CKME 136 Final Results

Step 1: Data Description

```
#Read data into a dataframe
data <- read.csv(file="C:/Users/krist/Desktop/CKME136/03. Final Project/Data
July 13/conposcovidloc (2).csv",header=T,sep=",",na.strings=c(""," ","NA"))

#summary statistics of dataset
summary(data)
```

```
##       Row_ID        Accurate_Episode_Date  Case_Reported_Date
##   Min.   :     1    2020-04-17:  658       2020-04-17:  750
##   1st Qu.: 9210     2020-04-13:  655       2020-04-15:  613
##   Median :18420     2020-04-15:  650       2020-04-13:  608
##   Mean   :18420     2020-04-14:  621       2020-05-29:  608
##   3rd Qu.:27630     2020-04-16:  603       2020-04-20:  604
##   Max.   :36839     (Other)   :33650       2020-04-18:  603
##                     NA's      :    2       (Other)   :33053
##    Test_Reported_Date      Specimen_Date      Age_Group          Client_Gender
##   2020-04-17:  689      2020-04-13:  746      50s   :5927     FEMALE     :19545
##   2020-04-18:  611      2020-04-15:  733      20s   :5730     MALE       :17014
##   2020-04-20:  607      2020-04-14:  690      40s   :5243     OTHER      :    9
##   2020-04-13:  604      2020-04-17:  687      30s   :5193     TRANSGENDER:    8
##   2020-04-15:  593      2020-04-16:  668      60s   :4150     UNKNOWN    :  263
##   (Other)   :33327     (Other)   :33030      80s   :3520
##   NA's      :  408     NA's      :  285      (Other):7076
##        Case_AcquisitionInfo         Outcome1       Outbreak_Related
##   CC            :11405       Fatal       : 2722    Yes :14593
##   No Epi-link   : 7313       Not Resolved: 1454    NA's:22246
##   No Info-Missing:  880      Resolved    :32663
##   No Info-Unk   :  851
##   OB            :14582
##   Travel        : 1808
##
##                                  Reporting_PHU
##   Toronto Public Health                   :13673
##   Peel Public Health                      : 6184
##   York Region Public Health Services: 3130
##   Ottawa Public Health                    : 2158
##   Windsor-Essex County Health Unit   : 1805
##   Durham Region Health Department    : 1751
##   (Other)                                 : 8138
##               Reporting_PHU_Address   Reporting_PHU_City
##   277 Victoria Street, 5th Floor:13673    Toronto    :13673
##   7120 Hurontario Street        : 6184    Mississauga: 6184
##   17250 Yonge Street            : 3130    Newmarket  : 3130
```

```
##  100 Constellation Drive      : 2158    Ottawa    : 2158
##  1005 Ouellette Avenue        : 1805    Windsor   : 1805
##  605 Rossland Road East       : 1751    Whitby    : 1751
##  (Other)                      : 8138    (Other)   : 8138
##  Reporting_PHU_Postal_Code
##  M5B 1W2:13673
##  L5W 1N4: 6184
##  L3Y 6Z1: 3130
##  K2G 6J8: 2158
##  N9A 4J8: 1805
##  L1N 0B2: 1751
##  (Other): 8138
##                                                        Reporting_PHU_Website
##  www.toronto.ca/community-people/health-wellness-care/       :13673
##  www.peelregion.ca/health/                                   : 6184
##  www.york.ca/wps/portal/yorkhome/health/                     : 3130
##  www.ottawapublichealth.ca                                   : 2158
##  www.wechu.org                                               : 1805
##  www.durham.ca/en/health-and-wellness/health-and-wellness.aspx: 1751
##  (Other)                                                     : 8138
##  Reporting_PHU_Latitude Reporting_PHU_Longitude
##  Min.   :42.31          Min.   :-94.49
##  1st Qu.:43.65          1st Qu.:-79.71
##  Median :43.66          Median :-79.38
##  Mean   :43.74          Mean   :-79.53
##  3rd Qu.:43.66          3rd Qu.:-79.38
##  Max.   :49.77          Max.   :-74.74
##
```

str(data)

```
## 'data.frame':    36839 obs. of  17 variables:
##  $ Row_ID               : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Accurate_Episode_Date : Factor w/ 158 levels "2020-01-01","2020-01-
10",..: 31 26 30 26 27 30 33 33 33 35 ...
##  $ Case_Reported_Date    : Factor w/ 141 levels "2020-01-23","2020-01-
24",..: 16 16 17 16 17 17 18 17 18 18 ...
##  $ Test_Reported_Date    : Factor w/ 140 levels "2020-01-27","2020-02-
03",..: 17 15 16 18 17 16 18 17 16 17 ...
##  $ Specimen_Date         : Factor w/ 144 levels "2020-01-23","2020-01-
24",..: 20 20 20 18 20 19 20 21 20 21 ...
##  $ Age_Group             : Factor w/ 10 levels "<20","20s","30s",..: 5
4 3 4 3 5 2 2 3 1 ...
##  $ Client_Gender         : Factor w/ 5 levels "FEMALE","MALE",..: 2 2 1
2 2 2 1 2 1 1 ...
##  $ Case_AcquisitionInfo  : Factor w/ 6 levels "CC","No Epi-link",..: 6
6 6 6 6 2 6 1 6 1 ...
##  $ Outcome1              : Factor w/ 3 levels "Fatal","Not
Resolved",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ Outbreak_Related      : Factor w/ 1 level "Yes": NA NA NA 1 NA NA NA
```

```
 NA NA NA ...
##  $ Reporting_PHU          : Factor w/ 34 levels "Algoma Public Health
Unit",..: 34 31 34 20 31 28 21 31 9 21 ...
##  $ Reporting_PHU_Address  : Factor w/ 34 levels "100 Constellation
Drive",..: 14 23 14 1 23 9 32 23 6 32 ...
##  $ Reporting_PHU_City     : Factor w/ 34 levels
"Barrie","Belleville",..: 14 31 14 17 31 27 12 31 16 12 ...
##  $ Reporting_PHU_Postal_Code: Factor w/ 34 levels "K2G 6J8","K6J 5T1",..:
11 16 11 1 16 30 13 16 14 13 ...
##  $ Reporting_PHU_Website  : Factor w/ 34 levels
"www.algomapublichealth.com",..: 34 31 34 19 31 22 20 31 6 20 ...
##  $ Reporting_PHU_Latitude : num  44 43.7 44 45.3 43.7 ...
##  $ Reporting_PHU_Longitude : num  -79.5 -79.4 -79.5 -75.8 -79.4 ...
```

Step 1: Data Preparation

```r
#Find missing values in data
sapply(data, function(x) sum(is.na(x)))
```

```
##                  Row_ID      Accurate_Episode_Date
##                       0                          2
##      Case_Reported_Date         Test_Reported_Date
##                       0                        408
##           Specimen_Date                  Age_Group
##                     285                          0
##           Client_Gender        Case_AcquisitionInfo
##                       0                          0
##                Outcome1            Outbreak_Related
##                       0                      22246
##           Reporting_PHU      Reporting_PHU_Address
##                       0                          0
##      Reporting_PHU_City Reporting_PHU_Postal_Code
##                       0                          0
##   Reporting_PHU_Website     Reporting_PHU_Latitude
##                       0                          0
##  Reporting_PHU_Longitude
##                       0
```

```r
sapply(data, function(x) length(unique(x)))
```

```
##                  Row_ID      Accurate_Episode_Date
##                   36839                        159
##      Case_Reported_Date         Test_Reported_Date
##                     141                        141
##           Specimen_Date                  Age_Group
##                     145                         10
##           Client_Gender        Case_AcquisitionInfo
##                       5                          6
##                Outcome1            Outbreak_Related
##                       3                          2
```

```
##              Reporting_PHU        Reporting_PHU_Address
##                          34                           34
##         Reporting_PHU_City  Reporting_PHU_Postal_Code
##                          34                           34
##      Reporting_PHU_Website      Reporting_PHU_Latitude
##                          34                           34
##   Reporting_PHU_Longitude
##                          34
```

```r
#install.packages("Amelia")
library(Amelia)
```

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

## Loading required package: Rcpp

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

## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.7.6, built: 2019-11-24)
## ## Copyright (C) 2005-2020 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##
```

```r
missmap(data, main = "Missing values vs observed")
#4% of data is missing

#Remove Not Resolved rows in Outcome1
dataclean <- droplevels(data[!data$Outcome1 == 'Not Resolved',])

#Outbreak Related has "missing values" but should be No
sum(is.na(dataclean$Outbreak_Related) == TRUE)
```

```
## [1] 21006
```

```r
length(dataclean$Outbreak_Related)
```

```
## [1] 35385
```

```r
#Replace NA with No
dataclean$Outbreak_Related <- factor(dataclean$Outbreak_Related, exclude =
NULL,
            levels = c("Yes", NA),
            labels = c("Yes", "No"))
table(dataclean$Outbreak_Related, useNA = "always")
```

```
##
##   Yes    No  <NA>
## 14379 21006     0
```

```r
str(dataclean$Outbreak_Related)
```

```
##  Factor w/ 2 levels "Yes","No": 2 2 2 1 2 2 2 2 2 2 ...

#Remove repetitive variables
dataclean$Reporting_PHU <- NULL
dataclean$Reporting_PHU_Address <- NULL
dataclean$Reporting_PHU_Postal_Code <- NULL
dataclean$Reporting_PHU_Website <- NULL

#summary statistics of clean dataset
summary(dataclean)

##       Row_ID       Accurate_Episode_Date  Case_Reported_Date
##   Min.   :    1    2020-04-17:  655        2020-04-17:  748
##   1st Qu.: 8872    2020-04-13:  653        2020-04-15:  611
##   Median :17759    2020-04-15:  647        2020-05-29:  608
##   Mean   :18081    2020-04-14:  617        2020-04-13:  607
##   3rd Qu.:27424    2020-04-16:  602        2020-04-20:  601
##   Max.   :36839    (Other)   :32209        2020-04-18:  599
##                    NA's      :    2        (Other)   :31611
##    Test_Reported_Date     Specimen_Date      Age_Group        Client_Gender
##   2020-04-17:  685     2020-04-13:  743     50s  :5721     FEMALE     :18903
##   2020-04-18:  606     2020-04-15:  728     20s  :5396     MALE       :16207
##   2020-04-20:  606     2020-04-14:  687     40s  :5037     OTHER      :    9
##   2020-04-13:  604     2020-04-17:  684     30s  :4956     TRANSGENDER:    8
##   2020-04-15:  590     2020-04-16:  667     60s  :4011     UNKNOWN    :  258
##   (Other)   :31903     (Other)   :31607     80s  :3457
##   NA's      :  391     NA's      :  269     (Other):6807
##       Case_AcquisitionInfo     Outcome1      Outbreak_Related
##   CC             :10820     Fatal   : 2722   Yes:14379
##   No Epi-link    : 6957     Resolved:32663   No :21006
##   No Info-Missing:  661
##   No Info-Unk    :  815
##   OB             :14368
##   Travel         : 1764
##
##    Reporting_PHU_City  Reporting_PHU_Latitude  Reporting_PHU_Longitude
##   Toronto     :13183   Min.   :42.31           Min.   :-94.49
##   Mississauga : 5838   1st Qu.:43.65           1st Qu.:-79.71
##   Newmarket   : 2997   Median :43.66           Median :-79.38
##   Ottawa      : 2104   Mean   :43.75           Mean   :-79.52
##   Whitby      : 1706   3rd Qu.:43.66           3rd Qu.:-79.38
##   Windsor     : 1600   Max.   :49.77           Max.   :-74.74
##   (Other)     : 7957

str(dataclean)

## 'data.frame':    35385 obs. of  13 variables:
##  $ Row_ID             : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Accurate_Episode_Date : Factor w/ 154 levels "2020-01-01","2020-01-
10",..: 31 26 30 26 27 30 33 33 33 35 ...
##  $ Case_Reported_Date    : Factor w/ 141 levels "2020-01-23","2020-01-
```

```
24",..: 16 16 17 16 17 17 18 17 18 18 ...
##  $ Test_Reported_Date    : Factor w/ 140 levels "2020-01-27","2020-02-
03",..: 17 15 16 18 17 16 18 17 16 17 ...
##  $ Specimen_Date         : Factor w/ 143 levels "2020-01-23","2020-01-
24",..: 20 20 20 18 20 19 20 21 20 21 ...
##  $ Age_Group             : Factor w/ 10 levels "<20","20s","30s",..: 5 4
3 4 3 5 2 2 3 1 ...
##  $ Client_Gender         : Factor w/ 5 levels "FEMALE","MALE",..: 2 2 1 2
2 2 1 2 1 1 ...
##  $ Case_AcquisitionInfo  : Factor w/ 6 levels "CC","No Epi-link",..: 6 6
6 6 6 2 6 1 6 1 ...
##  $ Outcome1              : Factor w/ 2 levels "Fatal","Resolved": 2 2 2 2
2 2 2 2 2 2 ...
##  $ Outbreak_Related      : Factor w/ 2 levels "Yes","No": 2 2 2 1 2 2 2 2
2 2 ...
##  $ Reporting_PHU_City     : Factor w/ 34 levels "Barrie","Belleville",..:
14 31 14 17 31 27 12 31 16 12 ...
##  $ Reporting_PHU_Latitude : num  44 43.7 44 45.3 43.7 ...
##  $ Reporting_PHU_Longitude: num  -79.5 -79.4 -79.5 -75.8 -79.4 ...
```

```
#bar charts
library(ggplot2)
```


Missing values vs observed
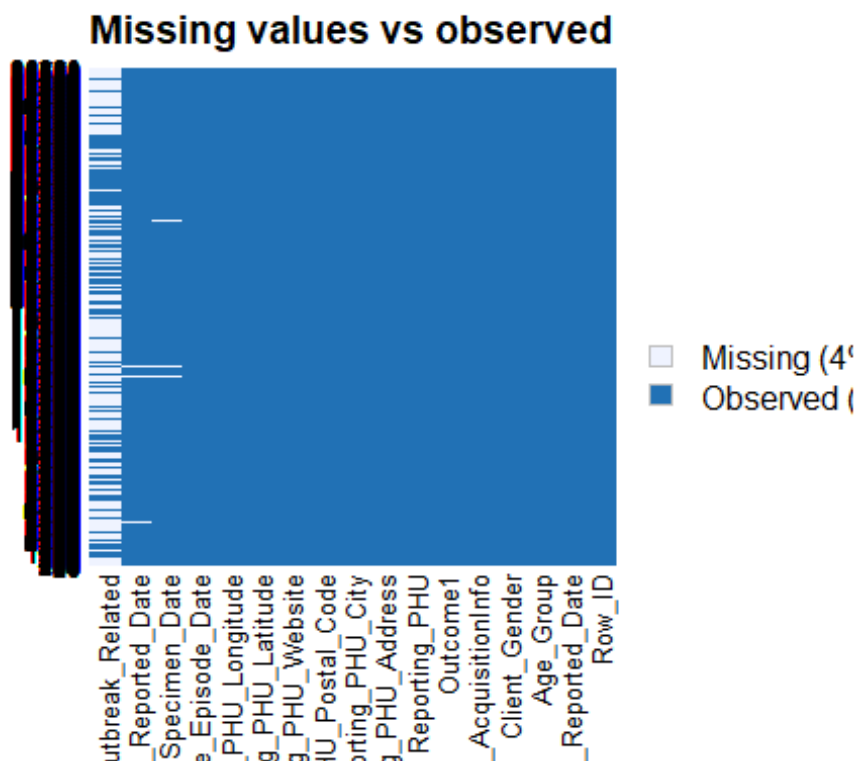
```
#Access ggplot-colors
gg_color_hue <- function(n) {
  hues = seq(15, 375, length=n+1)
  hcl(h=hues, l=65, c=100)[1:n]
```
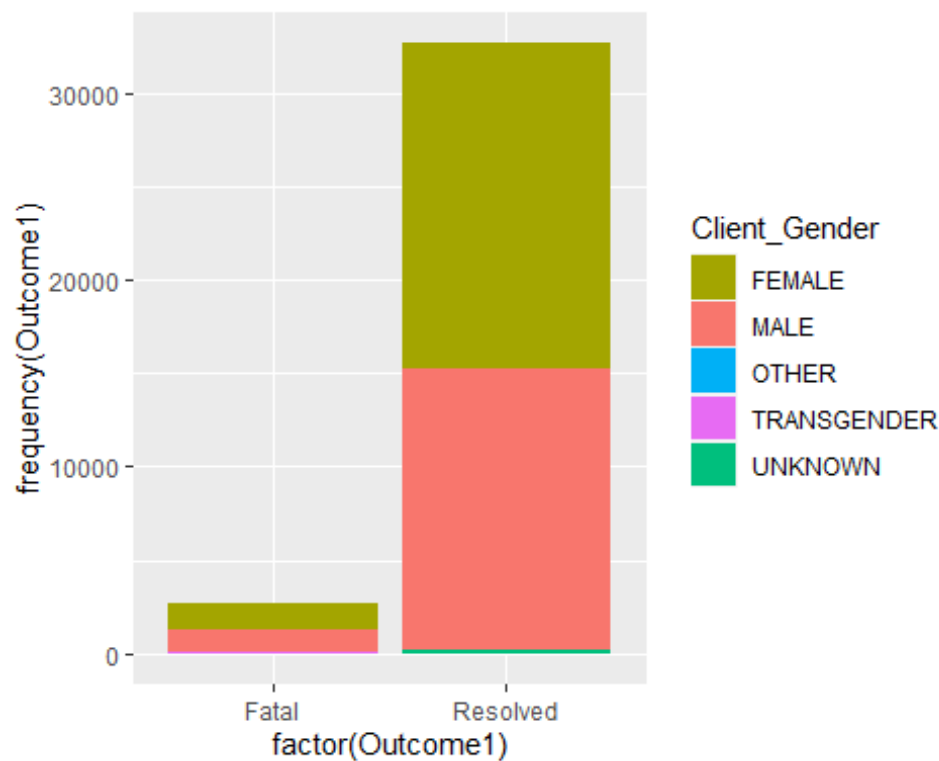
```
}

#create custom palette for Client Gender
mycols <- gg_color_hue(length(unique(dataclean$Client_Gender)))
names(mycols) <- unique(dataclean$Client_Gender)

#stacked bar chart for Client Genders in Outcome
ggplot(dataclean, aes(x = factor(Outcome1), y = frequency(Outcome1),
fill=Client_Gender)) + geom_bar(stat = 'identity') + scale_fill_manual(values
= mycols)
```



```
#Table of Client Genders in Outcome
Table.Gender <- table(dataclean$Outcome, dataclean$Client_Gender)
Table.Gender

##
##            FEMALE  MALE OTHER TRANSGENDER UNKNOWN
##    Fatal      1442  1240     0           1      39
##    Resolved  17461 14967     9           7     219

prop.table(Table.Gender, 1) #as percentage

##
##                    FEMALE          MALE        OTHER   TRANSGENDER
##    Fatal     0.5297575312 0.4555473916 0.0000000000 0.0003673769
##    Resolved  0.5345804121 0.4582249028 0.0002755411 0.0002143098
##
```
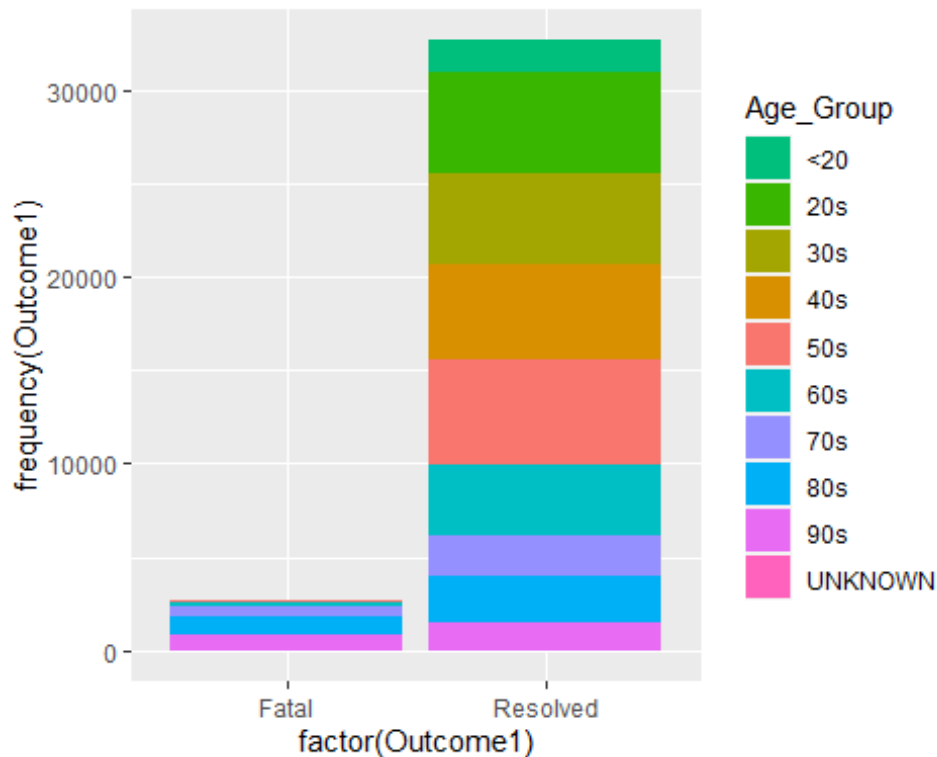
```
##                 UNKNOWN
##    Fatal     0.0143277002
##    Resolved  0.0067048342
```

```
#Age Group stacked bar chart
mycols <- gg_color_hue(length(unique(dataclean$Age_Group)))
names(mycols) <- unique(dataclean$Age_Group)
ggplot(dataclean, aes(x = factor(Outcome1), y = frequency(Outcome1),
fill=Age_Group)) + geom_bar(stat = 'identity') + scale_fill_manual(values =
mycols)
```



```
#Table of Age Group in Outcome
Table.Age <- table(dataclean$Outcome, dataclean$Age_Group)
Table.Age
```

```
##
##             <20   20s   30s   40s   50s   60s   70s   80s   90s UNKNOWN
##    Fatal      1     4     7    23    88   239   484   991   885        0
##    Resolved 1723  5392  4949  5014  5633  3772  2169  2466  1537        8
```

```
prop.table(Table.Age, 1) #as percentage
```

```
##
##                   <20          20s          30s          40s
##    Fatal    0.0003673769 0.0014695077 0.0025716385 0.0084496694
##    Resolved 0.0527508190 0.1650797538 0.1515170070 0.1535070263
##
##                   50s          60s          70s          80s
```

```
##    Fatal     0.0323291697 0.0878030860 0.1778104335 0.3640705364
##    Resolved 0.1724581331 0.1154823501 0.0664054129 0.0754982702
##
##                      90s        UNKNOWN
##    Fatal     0.3251285819 0.0000000000
##    Resolved 0.0470563022 0.0002449255
```

#Case Acquisition Info stacked bar chart
```
mycols <- gg_color_hue(length(unique(dataclean$Case_AcquisitionInfo)))
names(mycols) <- unique(dataclean$Case_AcquisitionInfo)
ggplot(dataclean, aes(x = factor(Outcome1), y = frequency(Outcome1),
fill=Case_AcquisitionInfo)) + geom_bar(stat = 'identity') +
scale_fill_manual(values = mycols)
```
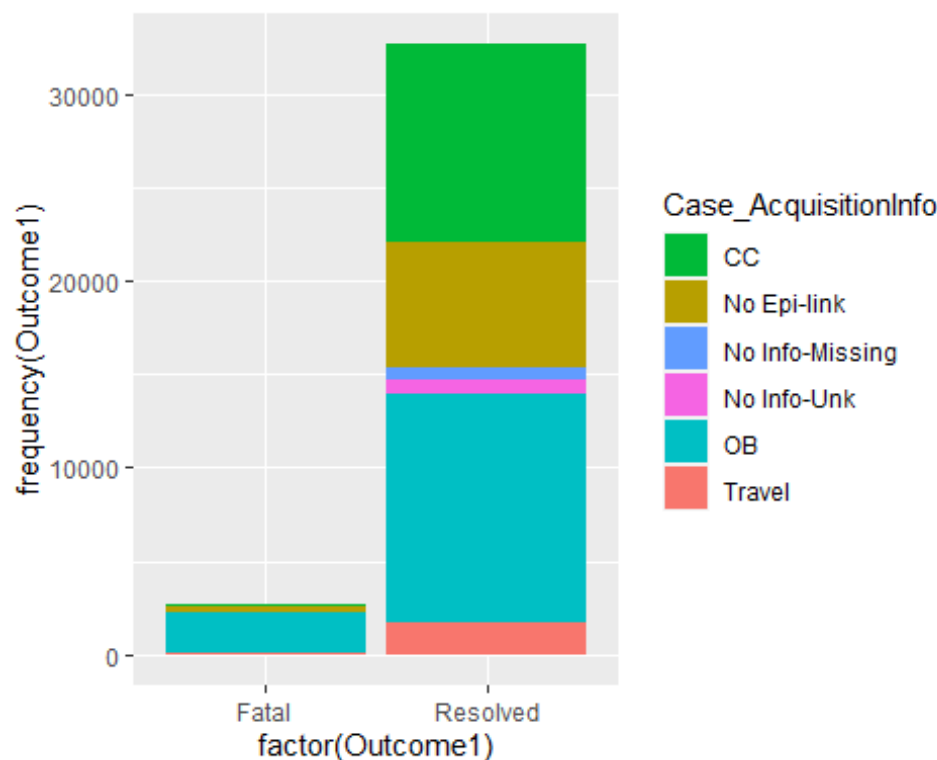

#Table of Case Acquisition Info in Outcome
```
Table.CAI <- table(dataclean$Outcome, dataclean$Case_AcquisitionInfo)
Table.CAI
```

```
##
##              CC No Epi-link No Info-Missing No Info-Unk    OB Travel
##    Fatal    178        252              23          54  2155     60
##    Resolved 10642      6705             638         761 12213   1704
```

```
prop.table(Table.CAI, 1) #as percentage
```

```
##
##                   CC No Epi-link No Info-Missing No Info-Unk          OB
##    Fatal    0.065393093 0.092578986      0.008449669 0.019838354 0.791697281
```

```
##    Resolved 0.325812081 0.205278143       0.019532805 0.023298534 0.373909316
##
##                Travel
##    Fatal     0.022042616
##    Resolved 0.052169121
```

#Outbreak Related stacked bar chart
```
mycols <- gg_color_hue(length(unique(dataclean$Outbreak_Related)))
names(mycols) <- unique(dataclean$Outbreak_Related)
ggplot(dataclean, aes(x = factor(Outcome1), y = frequency(Outcome1),
fill=Outbreak_Related)) + geom_bar(stat = 'identity') +
scale_fill_manual(values = mycols)
```
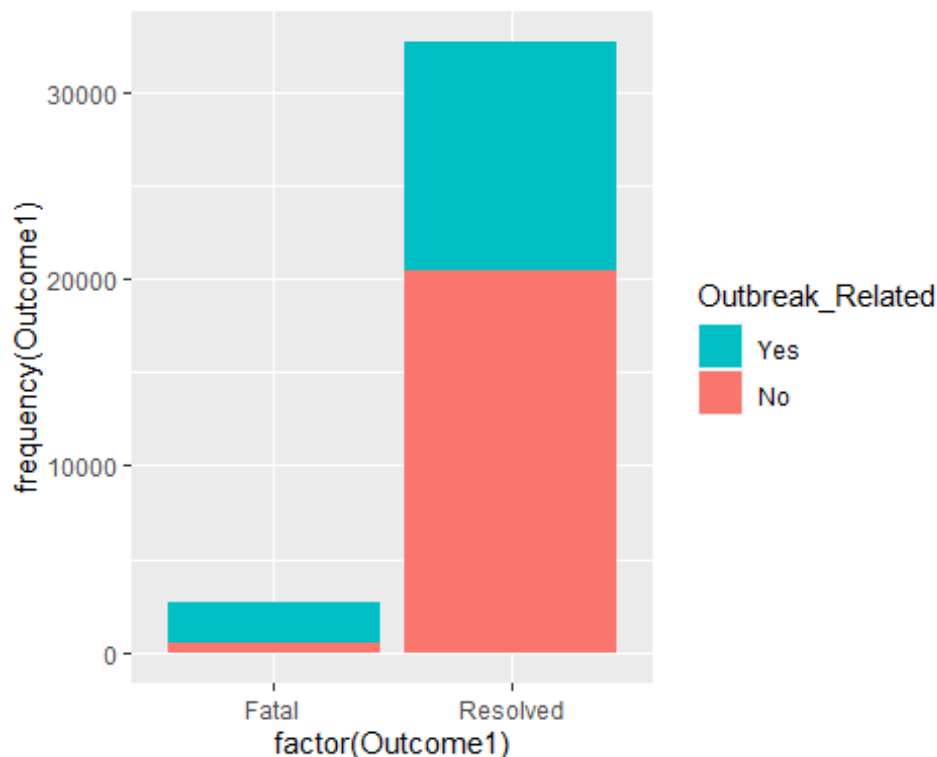

#Table of Outbreak Related in Outcome
```
Table.Outbreak <- table(dataclean$Outcome, dataclean$Outbreak_Related)
Table.Outbreak
```

```
##
##               Yes    No
##    Fatal      2155   567
##    Resolved 12224 20439
```

```
prop.table(Table.Outbreak, 1) #as percentage
```

```
##
##                 Yes        No
##    Fatal     0.7916973 0.2083027
##    Resolved 0.3742461 0.6257539
```

```
#Reporting City stacked bar chart
mycols <- gg_color_hue(length(unique(dataclean$Reporting_PHU_City)))
names(mycols) <- unique(dataclean$Reporting_PHU_City)
ggplot(dataclean, aes(x = factor(Outcome1), y = frequency(Outcome1),
fill=Reporting_PHU_City)) + geom_bar(stat = 'identity') +
scale_fill_manual(values = mycols)
```
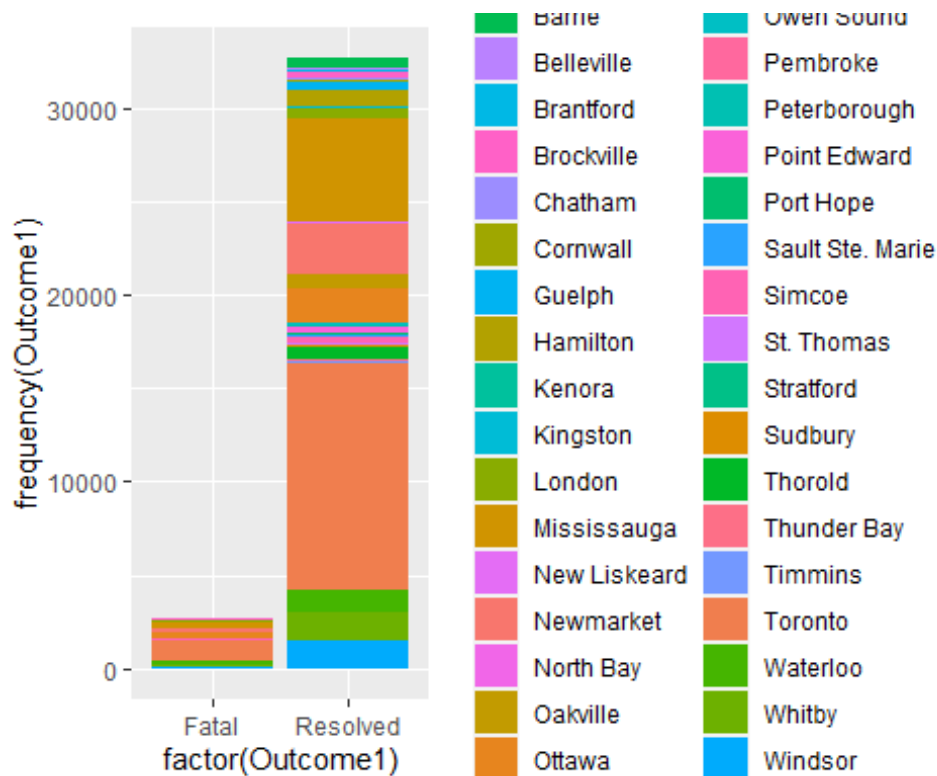


```
#Table of City in Outcome
Table.City <- table(dataclean$Outcome, dataclean$Reporting_PHU_City)
Table.City
```

```
##
##             Barrie Belleville Brantford Brockville Chatham Cornwall Guelph
##   Fatal         34          5         5         52       1       12     36
##   Resolved     572         39       127        300     157      151    446
##
##             Hamilton Kenora Kingston London Mississauga New Liskeard
##   Fatal           44      0        0     57         308            0
##   Resolved       792     40       96    569        5530           18
##
##             Newmarket North Bay Oakville Ottawa Owen Sound Pembroke
##   Fatal           249         1       25    263          0        1
##   Resolved       2748        33      737   1841        107       28
##
##             Peterborough Point Edward Port Hope Sault Ste. Marie Simcoe
##   Fatal                2           25        20                0     37
##   Resolved            93          260       181               25    394
```

```
## 
##           St. Thomas Stratford Sudbury Thorold Thunder Bay Timmins
##    Fatal              5        5       2      64           1       8
##    Resolved          79       54      65     699          91      59
## 
##           Toronto Waterloo Whitby Windsor
##    Fatal      1092      118    181      69
##    Resolved  12091     1185   1525    1531
```

```r
prop.table(Table.City, 1) #as percentage
```

```
## 
##                    Barrie     Belleville      Brantford     Brockville
##    Fatal     0.0124908156 0.0018368846 0.0018368846 0.0191036003
##    Resolved  0.0175121697 0.0011940116 0.0038881915 0.0091847044
## 
##                   Chatham       Cornwall         Guelph       Hamilton
##    Fatal     0.0003673769 0.0044085231 0.0132255694 0.0161645849
##    Resolved  0.0048066620 0.0046229679 0.0136545939 0.0242476196
## 
##                    Kenora       Kingston         London     Mississauga
##    Fatal     0.0000000000 0.0000000000 0.0209404849 0.1131520940
##    Resolved  0.0012246273 0.0029391054 0.0174203227 0.1693047179
## 
##              New Liskeard      Newmarket      North Bay       Oakville
##    Fatal     0.0000000000 0.0914768553 0.0003673769 0.0091844232
##    Resolved  0.0005510823 0.0841318924 0.0010103175 0.0225637572
## 
##                    Ottawa     Owen Sound       Pembroke  Peterborough
##    Fatal     0.0966201323 0.0000000000 0.0003673769 0.0007347539
##    Resolved  0.0563634694 0.0032758779 0.0008572391 0.0028472584
## 
##              Point Edward      Port Hope Sault Ste. Marie         Simcoe
##    Fatal     0.0091844232 0.0073475386    0.0000000000 0.0135929464
##    Resolved  0.0079600772 0.0055414383    0.0007653920 0.0120625785
## 
##                 St. Thomas      Stratford        Sudbury        Thorold
##    Fatal     0.0018368846 0.0018368846 0.0007347539 0.0235121234
##    Resolved  0.0024186388 0.0016532468 0.0019900193 0.0214003613
## 
##               Thunder Bay        Timmins        Toronto       Waterloo
##    Fatal     0.0003673769 0.0029390154 0.4011756062 0.0433504776
##    Resolved  0.0027860270 0.0018063252 0.3701742032 0.0362795824
## 
##                    Whitby         Windsor
##    Fatal     0.0664952241 0.0253490081
##    Resolved  0.0466889141 0.0468726081
```

```r
#install.packages("maps")
#install.packages("mapdata")
```

```r
#install.packages("mapproj")
library(maps)

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

library(mapdata)

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

library(mapproj)

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

map(database = "worldHires", "Canada", xlim=c(-140,-110),ylim=c(48,64),
col="grey80", fill=TRUE, projection="gilbert", orientation= c(90,0,225))
lon <- c(dataclean$Reporting_PHU_Longitude)
lat <- c(dataclean$Reporting_PHU_Latitude)
coord <- mapproject(lon, lat, proj="gilbert", orientation=c(90, 0, 225))
#convert points to projected lat/long
points(coord, pch=21, cex=1.2, col="red")
```
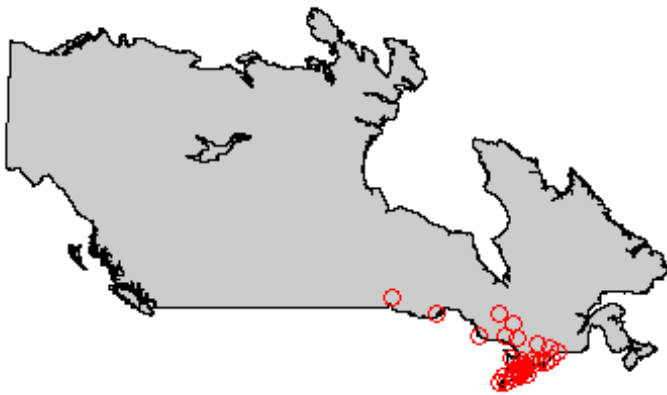

Step 3:

Experimental Design

```r
#check balance of data
table(dataclean$Outcome1)

##
##     Fatal Resolved
##      2722     32663
```

```r
#very imbalanced data

#balance data with function ROSE (Randomly Over Sampling Examples) which
creates a sample of synthetic data by enlarging the features space of
minority and majority class examples
#install.packages("ROSE")
library(ROSE)
```

Warning: package 'ROSE' was built under R version 3.5.3

Loaded ROSE 0.0-3

```r
dataclean.balanced <- ROSE(Outcome1~., data = dataclean, seed = 100)$data
table(dataclean.balanced$Outcome1)
```

```
##
## Resolved    Fatal
##    17585    17374
```

```r
#Split data into training (10%) and test (90%) sets
#createDataPartition function does stratified random sampling
set.seed(100)
#install.packages("caret")
library(caret)
```

Warning: package 'caret' was built under R version 3.5.3

Loading required package: lattice

```r
trainingRows <- createDataPartition(dataclean.balanced$Outcome1, p = 0.9,
list = FALSE)
training <- dataclean.balanced[trainingRows,]
test  <- dataclean.balanced[-trainingRows,]

#see if proportion of training and test set are the same
prop.table(table(training$Outcome1))
```

```
##
##  Resolved     Fatal
## 0.5030193 0.4969807
```

```r
prop.table(table(test$Outcome1))
```

```
##
##  Resolved     Fatal
## 0.5030043 0.4969957
```

```r
#proportions are the same
```

Step 4: Modeling - (1) Logistic Regression

```
#install.packages("caret")
library(caret)
#Fit training data into Multinomial Logistic Regression Model
model.LogReg <- glm(Outcome1 ~ Client_Gender + Age_Group +
Case_AcquisitionInfo + Outbreak_Related + Reporting_PHU_City, family =
binomial(link = "logit"), data=training)
summary (model.LogReg)

##
## Call:
## glm(formula = Outcome1 ~ Client_Gender + Age_Group + Case_AcquisitionInfo
+
##      Outbreak_Related + Reporting_PHU_City, family = binomial(link =
"logit"),
##      data = training)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.3330  -0.3838  -0.0003   0.6130   3.2751
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -1.865e+01  6.822e+02  -0.027  0.97819
## Client_GenderMALE                 5.980e-01  3.519e-02  16.996  < 2e-16
## Client_GenderOTHER               -1.046e+01  1.005e+03  -0.010  0.99169
## Client_GenderTRANSGENDER          1.200e+00  1.090e+00   1.101  0.27087
## Client_GenderUNKNOWN              2.678e-01  1.657e-01   1.616  0.10600
## Age_Group20s                     -7.323e-01  4.607e-01  -1.590  0.11190
## Age_Group30s                      3.465e-01  4.105e-01   0.844  0.39857
## Age_Group40s                      1.580e+00  3.903e-01   4.047 5.19e-05
## Age_Group50s                      2.677e+00  3.840e-01   6.970 3.16e-12
## Age_Group60s                      3.995e+00  3.827e-01  10.439  < 2e-16
## Age_Group70s                      5.264e+00  3.829e-01  13.747  < 2e-16
## Age_Group80s                      5.758e+00  3.830e-01  15.035  < 2e-16
## Age_Group90s                      6.137e+00  3.840e-01  15.983  < 2e-16
## Age_GroupUNKNOWN                 -1.139e+01  8.342e+02  -0.014  0.98910
## Case_AcquisitionInfoNo Epi-link   6.265e-01  6.271e-02   9.991  < 2e-16
## Case_AcquisitionInfoNo Info-Missing 8.431e-01 1.530e-01   5.511 3.57e-08
## Case_AcquisitionInfoNo Info-Unk   1.581e+00  1.185e-01  13.334  < 2e-16
## Case_AcquisitionInfoOB            1.398e+01  6.822e+02   0.020  0.98365
## Case_AcquisitionInfoTravel        2.371e-01  9.271e-02   2.557  0.01056
## Outbreak_RelatedNo                1.307e+01  6.822e+02   0.019  0.98472
## Reporting_PHU_CityBelleville     -1.997e-01  3.854e-01  -0.518  0.60428
## Reporting_PHU_CityBrantford       4.003e-01  3.172e-01   1.262  0.20690
## Reporting_PHU_CityBrockville      1.303e-01  1.742e-01   0.748  0.45449
## Reporting_PHU_CityChatham        -3.602e-01  4.483e-01  -0.803  0.42169
## Reporting_PHU_CityCornwall       -2.601e-01  2.778e-01  -0.936  0.34908
## Reporting_PHU_CityGuelph          1.114e-02  1.825e-01   0.061  0.95136
## Reporting_PHU_CityHamilton       -7.015e-02  1.702e-01  -0.412  0.68028
```

```
## Reporting_PHU_CityKenora            -1.420e+01  3.353e+02  -0.042  0.96621
## Reporting_PHU_CityKingston          -1.345e+01  1.719e+02  -0.078  0.93763
## Reporting_PHU_CityLondon             9.199e-01  1.751e-01   5.254 1.49e-07
## Reporting_PHU_CityMississauga        3.339e-01  1.333e-01   2.505  0.01225
## Reporting_PHU_CityNew Liskeard      -1.426e+01  4.206e+02  -0.034  0.97295
## Reporting_PHU_CityNewmarket          4.455e-01  1.369e-01   3.254  0.00114
## Reporting_PHU_CityNorth Bay          8.609e-01  5.667e-01   1.519  0.12872
## Reporting_PHU_CityOakville          -1.623e-01  1.909e-01  -0.850  0.39529
## Reporting_PHU_CityOttawa             5.878e-01  1.394e-01   4.216 2.49e-05
## Reporting_PHU_CityOwen Sound        -1.489e+01  1.439e+02  -0.103  0.91760
## Reporting_PHU_CityPembroke          -5.276e-01  8.331e-01  -0.633  0.52657
## Reporting_PHU_CityPeterborough      -5.224e-01  4.426e-01  -1.180  0.23794
## Reporting_PHU_CityPoint Edward       1.198e-01  2.116e-01   0.566  0.57113
## Reporting_PHU_CityPort Hope         -6.295e-02  2.156e-01  -0.292  0.77032
## Reporting_PHU_CitySault Ste. Marie  -1.396e+01  2.902e+02  -0.048  0.96163
## Reporting_PHU_CitySimcoe             4.881e-01  1.971e-01   2.476  0.01329
## Reporting_PHU_CitySt. Thomas         4.617e-01  3.305e-01   1.397  0.16239
## Reporting_PHU_CityStratford          8.231e-01  4.319e-01   1.906  0.05670
## Reporting_PHU_CitySudbury            5.871e-03  4.279e-01   0.014  0.98905
## Reporting_PHU_CityThorold            8.625e-02  1.639e-01   0.526  0.59872
## Reporting_PHU_CityThunder Bay       -5.812e-02  5.120e-01  -0.114  0.90962
## Reporting_PHU_CityTimmins            1.469e+00  3.266e-01   4.500 6.81e-06
## Reporting_PHU_CityToronto            2.679e-01  1.279e-01   2.094  0.03628
## Reporting_PHU_CityWaterloo           3.385e-01  1.489e-01   2.273  0.02300
## Reporting_PHU_CityWhitby             3.321e-01  1.430e-01   2.322  0.02026
## Reporting_PHU_CityWindsor            2.233e-01  1.582e-01   1.411  0.15821
##
## (Intercept)
## Client_GenderMALE                  ***
## Client_GenderOTHER
## Client_GenderTRANSGENDER
## Client_GenderUNKNOWN
## Age_Group20s
## Age_Group30s
## Age_Group40s                       ***
## Age_Group50s                       ***
## Age_Group60s                       ***
## Age_Group70s                       ***
## Age_Group80s                       ***
## Age_Group90s                       ***
## Age_GroupUNKNOWN
## Case_AcquisitionInfoNo Epi-link    ***
## Case_AcquisitionInfoNo Info-Missing ***
## Case_AcquisitionInfoNo Info-Unk    ***
## Case_AcquisitionInfoOB
## Case_AcquisitionInfoTravel         *
## Outbreak_RelatedNo
## Reporting_PHU_CityBelleville
## Reporting_PHU_CityBrantford
## Reporting_PHU_CityBrockville
```

```
## Reporting_PHU_CityChatham
## Reporting_PHU_CityCornwall
## Reporting_PHU_CityGuelph
## Reporting_PHU_CityHamilton
## Reporting_PHU_CityKenora
## Reporting_PHU_CityKingston
## Reporting_PHU_CityLondon                ***
## Reporting_PHU_CityMississauga           *
## Reporting_PHU_CityNew Liskeard
## Reporting_PHU_CityNewmarket             **
## Reporting_PHU_CityNorth Bay
## Reporting_PHU_CityOakville
## Reporting_PHU_CityOttawa                ***
## Reporting_PHU_CityOwen Sound
## Reporting_PHU_CityPembroke
## Reporting_PHU_CityPeterborough
## Reporting_PHU_CityPoint Edward
## Reporting_PHU_CityPort Hope
## Reporting_PHU_CitySault Ste. Marie
## Reporting_PHU_CitySimcoe                *
## Reporting_PHU_CitySt. Thomas
## Reporting_PHU_CityStratford             .
## Reporting_PHU_CitySudbury
## Reporting_PHU_CityThorold
## Reporting_PHU_CityThunder Bay
## Reporting_PHU_CityTimmins               ***
## Reporting_PHU_CityToronto               *
## Reporting_PHU_CityWaterloo              *
## Reporting_PHU_CityWhitby                *
## Reporting_PHU_CityWindsor
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43617  on 31463  degrees of freedom
## Residual deviance: 23564  on 31411  degrees of freedom
## AIC: 23670
##
## Number of Fisher Scoring iterations: 14

#ANOVA test
anova(model.LogReg, test="Chisq")

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Outcome1
##
```

```
## Terms added sequentially (first to last)
##
##
##                     Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                 31463      43617
## Client_Gender         4     72.7     31459      43545 6.189e-15 ***
## Age_Group             9  19308.4     31450      24236 < 2.2e-16 ***
## Case_AcquisitionInfo  5    436.4     31445      23800 < 2.2e-16 ***
## Outbreak_Related      1      1.3     31444      23798    0.2541
## Reporting_PHU_City   33    234.7     31411      23564 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Predict on test data
test.probs <- predict(model.LogReg, test, type = "response")
pred.log <- rep("Resolved", length(test.probs))
pred.log[test.probs>=0.5] <- "Fatal"
```

```
#Confusion Matrix
confusionMatrix(factor(pred.log),factor(test$Outcome1))
```

```
## Warning in confusionMatrix.default(factor(pred.log), factor(test
## $Outcome1)): Levels are not in the same order for reference and data.
## Refactoring data to match.
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Resolved Fatal
##   Resolved     1413   162
##   Fatal         345  1575
##
##                Accuracy : 0.8549
##                  95% CI : (0.8428, 0.8665)
##     No Information Rate : 0.503
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.71
##
##  Mcnemar's Test P-Value : 6.324e-16
##
##             Sensitivity : 0.8038
##             Specificity : 0.9067
##          Pos Pred Value : 0.8971
##          Neg Pred Value : 0.8203
##              Prevalence : 0.5030
##          Detection Rate : 0.4043
##    Detection Prevalence : 0.4506
##       Balanced Accuracy : 0.8552
##
```

```
##          'Positive' Class : Resolved
##
```

```r
#Recall = 0.9067
#Precision = 0.8203

#ROC Curve
#install.packages("pROC")
library(pROC)
```

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

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
roc.curve <- roc(test$Outcome1, test.probs)
```

```
## Setting levels: control = Resolved, case = Fatal

## Setting direction: controls < cases
```

```r
print(roc.curve)
```

```
##
## Call:
## roc.default(response = test$Outcome1, predictor = test.probs)
##
## Data: test.probs in 1758 controls (test$Outcome1 Resolved) < 1737 cases
(test$Outcome1 Fatal).
## Area under the curve: 0.9048
```

```r
plot(roc.curve, ylim=c(0,1), print.thres=TRUE)
abline(h=1,col='blue',lwd=2)
abline(h=0,col='red',lwd=2)
```
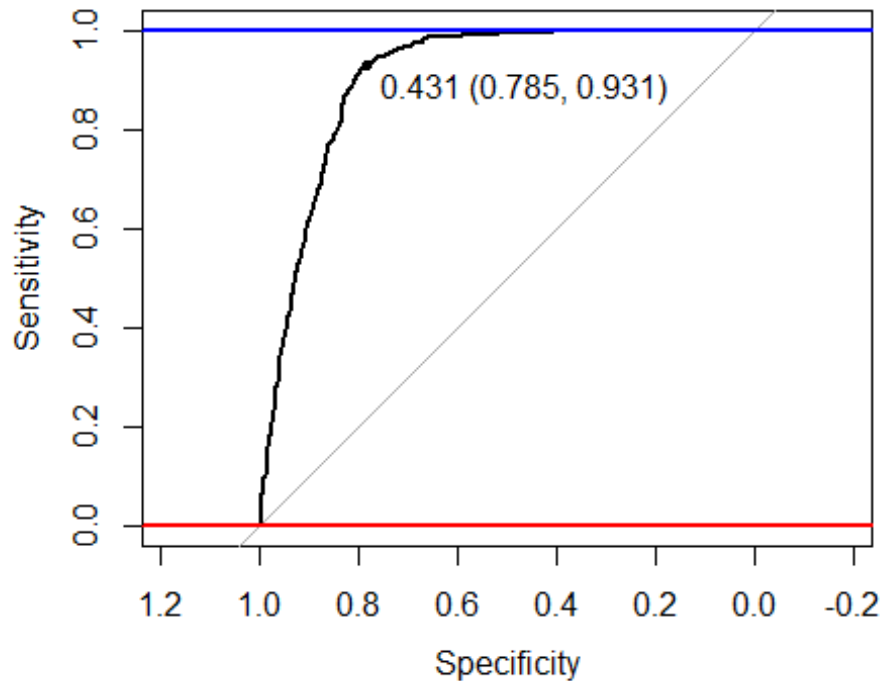
Step 4: Modeling - (2) Naive Bayes Classifier

```
#install.packages("e1071")
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 3.5.3
```

```
#install.packages("gmodels")
library(gmodels)
```

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

```
##
## Attaching package: 'gmodels'
```

```
## The following object is masked from 'package:pROC':
##
##     ci
```

```
#Building model on training set
NBC.classifier <- naiveBayes(training, training$Outcome1, laplace = 1)
NBC.classifier
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
```

```
## naiveBayes.default(x = training, y = training$Outcome1, laplace = 1)
##
## A-priori probabilities:
## training$Outcome1
##  Resolved      Fatal
## 0.5030193 0.4969807
##
## Conditional probabilities:
##                   Row_ID
## training$Outcome1     [,1]      [,2]
##          Resolved 18287.46 10894.11
##          Fatal    14611.89 10849.37
##
##                   Accurate_Episode_Date
## training$Outcome1   2020-01-01   2020-01-10   2020-01-21   2020-01-22
##          Resolved 1.251486e-04 1.251486e-04 1.251486e-04 6.257431e-05
##          Fatal    6.332721e-05 6.332721e-05 6.332721e-05 6.332721e-05
##                   Accurate_Episode_Date
## training$Outcome1   2020-01-24   2020-02-01   2020-02-05   2020-02-07
##          Resolved 1.877229e-04 6.257431e-05 1.251486e-04 1.251486e-04
##          Fatal    6.332721e-05 6.332721e-05 6.332721e-05 6.332721e-05
##                   Accurate_Episode_Date
## training$Outcome1   2020-02-10   2020-02-14   2020-02-15   2020-02-16
##          Resolved 1.251486e-04 6.257431e-05 6.257431e-05 6.257431e-05
##          Fatal    6.332721e-05 6.332721e-05 6.332721e-05 6.332721e-05
##                   Accurate_Episode_Date
## training$Outcome1   2020-02-17   2020-02-19   2020-02-20   2020-02-21
##          Resolved 6.257431e-05 6.257431e-05 1.877229e-04 1.877229e-04
##          Fatal    6.332721e-05 6.332721e-05 6.332721e-05 6.332721e-05
##                   Accurate_Episode_Date
## training$Outcome1   2020-02-22   2020-02-23   2020-02-24   2020-02-25
##          Resolved 2.502972e-04 6.257431e-05 1.251486e-04 6.257431e-05
##          Fatal    6.332721e-05 6.332721e-05 6.332721e-05 6.332721e-05
##                   Accurate_Episode_Date
## training$Outcome1   2020-02-26   2020-02-27   2020-02-28   2020-02-29
##          Resolved 6.257431e-05 1.251486e-04 1.251486e-04 3.754458e-04
##          Fatal    6.332721e-05 6.332721e-05 6.332721e-05 6.332721e-05
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-01   2020-03-02   2020-03-03   2020-03-04
##          Resolved 5.631688e-04 6.257431e-04 5.631688e-04 4.380201e-04
##          Fatal    1.013235e-03 6.332721e-05 6.332721e-05 5.066177e-04
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-05   2020-03-06   2020-03-07   2020-03-08
##          Resolved 1.063763e-03 1.188912e-03 8.134660e-04 8.760403e-04
##          Fatal    5.699449e-04 1.139890e-03 1.013235e-03 1.076563e-03
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-09   2020-03-10   2020-03-11   2020-03-12
##          Resolved 1.814655e-03 2.628121e-03 3.316438e-03 4.129904e-03
##          Fatal    9.499082e-04 1.773162e-03 1.013235e-03 3.166361e-04
##                   Accurate_Episode_Date
```

```
## training$Outcome1   2020-03-13   2020-03-14   2020-03-15   2020-03-16
##          Resolved 3.504161e-03 4.442776e-03 6.194856e-03 7.383768e-03
##          Fatal    2.216452e-03 3.356342e-03 4.052942e-03 4.812868e-03
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-17   2020-03-18   2020-03-19   2020-03-20
##          Resolved 8.197234e-03 8.822977e-03 7.821788e-03 9.135849e-03
##          Fatal    5.952758e-03 5.192831e-03 6.712684e-03 7.789247e-03
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-21   2020-03-22   2020-03-23   2020-03-24
##          Resolved 7.634065e-03 7.508917e-03 1.101308e-02 6.883174e-03
##          Fatal    5.319486e-03 7.282629e-03 7.852574e-03 7.155975e-03
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-25   2020-03-26   2020-03-27   2020-03-28
##          Resolved 8.697829e-03 6.883174e-03 7.696640e-03 8.697829e-03
##          Fatal    8.105883e-03 1.513520e-02 1.006903e-02 1.089228e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-03-29   2020-03-30   2020-03-31   2020-04-01
##          Resolved 6.820599e-03 1.088793e-02 8.760403e-03 1.176397e-02
##          Fatal    1.285542e-02 1.899816e-02 9.625736e-03 2.938383e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-02   2020-04-03   2020-04-04   2020-04-05
##          Resolved 1.032476e-02 1.270258e-02 1.076278e-02 8.572680e-03
##          Fatal    1.925147e-02 2.140460e-02 2.083465e-02 1.925147e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-06   2020-04-07   2020-04-08   2020-04-09
##          Resolved 1.170140e-02 1.245229e-02 1.070021e-02 1.126338e-02
##          Fatal    2.400101e-02 2.007473e-02 2.767399e-02 1.988474e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-10   2020-04-11   2020-04-12   2020-04-13
##          Resolved 1.432952e-02 1.695764e-02 1.376635e-02 1.683249e-02
##          Fatal    2.317776e-02 2.932050e-02 2.469761e-02 2.906719e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-14   2020-04-15   2020-04-16   2020-04-17
##          Resolved 1.883487e-02 1.733308e-02 1.620675e-02 1.839685e-02
##          Fatal    2.324109e-02 2.944715e-02 2.925717e-02 3.001710e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-18   2020-04-19   2020-04-20   2020-04-21
##          Resolved 1.389150e-02 9.886741e-03 1.345348e-02 9.448720e-03
##          Fatal    2.552087e-02 1.817491e-02 1.906149e-02 1.646508e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-22   2020-04-23   2020-04-24   2020-04-25
##          Resolved 1.201427e-02 1.132595e-02 1.088793e-02 1.157625e-02
##          Fatal    2.368438e-02 2.007473e-02 1.475524e-02 1.551517e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-26   2020-04-27   2020-04-28   2020-04-29
##          Resolved 1.051248e-02 1.301546e-02 1.145110e-02 1.101308e-02
##          Fatal    1.114559e-02 1.405864e-02 1.120892e-02 1.773162e-02
##                   Accurate_Episode_Date
## training$Outcome1   2020-04-30   2020-05-01   2020-05-02   2020-05-03
##          Resolved 1.145110e-02 1.301546e-02 1.019961e-02 7.446343e-03
```

```
##           Fatal    8.422519e-03 7.472611e-03 7.662593e-03 3.419669e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-04   2020-05-05   2020-05-06   2020-05-07
##           Resolved 8.948126e-03 1.038733e-02 8.885552e-03 1.007446e-02
##           Fatal    6.142740e-03 8.612501e-03 8.929137e-03 8.042556e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-08   2020-05-09   2020-05-10   2020-05-11
##           Resolved 1.051248e-02 7.508917e-03 9.135849e-03 1.157625e-02
##           Fatal    9.245773e-03 1.006903e-02 4.432905e-03 6.839339e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-12   2020-05-13   2020-05-14   2020-05-15
##           Resolved 9.198423e-03 7.634065e-03 1.013704e-02 1.307803e-02
##           Fatal    8.042556e-03 7.472611e-03 4.496232e-03 7.092648e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-16   2020-05-17   2020-05-18   2020-05-19
##           Resolved 1.038733e-02 8.510106e-03 1.163882e-02 8.322383e-03
##           Fatal    6.332721e-03 7.662593e-03 4.432905e-03 7.409284e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-20   2020-05-21   2020-05-22   2020-05-23
##           Resolved 1.195169e-02 8.760403e-03 8.072086e-03 7.696640e-03
##           Fatal    4.812868e-03 5.319486e-03 3.293015e-03 3.293015e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-24   2020-05-25   2020-05-26   2020-05-27
##           Resolved 8.384957e-03 1.163882e-02 1.013704e-02 9.949315e-03
##           Fatal    2.469761e-03 2.849725e-03 3.482997e-03 3.419669e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-05-28   2020-05-29   2020-05-30   2020-05-31
##           Resolved 7.446343e-03 8.822977e-03 8.572680e-03 6.632877e-03
##           Fatal    1.709835e-03 3.356342e-03 1.456526e-03 1.456526e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-06-01   2020-06-02   2020-06-03   2020-06-04
##           Resolved 1.057506e-02 7.133471e-03 7.759214e-03 7.070897e-03
##           Fatal    2.089798e-03 1.456526e-03 1.013235e-03 1.709835e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-06-05   2020-06-06   2020-06-07   2020-06-08
##           Resolved 6.257431e-03 4.567924e-03 4.505350e-03 6.069708e-03
##           Fatal    2.406434e-03 1.773162e-03 9.499082e-04 2.533088e-04
##                     Accurate_Episode_Date
## training$Outcome1   2020-06-09   2020-06-10   2020-06-11   2020-06-12
##           Resolved 4.255053e-03 6.883174e-03 6.069708e-03 6.069708e-03
##           Fatal    1.456526e-03 6.965993e-04 8.865810e-04 6.332721e-05
##                     Accurate_Episode_Date
## training$Outcome1   2020-06-13   2020-06-14   2020-06-15   2020-06-16
##           Resolved 4.630499e-03 4.567924e-03 6.007133e-03 5.819411e-03
##           Fatal    6.332721e-04 2.533088e-04 6.965993e-04 2.279780e-03
##                     Accurate_Episode_Date
## training$Outcome1   2020-06-17   2020-06-18   2020-06-19   2020-06-20
##           Resolved 4.630499e-03 4.192479e-03 5.443965e-03 4.880796e-03
##           Fatal    1.013235e-03 2.026471e-03 6.332721e-05 4.432905e-04
##                     Accurate_Episode_Date
```

```
## training$Outcome1    2020-06-21    2020-06-22    2020-06-23    2020-06-24
##         Resolved 4.004756e-03 5.193667e-03 4.630499e-03 5.443965e-03
##         Fatal    6.965993e-04 1.013235e-03 2.533088e-04 3.799633e-04
##                  Accurate_Episode_Date
## training$Outcome1    2020-06-25    2020-06-26    2020-06-27    2020-06-28
##         Resolved 9.949315e-03 2.878418e-03 3.066141e-03 5.631688e-04
##         Fatal    1.329871e-03 4.432905e-04 9.499082e-04 6.332721e-05
##                  Accurate_Episode_Date
## training$Outcome1    2020-06-29    2020-06-30    2020-07-01    2020-07-02
##         Resolved 3.128715e-04 2.502972e-04 6.257431e-05 1.877229e-04
##         Fatal    5.699449e-04 6.332721e-05 6.332721e-05 5.066177e-04
##                  Accurate_Episode_Date
## training$Outcome1    2020-07-03    2020-07-04    2020-07-05    2020-07-06
##         Resolved 6.257431e-05 1.251486e-04 6.257431e-05 1.251486e-04
##         Fatal    6.332721e-05 6.332721e-05 6.332721e-05 2.533088e-04
##                  Accurate_Episode_Date
## training$Outcome1    2020-07-07    2020-07-08
##         Resolved 6.257431e-05 1.877229e-04
##         Fatal    6.332721e-05 6.332721e-05
##
##                  Case_Reported_Date
## training$Outcome1    2020-01-23    2020-01-24    2020-02-21    2020-02-25
##         Resolved 1.252505e-04 1.878758e-04 1.252505e-04 6.262525e-05
##         Fatal    6.337939e-05 6.337939e-05 6.337939e-05 6.337939e-05
##                  Case_Reported_Date
## training$Outcome1    2020-02-26    2020-02-27    2020-02-28    2020-02-29
##         Resolved 6.262525e-05 6.262525e-05 1.252505e-04 1.878758e-04
##         Fatal    6.337939e-05 6.337939e-05 6.337939e-05 6.337939e-05
##                  Case_Reported_Date
## training$Outcome1    2020-03-01    2020-03-03    2020-03-04    2020-03-05
##         Resolved 6.262525e-05 1.878758e-04 1.878758e-04 1.878758e-04
##         Fatal    6.337939e-05 6.337939e-05 6.337939e-05 6.337939e-05
##                  Case_Reported_Date
## training$Outcome1    2020-03-06    2020-03-07    2020-03-08    2020-03-09
##         Resolved 1.878758e-04 6.262525e-05 1.878758e-04 1.252505e-04
##         Fatal    6.337939e-05 6.337939e-05 3.168969e-04 6.337939e-05
##                  Case_Reported_Date
## training$Outcome1    2020-03-10    2020-03-11    2020-03-12    2020-03-13
##         Resolved 1.252505e-04 6.888778e-04 1.189880e-03 8.767535e-04
##         Fatal    6.337939e-05 6.337939e-05 5.070351e-04 7.605527e-04
##                  Case_Reported_Date
## training$Outcome1    2020-03-14    2020-03-15    2020-03-16    2020-03-17
##         Resolved 1.002004e-03 1.002004e-03 2.066633e-03 1.878758e-03
##         Fatal    6.337939e-05 6.337939e-05 5.704145e-04 6.971733e-04
##                  Case_Reported_Date
## training$Outcome1    2020-03-18    2020-03-19    2020-03-20    2020-03-21
##         Resolved 1.565631e-03 2.129259e-03 2.004008e-03 2.191884e-03
##         Fatal    8.239321e-04 7.605527e-04 2.028140e-03 3.612625e-03
##                  Case_Reported_Date
## training$Outcome1    2020-03-22    2020-03-23    2020-03-24    2020-03-25
```

```
##            Resolved 2.755511e-03 4.509018e-03 4.759519e-03 3.193888e-03
##            Fatal    2.091520e-03 2.915452e-03 2.725314e-03 4.436557e-03
##                      Case_Reported_Date
## training$Outcome1   2020-03-26  2020-03-27  2020-03-28  2020-03-29
##            Resolved 6.826152e-03 8.329158e-03 6.575651e-03 6.325150e-03
##            Fatal    4.880213e-03 6.591456e-03 5.830904e-03 8.809735e-03
##                      Case_Reported_Date
## training$Outcome1   2020-03-30  2020-03-31  2020-04-01  2020-04-02
##            Resolved 1.321393e-02 1.102204e-02 1.290080e-02 1.296343e-02
##            Fatal    8.492838e-03 1.311953e-02 1.090125e-02 2.028140e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-03  2020-04-04  2020-04-05  2020-04-06
##            Resolved 1.164830e-02 9.706914e-03 1.058367e-02 1.114729e-02
##            Fatal    2.877424e-02 1.349981e-02 1.761947e-02 3.143618e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-07  2020-04-08  2020-04-09  2020-04-10
##            Resolved 1.221192e-02 1.202405e-02 1.283818e-02 1.020792e-02
##            Fatal    2.592217e-02 2.376727e-02 1.736595e-02 2.104196e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-11  2020-04-12  2020-04-13  2020-04-14
##            Resolved 1.277555e-02 1.358968e-02 1.672094e-02 1.596944e-02
##            Fatal    2.585879e-02 2.915452e-02 2.712638e-02 2.997845e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-15  2020-04-16  2020-04-17  2020-04-18
##            Resolved 1.534319e-02 1.603206e-02 1.985220e-02 1.647044e-02
##            Fatal    3.308404e-02 2.940804e-02 3.948536e-02 2.890100e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-19  2020-04-20  2020-04-21  2020-04-22
##            Resolved 1.515531e-02 1.734719e-02 1.434118e-02 1.584419e-02
##            Fatal    2.262644e-02 1.888706e-02 2.224617e-02 2.573203e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-23  2020-04-24  2020-04-25  2020-04-26
##            Resolved 1.352705e-02 1.120992e-02 1.202405e-02 1.008267e-02
##            Fatal    2.199265e-02 1.527443e-02 2.072506e-02 1.280264e-02
##                      Case_Reported_Date
## training$Outcome1   2020-04-27  2020-04-28  2020-04-29  2020-04-30
##            Resolved 1.196142e-02 1.077154e-02 1.146042e-02 1.120992e-02
##            Fatal    1.597161e-02 1.438712e-02 1.470402e-02 1.527443e-02
##                      Case_Reported_Date
## training$Outcome1   2020-05-01  2020-05-02  2020-05-03  2020-05-04
##            Resolved 1.277555e-02 1.064629e-02 9.393788e-03 9.456413e-03
##            Fatal    1.571809e-02 7.352009e-03 8.619597e-03 7.478768e-03
##                      Case_Reported_Date
## training$Outcome1   2020-05-05  2020-05-06  2020-05-07  2020-05-08
##            Resolved 1.083417e-02 1.070892e-02 9.644289e-03 8.767535e-03
##            Fatal    8.873114e-03 1.026746e-02 5.957663e-03 6.528077e-03
##                      Case_Reported_Date
## training$Outcome1   2020-05-09  2020-05-10  2020-05-11  2020-05-12
##            Resolved 8.642285e-03 1.058367e-02 5.761523e-03 6.638277e-03
##            Fatal    8.809735e-03 1.172519e-02 5.197110e-03 7.542147e-03
```

```
##                 Case_Reported_Date
## training$Outcome1   2020-05-13   2020-05-14   2020-05-15   2020-05-16
##          Resolved 1.164830e-02 9.894790e-03 1.108467e-02 1.014529e-02
##          Fatal    9.380150e-03 1.052098e-02 1.033084e-02 6.211180e-03
##                 Case_Reported_Date
## training$Outcome1   2020-05-17   2020-05-18   2020-05-19   2020-05-20
##          Resolved 1.152305e-02 9.080661e-03 9.268537e-03 1.039579e-02
##          Fatal    8.239321e-03 5.830904e-03 4.373178e-03 8.492838e-03
##                 Case_Reported_Date
## training$Outcome1   2020-05-21   2020-05-22   2020-05-23   2020-05-24
##          Resolved 9.957415e-03 1.064629e-02 1.171092e-02 9.519038e-03
##          Fatal    3.422487e-03 8.112562e-03 6.718215e-03 3.929522e-03
##                 Case_Reported_Date
## training$Outcome1   2020-05-25   2020-05-26   2020-05-27   2020-05-28
##          Resolved 1.127255e-02 8.203908e-03 1.008267e-02 9.519038e-03
##          Fatal    4.183040e-03 2.725314e-03 1.964761e-03 4.690075e-03
##                 Case_Reported_Date
## training$Outcome1   2020-05-29   2020-05-30   2020-05-31   2020-06-01
##          Resolved 1.922595e-02 1.095942e-02 1.177355e-02 1.058367e-02
##          Fatal    6.021042e-03 4.563316e-03 3.549246e-03 1.647864e-03
##                 Case_Reported_Date
## training$Outcome1   2020-06-02   2020-06-03   2020-06-04   2020-06-05
##          Resolved 1.033317e-02 1.083417e-02 8.892786e-03 8.830160e-03
##          Fatal    6.337939e-05 2.978831e-03 1.584485e-03 1.711244e-03
##                 Case_Reported_Date
## training$Outcome1   2020-06-06   2020-06-07   2020-06-08   2020-06-09
##          Resolved 6.826152e-03 6.074649e-03 6.012024e-03 6.012024e-03
##          Fatal    1.711244e-03 3.168969e-04 1.901382e-03 1.838002e-03
##                 Case_Reported_Date
## training$Outcome1   2020-06-10   2020-06-11   2020-06-12   2020-06-13
##          Resolved 5.886774e-03 4.571643e-03 7.640281e-03 6.387776e-03
##          Fatal    1.330967e-03 2.218279e-03 5.070351e-04 8.873114e-04
##                 Case_Reported_Date
## training$Outcome1   2020-06-14   2020-06-15   2020-06-16   2020-06-17
##          Resolved 3.694890e-03 5.949399e-03 5.135271e-03 6.826152e-03
##          Fatal    6.337939e-05 7.605527e-04 9.506908e-04 1.711244e-03
##                 Case_Reported_Date
## training$Outcome1   2020-06-18   2020-06-19   2020-06-20   2020-06-21
##          Resolved 6.638277e-03 4.822144e-03 3.945391e-03 5.949399e-03
##          Fatal    2.598555e-03 1.140829e-03 1.330967e-03 1.267588e-03
##                 Case_Reported_Date
## training$Outcome1   2020-06-22   2020-06-23   2020-06-24   2020-06-25
##          Resolved 5.260521e-03 4.571643e-03 4.759519e-03 4.321142e-03
##          Fatal    1.077450e-03 5.070351e-04 6.971733e-04 7.605527e-04
##                 Case_Reported_Date
## training$Outcome1   2020-06-26   2020-06-27   2020-06-28   2020-06-29
##          Resolved 5.010020e-03 7.640281e-03 7.014028e-03 3.319138e-03
##          Fatal    9.506908e-04 6.337939e-05 1.584485e-03 6.337939e-05
##                 Case_Reported_Date
## training$Outcome1   2020-06-30   2020-07-01   2020-07-02   2020-07-03
```

```
##               Resolved 3.193888e-03 2.129259e-03 1.628257e-03 1.189880e-03
##               Fatal    2.535176e-04 6.337939e-05 1.457726e-03 6.337939e-05
##                        Case_Reported_Date
## training$Outcome1  2020-07-04   2020-07-05   2020-07-06   2020-07-07
##               Resolved 1.565631e-03 7.515030e-04 4.383768e-04 3.757515e-04
##               Fatal    5.070351e-04 6.337939e-05 6.337939e-05 6.337939e-05
##                        Case_Reported_Date
## training$Outcome1  2020-07-08   2020-07-09   2020-07-10   2020-07-11
##               Resolved 2.505010e-04 3.131263e-04 2.505010e-04 6.262525e-05
##               Fatal    6.337939e-05 6.337939e-05 2.535176e-04 6.337939e-05
##                        Case_Reported_Date
## training$Outcome1  2020-07-12
##               Resolved 6.262525e-05
##               Fatal    6.337939e-05
##
##                        Test_Reported_Date
## training$Outcome1  2020-01-27   2020-02-03   2020-02-24   2020-02-25
##               Resolved 1.252583e-04 1.878875e-04 1.252583e-04 6.262917e-05
##               Fatal    6.338341e-05 6.338341e-05 6.338341e-05 6.338341e-05
##                        Test_Reported_Date
## training$Outcome1  2020-02-27   2020-02-28   2020-02-29   2020-03-01
##               Resolved 6.262917e-05 1.252583e-04 1.878875e-04 2.505167e-04
##               Fatal    6.338341e-05 6.338341e-05 6.338341e-05 6.338341e-05
##                        Test_Reported_Date
## training$Outcome1  2020-03-02   2020-03-03   2020-03-04   2020-03-05
##               Resolved 6.262917e-05 6.262917e-05 1.878875e-04 1.878875e-04
##               Fatal    6.338341e-05 6.338341e-05 6.338341e-05 6.338341e-05
##                        Test_Reported_Date
## training$Outcome1  2020-03-07   2020-03-08   2020-03-09   2020-03-10
##               Resolved 6.262917e-05 6.262917e-05 2.505167e-04 1.252583e-04
##               Fatal    6.338341e-05 6.338341e-05 6.338341e-05 6.338341e-05
##                        Test_Reported_Date
## training$Outcome1  2020-03-11   2020-03-12   2020-03-13   2020-03-14
##               Resolved 3.757750e-04 5.636626e-04 8.768084e-04 1.252583e-03
##               Fatal    6.338341e-05 6.338341e-05 1.204285e-03 6.338341e-05
##                        Test_Reported_Date
## training$Outcome1  2020-03-15   2020-03-16   2020-03-17   2020-03-18
##               Resolved 6.262917e-04 8.141792e-04 1.690988e-03 1.565729e-03
##               Fatal    6.338341e-05 5.704507e-04 1.204285e-03 1.077518e-03
##                        Test_Reported_Date
## training$Outcome1  2020-03-19   2020-03-20   2020-03-21   2020-03-22
##               Resolved 1.565729e-03 2.379909e-03 1.315213e-03 2.630425e-03
##               Fatal    6.338341e-05 8.239843e-04 3.359321e-03 1.838119e-03
##                        Test_Reported_Date
## training$Outcome1  2020-03-23   2020-03-24   2020-03-25   2020-03-26
##               Resolved 4.509300e-03 4.822446e-03 4.321413e-03 7.139726e-03
##               Fatal    2.725486e-03 9.507511e-04 5.387590e-03 4.500222e-03
##                        Test_Reported_Date
## training$Outcome1  2020-03-27   2020-03-28   2020-03-29   2020-03-30
##               Resolved 7.640759e-03 5.699255e-03 6.889209e-03 1.415419e-02
```

```
##            Fatal      6.084807e-03 4.310072e-03 6.465107e-03 1.140901e-02
##                        Test_Reported_Date
## training$Outcome1   2020-03-31   2020-04-01   2020-04-02   2020-04-03
##           Resolved 1.045907e-02 1.196217e-02 1.233795e-02 1.114799e-02
##            Fatal    1.331052e-02 1.267668e-02 2.193066e-02 2.674780e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-04   2020-04-05   2020-04-06   2020-04-07
##           Resolved 1.033381e-02 9.770151e-03 1.020856e-02 1.240058e-02
##            Fatal    2.408569e-02 1.705014e-02 2.548013e-02 2.845915e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-08   2020-04-09   2020-04-10   2020-04-11
##           Resolved 1.334001e-02 1.371579e-02 1.052170e-02 1.227532e-02
##            Fatal    1.895164e-02 2.155036e-02 2.161374e-02 3.238892e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-12   2020-04-13   2020-04-14   2020-04-15
##           Resolved 1.371579e-02 1.628358e-02 1.722302e-02 1.490574e-02
##            Fatal    1.787412e-02 2.516321e-02 2.839577e-02 3.270584e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-16   2020-04-17   2020-04-18   2020-04-19
##           Resolved 1.640884e-02 1.722302e-02 1.622096e-02 1.471786e-02
##            Fatal    3.213539e-02 3.745959e-02 3.105787e-02 2.078976e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-20   2020-04-21   2020-04-22   2020-04-23
##           Resolved 1.841298e-02 1.471786e-02 1.484311e-02 1.446734e-02
##            Fatal    2.059961e-02 2.212081e-02 2.478291e-02 2.345186e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-24   2020-04-25   2020-04-26   2020-04-27
##           Resolved 1.227532e-02 1.409156e-02 9.206488e-03 1.196217e-02
##            Fatal    1.793750e-02 2.433923e-02 1.216961e-02 1.762059e-02
##                        Test_Reported_Date
## training$Outcome1   2020-04-28   2020-04-29   2020-04-30   2020-05-01
##           Resolved 1.033381e-02 1.139851e-02 1.139851e-02 1.233795e-02
##            Fatal    1.083856e-02 1.388097e-02 1.508525e-02 1.413450e-02
##                        Test_Reported_Date
## training$Outcome1   2020-05-02   2020-05-03   2020-05-04   2020-05-05
##           Resolved 1.033381e-02 1.014593e-02 9.958038e-03 1.133588e-02
##            Fatal    8.239843e-03 8.873677e-03 6.845408e-03 1.033150e-02
##                        Test_Reported_Date
## training$Outcome1   2020-05-06   2020-05-07   2020-05-08   2020-05-09
##           Resolved 9.895409e-03 9.206488e-03 9.081230e-03 8.204422e-03
##            Fatal    8.620143e-03 9.190594e-03 4.626989e-03 1.033150e-02
##                        Test_Reported_Date
## training$Outcome1   2020-05-10   2020-05-11   2020-05-12   2020-05-13
##           Resolved 1.052170e-02 6.388176e-03 7.515501e-03 1.108536e-02
##            Fatal    7.162325e-03 6.148190e-03 7.606009e-03 6.655258e-03
##                        Test_Reported_Date
## training$Outcome1   2020-05-14   2020-05-15   2020-05-16   2020-05-17
##           Resolved 1.020856e-02 1.058433e-02 9.958038e-03 1.177428e-02
##            Fatal    1.134563e-02 1.064841e-02 6.465107e-03 7.859542e-03
##                        Test_Reported_Date
```

```
## training$Outcome1   2020-05-18   2020-05-19   2020-05-20   2020-05-21
##          Resolved 9.081230e-03 8.768084e-03 1.002067e-02 1.027118e-02
##          Fatal    4.817139e-03 6.211574e-03 7.479242e-03 5.007289e-03
##                   Test_Reported_Date
## training$Outcome1   2020-05-22   2020-05-23   2020-05-24   2020-05-25
##          Resolved 1.058433e-02 1.215006e-02 9.958038e-03 1.484311e-02
##          Fatal    6.782024e-03 7.415859e-03 4.880522e-03 6.528491e-03
##                   Test_Reported_Date
## training$Outcome1   2020-05-26   2020-05-27   2020-05-28   2020-05-29
##          Resolved 7.953905e-03 8.705455e-03 9.394376e-03 1.841298e-02
##          Fatal    6.972175e-04 1.964886e-03 4.690372e-03 6.401724e-03
##                   Test_Reported_Date
## training$Outcome1   2020-05-30   2020-05-31   2020-06-01   2020-06-02
##          Resolved 1.315213e-02 9.707522e-03 9.958038e-03 1.102273e-02
##          Fatal    4.690372e-03 2.535336e-03 2.915637e-03 5.704507e-04
##                   Test_Reported_Date
## training$Outcome1   2020-06-03   2020-06-04   2020-06-05   2020-06-06
##          Resolved 1.133588e-02 9.018601e-03 8.329680e-03 6.513434e-03
##          Fatal    2.852253e-03 1.267668e-03 1.711352e-03 1.711352e-03
##                   Test_Reported_Date
## training$Outcome1   2020-06-07   2020-06-08   2020-06-09   2020-06-10
##          Resolved 6.075030e-03 6.450805e-03 6.325546e-03 5.887142e-03
##          Fatal    6.338341e-05 2.915637e-03 1.394435e-03 1.521202e-03
##                   Test_Reported_Date
## training$Outcome1   2020-06-11   2020-06-12   2020-06-13   2020-06-14
##          Resolved 4.947705e-03 7.264984e-03 6.012401e-03 4.133525e-03
##          Fatal    2.028269e-03 5.070672e-04 8.873677e-04 6.338341e-05
##                   Test_Reported_Date
## training$Outcome1   2020-06-15   2020-06-16   2020-06-17   2020-06-18
##          Resolved 5.949771e-03 5.260851e-03 6.701321e-03 6.701321e-03
##          Fatal    7.606009e-04 1.140901e-03 1.711352e-03 2.408569e-03
##                   Test_Reported_Date
## training$Outcome1   2020-06-19   2020-06-20   2020-06-21   2020-06-22
##          Resolved 5.386109e-03 3.632492e-03 6.262917e-03 5.323480e-03
##          Fatal    1.140901e-03 1.331052e-03 1.267668e-03 1.077518e-03
##                   Test_Reported_Date
## training$Outcome1   2020-06-23   2020-06-24   2020-06-25   2020-06-26
##          Resolved 4.885075e-03 4.759817e-03 4.196155e-03 4.885075e-03
##          Fatal    5.070672e-04 7.606009e-04 6.972175e-04 9.507511e-04
##                   Test_Reported_Date
## training$Outcome1   2020-06-27   2020-06-28   2020-06-29   2020-06-30
##          Resolved 7.390242e-03 7.077097e-03 3.820380e-03 2.693054e-03
##          Fatal    6.338341e-05 1.584585e-03 6.338341e-05 2.535336e-04
##                   Test_Reported_Date
## training$Outcome1   2020-07-01   2020-07-02   2020-07-03   2020-07-04
##          Resolved 2.004134e-03 1.628358e-03 1.377842e-03 1.440471e-03
##          Fatal    6.338341e-05 1.457818e-03 6.338341e-05 5.070672e-04
##                   Test_Reported_Date
## training$Outcome1   2020-07-05   2020-07-06   2020-07-07   2020-07-08
##          Resolved 6.889209e-04 5.010334e-04 3.757750e-04 3.757750e-04
```

```
##           Fatal     6.338341e-05 6.338341e-05 6.338341e-05 6.338341e-05
##                   Test_Reported_Date
## training$Outcome1   2020-07-09   2020-07-10   2020-07-11   2020-07-12
##           Resolved 3.131459e-04 1.878875e-04 6.262917e-05 6.262917e-05
##           Fatal     6.338341e-05 2.535336e-04 6.338341e-05 6.338341e-05
##
##                   Specimen_Date
## training$Outcome1   2020-01-23   2020-01-24   2020-01-25   2020-02-20
##           Resolved 1.252348e-04 1.878522e-04 6.261741e-05 6.261741e-05
##           Fatal     6.337136e-05 6.337136e-05 6.337136e-05 6.337136e-05
##                   Specimen_Date
## training$Outcome1   2020-02-22   2020-02-23   2020-02-25   2020-02-26
##           Resolved 1.252348e-04 6.261741e-05 6.261741e-05 6.261741e-05
##           Fatal     6.337136e-05 6.337136e-05 6.337136e-05 6.337136e-05
##                   Specimen_Date
## training$Outcome1   2020-02-27   2020-02-28   2020-02-29   2020-03-01
##           Resolved 6.261741e-05 2.504696e-04 1.878522e-04 6.261741e-05
##           Fatal     6.337136e-05 6.337136e-05 6.337136e-05 6.337136e-05
##                   Specimen_Date
## training$Outcome1   2020-03-02   2020-03-03   2020-03-04   2020-03-05
##           Resolved 1.252348e-04 2.504696e-04 3.130870e-04 6.261741e-05
##           Fatal     6.337136e-05 6.337136e-05 6.337136e-05 6.337136e-05
##                   Specimen_Date
## training$Outcome1   2020-03-06   2020-03-07   2020-03-08   2020-03-09
##           Resolved 1.252348e-04 6.261741e-05 1.252348e-04 2.504696e-04
##           Fatal     6.337136e-05 6.337136e-05 6.337136e-05 6.337136e-05
##                   Specimen_Date
## training$Outcome1   2020-03-10   2020-03-11   2020-03-12   2020-03-13
##           Resolved 4.383219e-04 9.392611e-04 1.628053e-03 1.377583e-03
##           Fatal     6.337136e-05 5.069708e-04 4.435995e-04 5.703422e-04
##                   Specimen_Date
## training$Outcome1   2020-03-14   2020-03-15   2020-03-16   2020-03-17
##           Resolved 1.502818e-03 1.440200e-03 3.193488e-03 4.007514e-03
##           Fatal     6.337136e-05 9.505703e-04 1.140684e-03 1.457541e-03
##                   Specimen_Date
## training$Outcome1   2020-03-18   2020-03-19   2020-03-20   2020-03-21
##           Resolved 3.443957e-03 5.009393e-03 6.073889e-03 5.134627e-03
##           Fatal     1.964512e-03 2.724968e-03 3.675539e-03 2.281369e-03
##                   Specimen_Date
## training$Outcome1   2020-03-22   2020-03-23   2020-03-24   2020-03-25
##           Resolved 5.322480e-03 7.201002e-03 5.510332e-03 7.138384e-03
##           Fatal     3.422053e-03 6.147022e-03 3.041825e-03 4.562738e-03
##                   Specimen_Date
## training$Outcome1   2020-03-26   2020-03-27   2020-03-28   2020-03-29
##           Resolved 6.762680e-03 6.825297e-03 1.039449e-02 5.635567e-03
##           Fatal     8.365019e-03 6.463878e-03 1.001267e-02 1.020279e-02
##                   Specimen_Date
## training$Outcome1   2020-03-30   2020-03-31   2020-04-01   2020-04-02
##           Resolved 1.120852e-02 1.139637e-02 1.077019e-02 1.214778e-02
##           Fatal     1.356147e-02 1.723701e-02 2.712294e-02 2.503169e-02
```

```
##                Specimen_Date
## training$Outcome1   2020-04-03   2020-04-04   2020-04-05   2020-04-06
##          Resolved 9.079524e-03 8.641202e-03 7.639324e-03 1.421415e-02
##             Fatal 1.926489e-02 2.281369e-02 2.027883e-02 2.775665e-02
##                Specimen_Date
## training$Outcome1   2020-04-07   2020-04-08   2020-04-09   2020-04-10
##          Resolved 1.314966e-02 1.095805e-02 1.333751e-02 1.653100e-02
##             Fatal 2.122940e-02 2.598226e-02 2.674271e-02 2.503169e-02
##                Specimen_Date
## training$Outcome1   2020-04-11   2020-04-12   2020-04-13   2020-04-14
##          Resolved 1.565435e-02 1.095805e-02 1.953663e-02 2.053851e-02
##             Fatal 2.934094e-02 1.844106e-02 3.498099e-02 3.257288e-02
##                Specimen_Date
## training$Outcome1   2020-04-15   2020-04-16   2020-04-17   2020-04-18
##          Resolved 1.978710e-02 1.778334e-02 1.922354e-02 1.415153e-02
##             Fatal 3.314322e-02 3.948035e-02 2.509506e-02 2.585551e-02
##                Specimen_Date
## training$Outcome1   2020-04-19   2020-04-20   2020-04-21   2020-04-22
##          Resolved 1.014402e-02 1.696932e-02 1.252348e-02 1.264872e-02
##             Fatal 8.491762e-03 2.256020e-02 1.812421e-02 2.572877e-02
##                Specimen_Date
## training$Outcome1   2020-04-23   2020-04-24   2020-04-25   2020-04-26
##          Resolved 1.314966e-02 1.327489e-02 1.183469e-02 6.700063e-03
##             Fatal 1.641318e-02 1.679341e-02 1.679341e-02 8.871990e-03
##                Specimen_Date
## training$Outcome1   2020-04-27   2020-04-28   2020-04-29   2020-04-30
##          Resolved 1.371321e-02 1.227301e-02 1.383845e-02 1.252348e-02
##             Fatal 2.072243e-02 1.730038e-02 1.736375e-02 7.287706e-03
##                Specimen_Date
## training$Outcome1   2020-05-01   2020-05-02   2020-05-03   2020-05-04
##          Resolved 1.246086e-02 9.204759e-03 6.199123e-03 1.020664e-02
##             Fatal 1.026616e-02 5.006337e-03 3.738910e-03 8.998733e-03
##                Specimen_Date
## training$Outcome1   2020-05-05   2020-05-06   2020-05-07   2020-05-08
##          Resolved 1.346274e-02 9.830933e-03 1.246086e-02 1.051972e-02
##             Fatal 8.301648e-03 7.984791e-03 1.108999e-02 8.935361e-03
##                Specimen_Date
## training$Outcome1   2020-05-09   2020-05-10   2020-05-11   2020-05-12
##          Resolved 7.952411e-03 6.637445e-03 1.252348e-02 9.893550e-03
##             Fatal 1.305450e-02 5.386565e-03 1.102662e-02 1.121673e-02
##                Specimen_Date
## training$Outcome1   2020-05-13   2020-05-14   2020-05-15   2020-05-16
##          Resolved 9.705698e-03 9.455229e-03 9.329994e-03 7.827176e-03
##             Fatal 6.970849e-03 4.752852e-03 4.435995e-03 5.323194e-03
##                Specimen_Date
## training$Outcome1   2020-05-17   2020-05-18   2020-05-19   2020-05-20
##          Resolved 8.077646e-03 6.512210e-03 1.252348e-02 1.252348e-02
##             Fatal 6.210393e-03 4.499366e-03 6.083650e-03 9.632446e-03
##                Specimen_Date
## training$Outcome1   2020-05-21   2020-05-22   2020-05-23   2020-05-24
```

```
##             Resolved 1.058234e-02 9.830933e-03 7.576706e-03 8.265498e-03
##             Fatal    3.992395e-03 4.372624e-03 5.259823e-03 3.295311e-03
##                      Specimen_Date
## training$Outcome1    2020-05-25   2020-05-26   2020-05-27   2020-05-28
##             Resolved 1.158422e-02 1.202254e-02 1.077019e-02 1.051972e-02
##             Fatal    3.992395e-03 2.915082e-03 3.802281e-03 4.182510e-03
##                      Specimen_Date
## training$Outcome1    2020-05-29   2020-05-30   2020-05-31   2020-06-01
##             Resolved 1.340013e-02 7.764559e-03 6.825297e-03 1.120852e-02
##             Fatal    2.408112e-03 3.231939e-03 8.871990e-04 2.217997e-03
##                      Specimen_Date
## training$Outcome1    2020-06-02   2020-06-03   2020-06-04   2020-06-05
##             Resolved 9.204759e-03 8.328115e-03 8.766437e-03 6.950532e-03
##             Fatal    1.013942e-03 1.267427e-03 2.344740e-03 1.204056e-03
##                      Specimen_Date
## training$Outcome1    2020-06-06   2020-06-07   2020-06-08   2020-06-09
##             Resolved 5.510332e-03 4.633688e-03 5.572949e-03 5.197245e-03
##             Fatal    1.140684e-03 1.394170e-03 8.238276e-04 1.330798e-03
##                      Specimen_Date
## training$Outcome1    2020-06-10   2020-06-11   2020-06-12   2020-06-13
##             Resolved 6.887915e-03 7.764559e-03 5.197245e-03 4.946775e-03
##             Fatal    1.013942e-03 6.337136e-04 9.505703e-04 1.330798e-03
##                      Specimen_Date
## training$Outcome1    2020-06-14   2020-06-15   2020-06-16   2020-06-17
##             Resolved 4.007514e-03 7.201002e-03 5.823419e-03 6.011271e-03
##             Fatal    2.534854e-04 6.970849e-04 3.105196e-03 4.435995e-04
##                      Specimen_Date
## training$Outcome1    2020-06-18   2020-06-19   2020-06-20   2020-06-21
##             Resolved 3.757044e-03 5.823419e-03 5.760802e-03 2.880401e-03
##             Fatal    1.330798e-03 7.604563e-04 4.435995e-04 1.140684e-03
##                      Specimen_Date
## training$Outcome1    2020-06-22   2020-06-23   2020-06-24   2020-06-25
##             Resolved 5.072010e-03 4.884158e-03 5.886036e-03 1.077019e-02
##             Fatal    1.204056e-03 6.970849e-04 3.802281e-04 1.520913e-03
##                      Specimen_Date
## training$Outcome1    2020-06-26   2020-06-27   2020-06-28   2020-06-29
##             Resolved 4.257984e-03 3.068253e-03 2.191609e-03 1.878522e-03
##             Fatal    6.337136e-05 2.534854e-04 2.534854e-04 5.703422e-04
##                      Specimen_Date
## training$Outcome1    2020-06-30   2020-07-01   2020-07-02   2020-07-03
##             Resolved 2.191609e-03 6.887915e-04 1.127113e-03 3.757044e-04
##             Fatal    9.505703e-04 6.337136e-05 5.069708e-04 6.337136e-05
##                      Specimen_Date
## training$Outcome1    2020-07-04   2020-07-05   2020-07-06   2020-07-07
##             Resolved 6.261741e-04 2.504696e-04 3.130870e-04 1.252348e-04
##             Fatal    6.337136e-05 6.337136e-05 2.534854e-04 6.337136e-05
##                      Specimen_Date
## training$Outcome1    2020-07-08   2020-07-09   2020-07-11
##             Resolved 3.757044e-04 6.261741e-05 6.261741e-05
##             Fatal    6.337136e-05 6.337136e-05 6.337136e-05
```

```
## 
##                Age_Group
## training$Outcome1          <20          20s          30s          40s
##           Resolved 4.988318e-02 1.651197e-01 1.497127e-01 1.516070e-01
##           Fatal    5.112801e-04 1.022560e-03 2.875951e-03 9.586502e-03
##                Age_Group
## training$Outcome1          50s          60s          70s          80s
##           Resolved 1.739597e-01 1.171308e-01 6.459557e-02 7.772937e-02
##           Fatal    3.253020e-02 8.570333e-02 1.722375e-01 3.628172e-01
##                Age_Group
## training$Outcome1          90s      UNKNOWN
##           Resolved 5.000947e-02 2.525731e-04
##           Fatal    3.326516e-01 6.391001e-05
## 
##                Client_Gender
## training$Outcome1       FEMALE         MALE        OTHER  TRANSGENDER
##           Resolved 5.404876e-01 4.535750e-01 1.894896e-04 1.894896e-04
##           Fatal    5.358650e-01 4.489835e-01 6.393044e-05 3.835827e-04
##                Client_Gender
## training$Outcome1      UNKNOWN
##           Resolved 5.558363e-03
##           Fatal    1.470400e-02
## 
##                Case_AcquisitionInfo
## training$Outcome1          CC No Epi-link No Info-Missing No Info-Unk
##           Resolved 0.326659509 0.206530664    0.020021474 0.022105729
##           Fatal    0.063223167 0.092948923    0.007479384 0.020584287
##                Case_AcquisitionInfo
## training$Outcome1          OB     Travel
##           Resolved 0.371818354 0.052864271
##           Fatal    0.790896887 0.024867353
## 
##                Outcome1
## training$Outcome1     Resolved        Fatal
##           Resolved 9.999368e-01 6.317518e-05
##           Fatal    6.394271e-05 9.999361e-01
## 
##                Outbreak_Related
## training$Outcome1       Yes          No
##           Resolved 0.3721650 0.6278350
##           Fatal    0.7910992 0.2089008
## 
##                Reporting_PHU_City
## training$Outcome1       Barrie   Belleville    Brantford   Brockville
##           Resolved 1.872517e-02 1.450098e-03 4.413341e-03 9.583255e-03
##           Fatal    1.352817e-02 1.722928e-03 2.361049e-03 1.971795e-02
##                Reporting_PHU_City
## training$Outcome1      Chatham     Cornwall       Guelph     Hamilton
##           Resolved 4.791627e-03 4.854675e-03 1.506841e-02 2.553433e-02
##           Fatal    6.381214e-04 2.871546e-03 1.269862e-02 1.525110e-02
```

```
##                  Reporting_PHU_City
## training$Outcome1        Kenora      Kingston        London  Mississauga
##           Resolved 9.457159e-04 3.215434e-03 1.595107e-02 1.706071e-01
##           Fatal    6.381214e-05 6.381214e-05 2.361049e-02 1.132027e-01
##                  Reporting_PHU_City
## training$Outcome1 New Liskeard    Newmarket    North Bay     Oakville
##           Resolved 5.674295e-04 8.461005e-02 1.134859e-03 2.074270e-02
##           Fatal    6.381214e-05 9.322953e-02 5.104971e-04 8.295578e-03
##                  Reporting_PHU_City
## training$Outcome1       Ottawa   Owen Sound     Pembroke Peterborough
##           Resolved 5.459933e-02 4.161150e-03 7.565727e-04 2.837148e-03
##           Fatal    9.648395e-02 6.381214e-05 3.190607e-04 8.933699e-04
##                  Reporting_PHU_City
## training$Outcome1 Point Edward    Port Hope Sault Ste. Marie       Simcoe
##           Resolved 8.133157e-03 6.809155e-03    1.197907e-03 1.166383e-02
##           Fatal    7.721269e-03 7.019335e-03    6.381214e-05 1.340055e-02
##                  Reporting_PHU_City
## training$Outcome1   St. Thomas    Stratford      Sudbury      Thorold
##           Resolved 2.711052e-03 1.134859e-03 2.143623e-03 2.067965e-02
##           Fatal    1.722928e-03 1.850552e-03 1.020994e-03 2.444005e-02
##                  Reporting_PHU_City
## training$Outcome1 Thunder Bay      Timmins      Toronto     Waterloo
##           Resolved 3.404577e-03 1.891432e-03 3.672530e-01 3.946788e-02
##           Fatal    4.466850e-04 3.637292e-03 4.026546e-01 4.300938e-02
##                  Reporting_PHU_City
## training$Outcome1       Whitby      Windsor
##           Resolved 4.495303e-02 4.400731e-02
##           Fatal    6.196159e-02 2.546104e-02
##
##                  Reporting_PHU_Latitude
## training$Outcome1     [,1]       [,2]
##           Resolved 43.74427 0.7380921
##           Fatal    43.82880 0.7294835
##
##                  Reporting_PHU_Longitude
## training$Outcome1     [,1]       [,2]
##           Resolved -79.53692 1.598084
##           Fatal    -79.20676 1.574879

#Confusion matrix
NBC.predict <- predict(NBC.classifier, test)
CrossTable(NBC.predict, test$Outcome1, prop.chisq = FALSE, prop.t = FALSE,
prop.r = FALSE,
dnn = c('predicted', 'actual'))

##
##
##    Cell Contents
## |-------------------------|
## |                       N |
```

```
## |                 N / Col Total |
## |-------------------------------|
##
##
## Total Observations in Table:  3495
##
##
##               | actual
##     predicted |  Resolved |      Fatal | Row Total |
## --------------|-----------|-----------|-----------|
##      Resolved |      1758 |         1 |      1759 |
##               |     1.000 |     0.001 |           |
## --------------|-----------|-----------|-----------|
##         Fatal |         0 |      1736 |      1736 |
##               |     0.000 |     0.999 |           |
## --------------|-----------|-----------|-----------|
## Column Total |       1758 |      1737 |      3495 |
##               |     0.503 |     0.497 |           |
## --------------|-----------|-----------|-----------|
##
##

#recall = 0.9994
#precision = 1.000
```

Step 4: Modeling - (3) Random Forest

```
#install.packages("randomForest")
library(randomForest)

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

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

#Building Random Forest model on training set
model.randomForest <- randomForest(Outcome1 ~ Client_Gender + Age_Group +
Case_AcquisitionInfo + Outbreak_Related + Reporting_PHU_City, data=training,
importance = TRUE)
model.randomForest

##
## Call:
```

```
##  randomForest(formula = Outcome1 ~ Client_Gender + Age_Group +
Case_AcquisitionInfo + Outbreak_Related + Reporting_PHU_City,      data =
training, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 14.72%
## Confusion matrix:
##            Resolved Fatal class.error
## Resolved     12335  3492  0.22063562
## Fatal         1140 14497  0.07290401
```

```
#Predicting on test set
predict.RF <- predict(model.randomForest, test, type = "class")
head(predict.RF)
```

```
##       32        39        41        81        96       114
## Resolved     Fatal     Fatal    Fatal Resolved Resolved
## Levels: Resolved Fatal
```

```
#Confusion Matrix
table(predict.RF, test$Outcome1)
```

```
##
## predict.RF Resolved Fatal
##    Resolved     1393   119
##    Fatal         365  1618
```

```
#recall = 0.9315
#precision = 0.8159
```

```
#importance that model has assigned to each variable
varImpPlot(model.randomForest)
```

model.randomForest

Age_Group	Age_Group
Reporting_PHU_City	Case_AcquisitionInfo
Client_Gender	Outbreak_Related
Case_AcquisitionInfo	Reporting_PHU_City
Outbreak_Related	Client_Gender

50   0   6000

MeanDecreaseAc   MeanDecrease