Multinomial_Regression

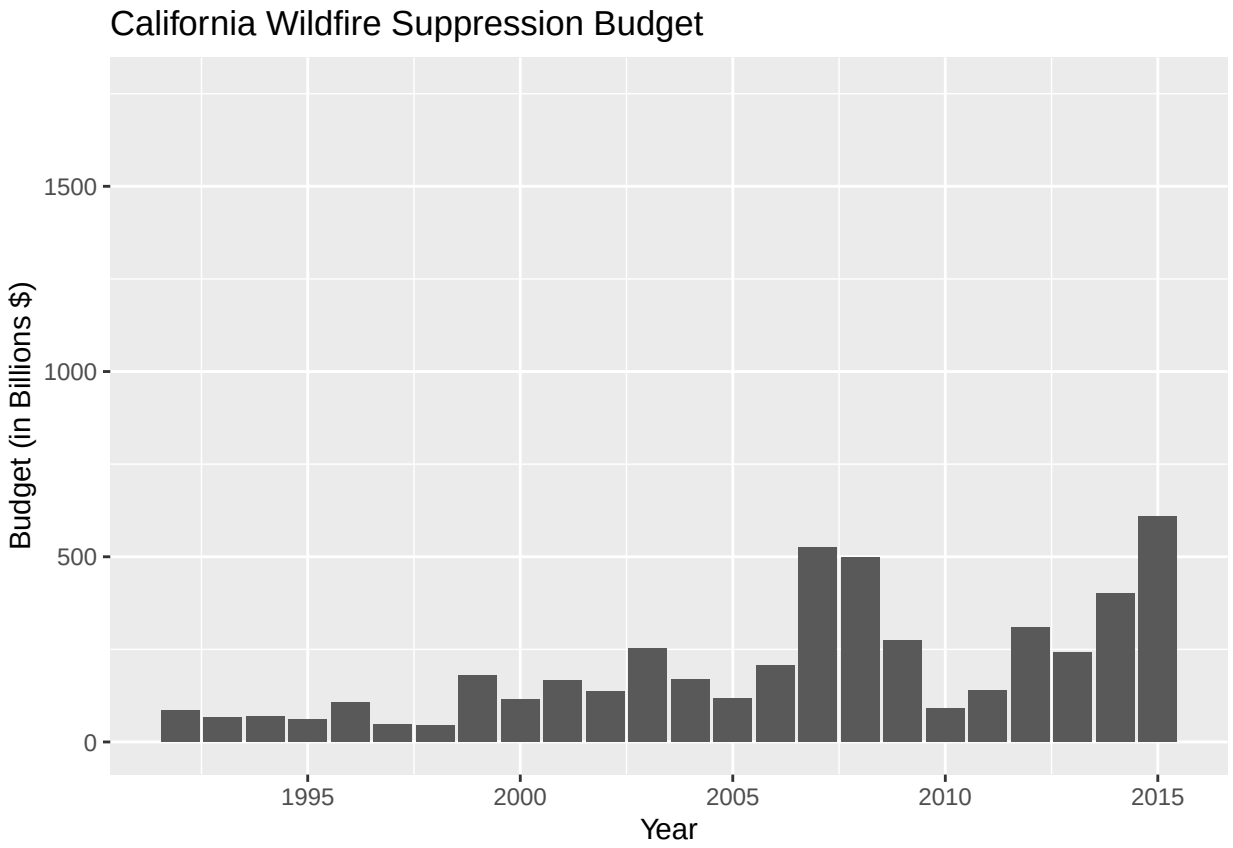
Sumanth Nallamotu

4/30/2022

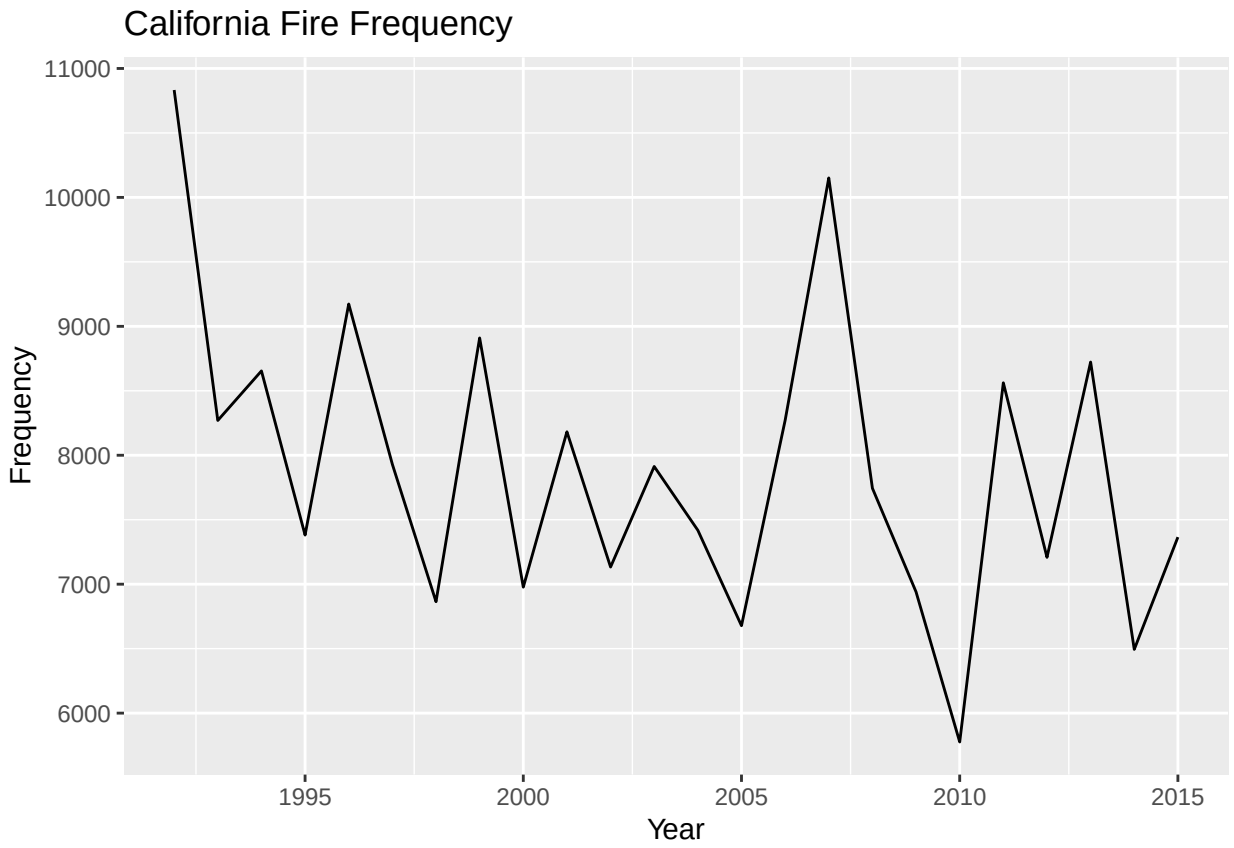
R Markdown

```
fires <- read.csv("/Users/sumanth/Documents/GitHub/6101_Group5_Wildfires/final_wildfire.csv")

fires$STAT_CAUSE_CODE <- as.factor(fires$STAT_CAUSE_CODE)
fires$FIRE_SIZE_CLASS <- as.factor(fires$FIRE_SIZE_CLASS)

split <- createDataPartition(fires$FIRE_SIZE_CLASS, p = .70, list = FALSE)
train <- fires[split,]
test <- fires[-split,]
```

```
budge <- read.csv("/Users/sumanth/Documents/GitHub/6101_Group5_Wildfires/fire_suppression.csv")
budge$Budget <- as.numeric(gsub("[$,]", "", budge$Budget))
budge$Budget <- budge$Budget / 1000000

firesByYear <- fires %>% count(Year)
firesBudget <- merge(firesByYear, budge, by="Year")

ggplot(firesBudget, aes(x=Year, y=Budget)) + geom_bar(stat='identity') + ylim(0, max(budge$Budget)) + xl
  ylab("Budget (in Billions $)") + ggtitle("California Wildfire Suppression Budget")
```




```
ggplot(firesBudget) + geom_line(aes(x=Year, y=n), stat='identity', group=1) + xlab("Year") + ylab("Freq
  ggtitle("California Fire Frequency")
```




```
model1 <- multinom(FIRE_SIZE_CLASS ~ tair_day_livneh_vic +
                     soilmoist1_day_livneh_vic +
                     rainfall_day_livneh_vic, data = fires)
```

```
## # weights:  35 (24 variable)
## initial  value 341878.899998
## iter  10 value 183964.011779
## iter  20 value 180022.515724
## iter  30 value 172387.834134
## iter  40 value 172152.087864
## final  value 172151.836882
## converged
```

```
summary(model1)
```

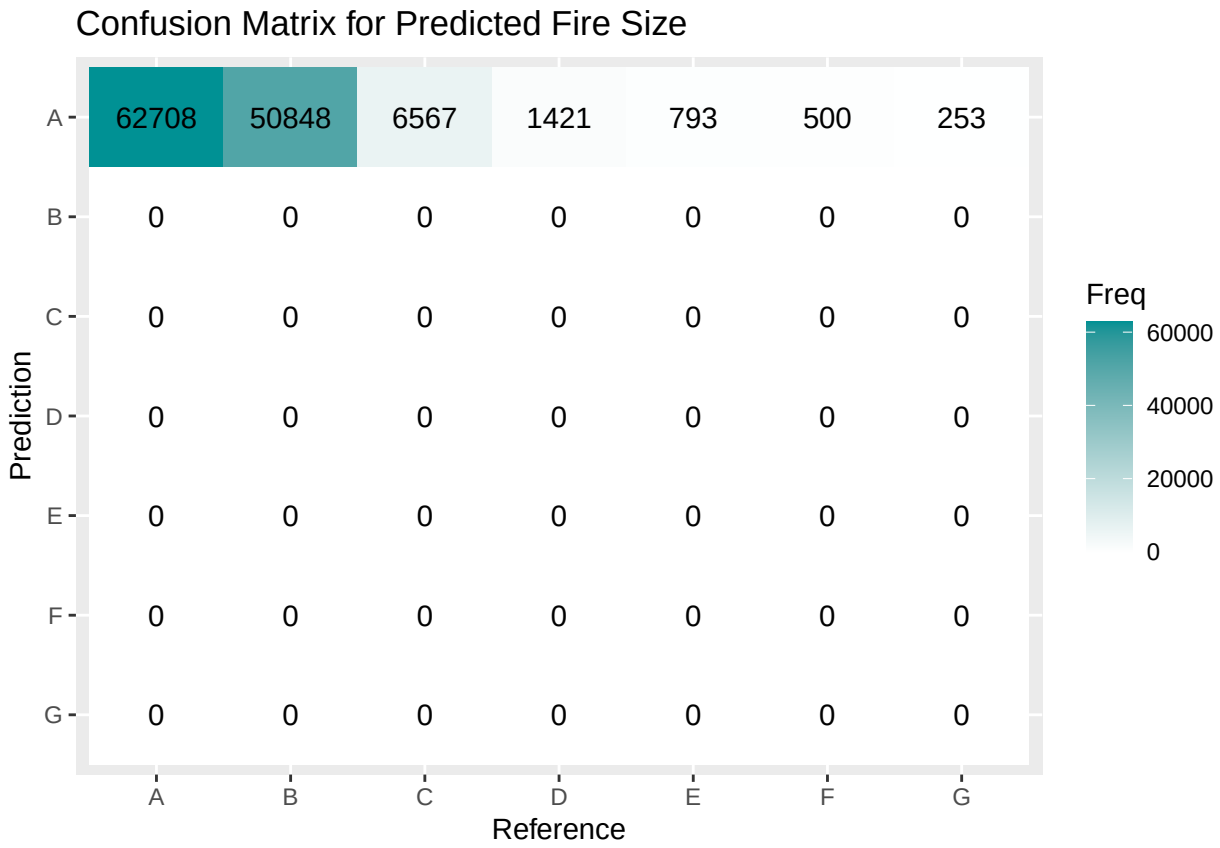
```
## Call:
## multinom(formula = FIRE_SIZE_CLASS ~ tair_day_livneh_vic + soilmoist1_day_livneh_vic +
##     rainfall_day_livneh_vic, data = fires)
##
## Coefficients:
##   (Intercept) tair_day_livneh_vic soilmoist1_day_livneh_vic
## B   0.3004042        -0.006758249              -0.026042456
## C  -1.9555961         0.019870821              -0.049524776
## D  -3.7888146         0.030926364              -0.043931875
## E  -4.7896124         0.057208376              -0.053905459
## F  -6.7093016         0.095029479               0.003408496
## G  -5.8488367         0.090499419              -0.113385372
##   rainfall_day_livneh_vic
```

```
## B             -0.10832070
## C             -0.12486710
## D             -0.10371291
## E             -0.05670748
## F             -0.17859896
## G             -0.42030231
##
## Std. Errors:
##   (Intercept) tair_day_livneh_vic soilmoist1_day_livneh_vic
## B  0.06783931         0.001427835               0.003554221
## C  0.15610215         0.003264026               0.008340375
## D  0.32982194         0.006908733               0.017640060
## E  0.45918391         0.009632546               0.024811319
## F  0.59938981         0.012785392               0.031885844
## G  0.89960621         0.018810195               0.050822294
##   rainfall_day_livneh_vic
## B             0.006549645
## C             0.017697053
## D             0.036637381
## E             0.048671895
## F             0.074055428
## G             0.151514548
##
## Residual Deviance: 344303.7
## AIC: 344351.7
```

```
train$predictions <- predict(model1, newdata = train, "class")
tab <- table(train$FIRE_SIZE_CLASS, train$predictions)
round((sum(diag(tab))/sum(tab))*100,2)
```

```
## [1] 50.94
```

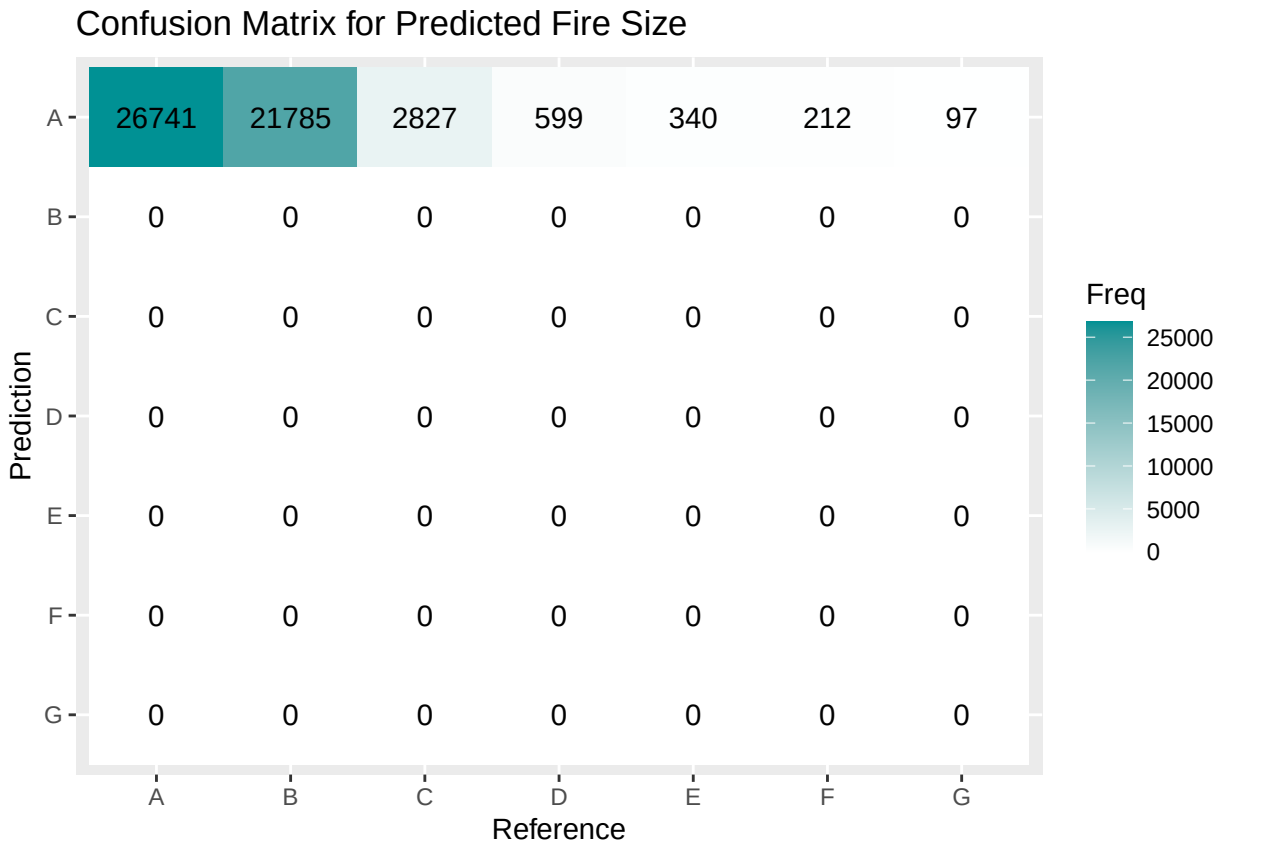
```
cm <- confusionMatrix(train$predictions, train$FIRE_SIZE_CLASS)

cmdf <- as.data.frame(cm$table)
cmdf$Prediction <- factor(cmdf$Prediction, levels=rev(levels(cmdf$Prediction)))

ggplot(cmdf, aes(Reference,Prediction, fill= Freq)) +
        geom_tile() + geom_text(aes(label=Freq)) +
        scale_fill_gradient(low="white", high="#009194") +
        labs(x = "Reference",y = "Prediction", title="Confusion Matrix for Predicted Fire Size")
```




```
test$predictions <- predict(model1, newdata = test, "class")
tab <- table(test$FIRE_SIZE_CLASS, test$predictions)
round((sum(diag(tab))/sum(tab))*100,2)
```

```
## [1] 50.84
```

```
cm <- confusionMatrix(test$predictions, test$FIRE_SIZE_CLASS)

cmdf <- as.data.frame(cm$table)
cmdf$Prediction <- factor(cmdf$Prediction, levels=rev(levels(cmdf$Prediction)))

ggplot(cmdf, aes(Reference,Prediction, fill= Freq)) +
        geom_tile() + geom_text(aes(label=Freq)) +
        scale_fill_gradient(low="white", high="#009194") +
        labs(x = "Reference",y = "Prediction", title="Confusion Matrix for Predicted Fire Size")
```




```
model2 <- multinom(STAT_CAUSE_CODE ~ month + tair_day_livneh_vic +
                     soilmoist1_day_livneh_vic +
                     rainfall_day_livneh_vic, data = fires)
```

```
## # weights:  78 (60 variable)
## initial  value 450638.517563
## iter  10 value 351990.198964
## iter  20 value 350236.347991
## iter  30 value 349300.527741
## iter  40 value 348573.289051
## iter  50 value 347430.734728
## iter  60 value 339582.375984
## iter  70 value 338693.455512
## iter  80 value 338687.765856
## final  value 338687.730830
## converged
```

```
summary(model1)
```

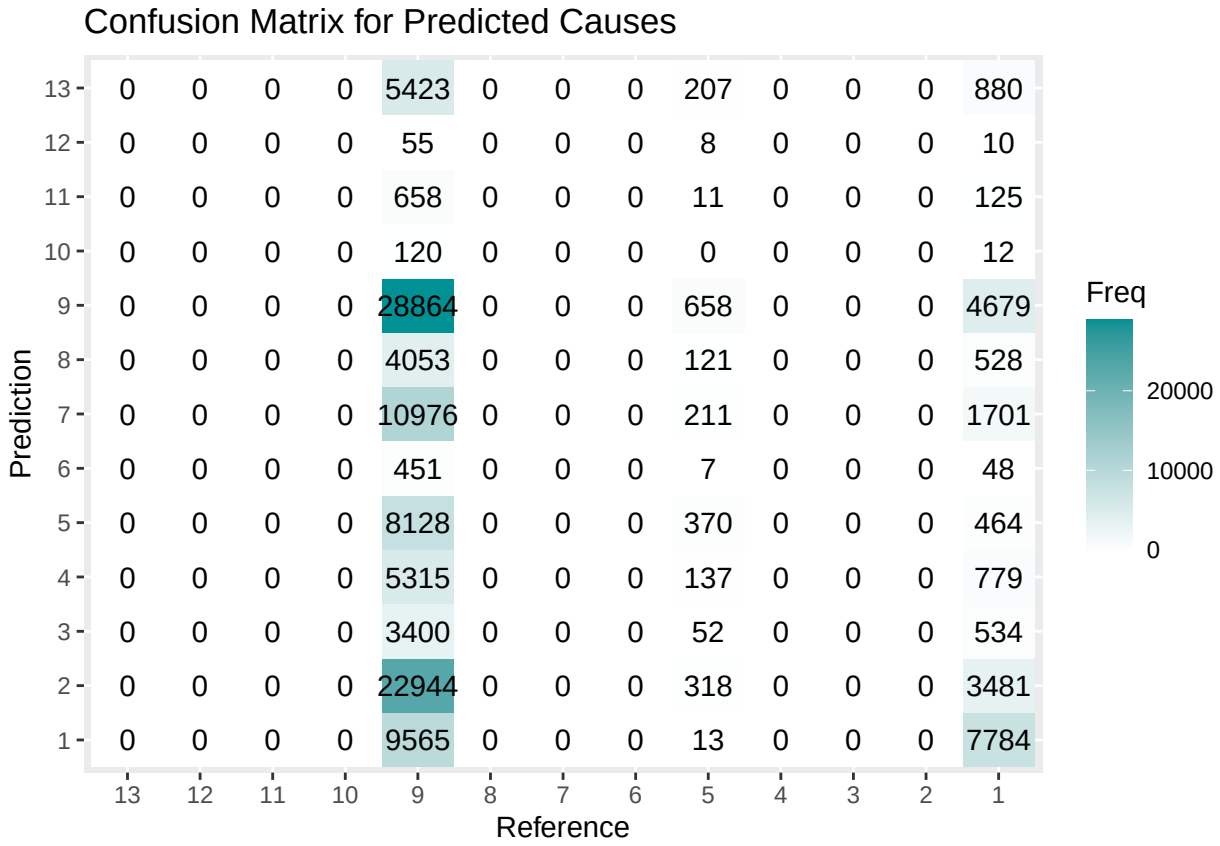
```
## Call:
## multinom(formula = FIRE_SIZE_CLASS ~ tair_day_livneh_vic + soilmoist1_day_livneh_vic +
##     rainfall_day_livneh_vic, data = fires)
##
## Coefficients:
##   (Intercept) tair_day_livneh_vic soilmoist1_day_livneh_vic
## B   0.3004042        -0.006758249              -0.026042456
## C  -1.9555961         0.019870821              -0.049524776
## D  -3.7888146         0.030926364              -0.043931875
```

```
## E  -4.7896124         0.057208376              -0.053905459
## F  -6.7093016         0.095029479               0.003408496
## G  -5.8488367         0.090499419              -0.113385372
##   rainfall_day_livneh_vic
## B             -0.10832070
## C             -0.12486710
## D             -0.10371291
## E             -0.05670748
## F             -0.17859896
## G             -0.42030231
##
## Std. Errors:
##   (Intercept) tair_day_livneh_vic soilmoist1_day_livneh_vic
## B  0.06783931         0.001427835               0.003554221
## C  0.15610215         0.003264026               0.008340375
## D  0.32982194         0.006908733               0.017640060
## E  0.45918391         0.009632546               0.024811319
## F  0.59938981         0.012785392               0.031885844
## G  0.89960621         0.018810195               0.050822294
##   rainfall_day_livneh_vic
## B             0.006549645
## C             0.017697053
## D             0.036637381
## E             0.048671895
## F             0.074055428
## G             0.151514548
##
## Residual Deviance: 344303.7
## AIC: 344351.7
```

```
# Predicting the values for train dataset
train$predictions <- as.factor(predict(model2, newdata = train, "class"))

# Building classification table
tab <- table(train$STAT_CAUSE_CODE, train$predictions)

# Calculating accuracy - sum of diagonal elements divided by total obs
round((sum(diag(tab))/sum(tab))*100,2)
```

```
## [1] 30.07
```

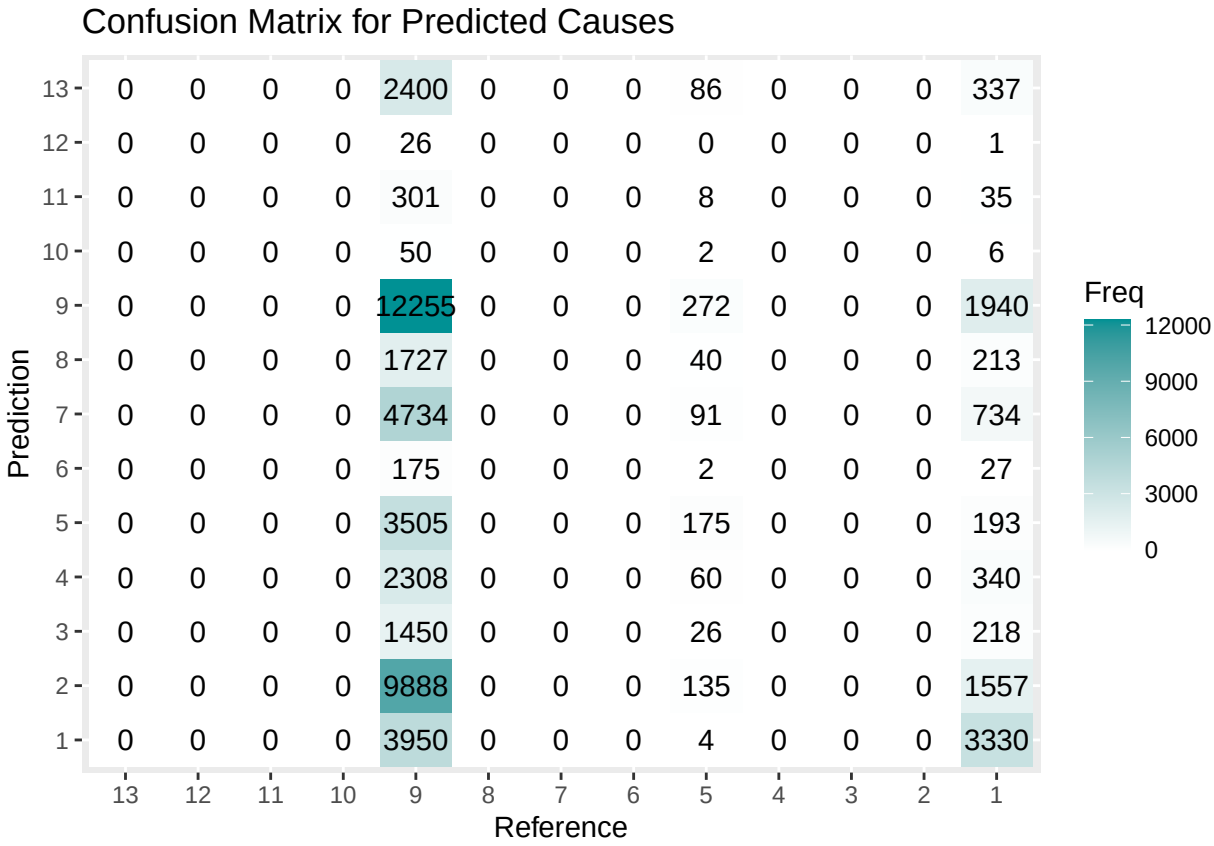
```
cm2 <- confusionMatrix(train$predictions, train$STAT_CAUSE_CODE)

cm2df <- as.data.frame(cm2$table)
cm2df$Prediction <- factor(cm2df$Prediction, levels=rev(levels(cm2df$Prediction)))

ggplot(cm2df, aes(Prediction,Reference, fill= Freq)) +
        geom_tile() + geom_text(aes(label=Freq)) +
        scale_fill_gradient(low="white", high="#009194") +
        labs(x = "Reference",y = "Prediction", title="Confusion Matrix for Predicted Causes")
```




```
# Predicting the values for train dataset
test$predictions <- as.factor(predict(model2, newdata = test, "class"))

# Building classification table
tab <- table(test$STAT_CAUSE_CODE, test$predictions)

# Calculating accuracy - sum of diagonal elements divided by total obs
round((sum(diag(tab))/sum(tab))*100,2)
```

```
## [1] 29.96
```

```
cm2 <- confusionMatrix(test$predictions, test$STAT_CAUSE_CODE)

cm2df <- as.data.frame(cm2$table)
cm2df$Prediction <- factor(cm2df$Prediction, levels=rev(levels(cm2df$Prediction)))

ggplot(cm2df, aes(Prediction,Reference, fill= Freq)) +
        geom_tile() + geom_text(aes(label=Freq)) +
        scale_fill_gradient(low="white", high="#009194") +
        labs(x = "Reference",y = "Prediction", title="Confusion Matrix for Predicted Causes")
```


Confusion Matrix for Predicted Causes
Prediction
Reference
Freq
12000
9000
6000
3000
0
13 12 11 10 9 8 7 6 5 4 3 2 1
13: 0 0 0 0 2400 0 0 0 86 0 0 0 337
12: 0 0 0 0 26 0 0 0 0 0 0 0 1
11: 0 0 0 0 301 0 0 0 8 0 0 0 35
10: 0 0 0 0 50 0 0 0 2 0 0 0 6
9: 0 0 0 0 12255 0 0 0 272 0 0 0 1940
8: 0 0 0 0 1727 0 0 0 40 0 0 0 213
7: 0 0 0 0 4734 0 0 0 91 0 0 0 734
6: 0 0 0 0 175 0 0 0 2 0 0 0 27
5: 0 0 0 0 3505 0 0 0 175 0 0 0 193
4: 0 0 0 0 2308 0 0 0 60 0 0 0 340
3: 0 0 0 0 1450 0 0 0 26 0 0 0 218
2: 0 0 0 0 9888 0 0 0 135 0 0 0 1557
1: 0 0 0 0 3950 0 0 0 4 0 0 0 3330