  122.00
```

```r
#Entering the predicted 'total_cases' from the test-set into the submission form
Submitformat <- getURL("https://s3.amazonaws.com/drivendata/data/44/public/submission_format.csv")
submitformat <- read.csv(text=Submitformat)
submitformat$total_cases<- dengue_test$total_cases

# Exporting the output (total cases) to local drive as an Excel file
write.csv(submitformat, "D://STUDY//MSIS//DM//submit0407620xgb_send.csv", row.names = FALSE)
```

Our Current Ranking in the DengAI Competition at DrivenData.org