Baseline Model Performance

Gwen Rino

April 11, 2018

Model Description

In order to establish a baseline measure of Mean Absolute Error (MAE) for dengue prediction models, I created a naive model that predicts the number of cases of dengue per week to be equal to the average number of cases for that season of the year.

```
# Find mean total_cases for fall, spring, summer, winter
dengue %>% dplyr::group_by(season) %>%
  summarize(mean = mean(total_cases, na.rm = TRUE))
```

```
## # A tibble: 4 x 2
##   season  mean
##   <chr>  <dbl>
## 1 fall    59.2
## 2 spring  11.9
## 3 summer  41.6
## 4 winter  23.9
```

```
# Assign seasonal mean as prediction for each observation
dengue.naive <- dengue %>%
  mutate(predictions = case_when(season == "fall" ~ 59.2,
                                 season == "spring" ~ 11.9,
                                 season == "summer" ~ 41.6,
                                 season == "winter" ~ 23.9))

predictions <- dengue.naive$predictions
```

Model Evaluation

The baseline MAE is 25.5.

```
# Error = actual number of cases - predicted number of cases
error <- dengue.naive$total_cases - predictions

# Find MAE
print(paste("Naive model MAE = ", mean(abs(error))))
```

```
## [1] "Naive model MAE =  25.4959401709402"
```