ECO 395M Homework 3

Haokun Zhang (hz7297), Jyun-Yu Cheng (jc95272) and Lu Zhang (lz7445)

2023-03-22

What causes what?

1. The number of cops in the city doesn't direct cause the reduction of crime rate on streets. There are many other factors and events that have correlation to crime rates, such as terror alert level, and certain events and festivals that increase the number of cops deplied. Also, high crime rate in one area will naturally increase the cop forces deployed, thus will reduce the criminal rate. Focus only on the relationship between numbers of cops on the street and crime rate will lead to a endogeneity biased conclusion on causation of numbers of cops to the crime.

2. The researchers used daily police reports of crime from the Metropolitan Police Department of the District of Columbia that cover the time period of 506 days since the HSAS terror alert system began. During high alert level, the D.C. police forces increased their presence on the streets.The researchers used the high-alert periods to estimate the effect of police on crime and break the circle of endogenous relationship between police presence and crime rate. In Table 2, the daily total number of crimes in D.C. decreased by an average of seven crimes per day on high-alert days.

3. The researchers included metro ridership as a variable to test their hypothesize that tourism is reduced on high-alert days, therefore there are less crimes on street. They added logged midday Metro ridership to the regression and captured the percentage of change on number of crimes based on the change of Metro ridership

4. Table 4 presents reduction in crime on high-alert days using police patrol concentration on hte national mall. The first column presents robust coefficient of estimation of crime in the National Mall area and the other districts during periods of high alert. 2.62 crimes decreased in the National Mall area, implying 15 percent of decline during high-alert days. Crime also decreases in the other districts, though the effect is not statistically significant. Lastly, ten percentage of increase on midday ridership increases 0.24 percent of crime rate.

Question 2 : Tree modeling: dengue cases

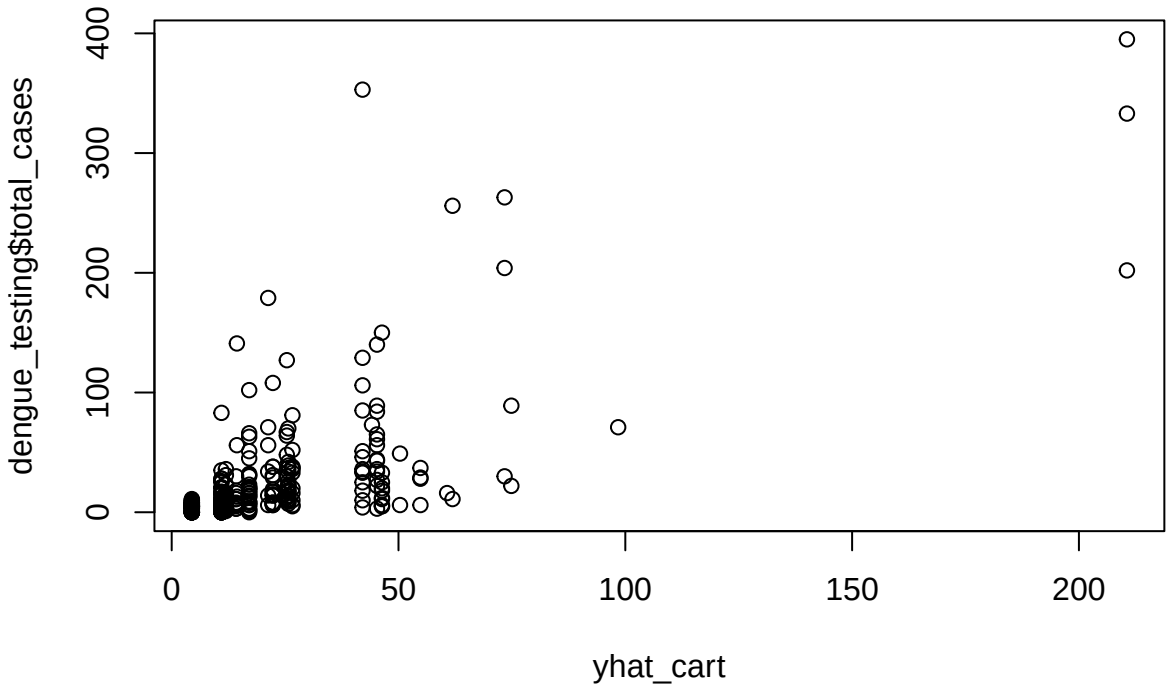
```
##   city season total_cases   ndvi_ne   ndvi_nw   ndvi_se   ndvi_sw
## 1   sj spring           4 0.1226000 0.1037250 0.1984833 0.1776167
## 2   sj spring           5 0.1699000 0.1421750 0.1623571 0.1554857
## 3   sj spring           4 0.0322500 0.1729667 0.1572000 0.1708429
## 4   sj spring           3 0.1286333 0.2450667 0.2275571 0.2358857
## 5   sj spring           6 0.1962000 0.2622000 0.2512000 0.2473400
## 6   sj summer           2        NA 0.1748500 0.2543143 0.1817429
##   precipitation_amt air_temp_k avg_temp_k dew_point_temp_k max_air_temp_k
## 1             12.42   297.5729   297.7429         292.4143          299.8
## 2             22.82   298.2114   298.4429         293.9514          300.9
## 3             34.54   298.7814   298.8786         295.4343          300.5
## 4             15.36   298.9871   299.2286         295.3100          301.4
```

```
## 5              7.52   299.5186   299.6643         295.8214          301.9
## 6              9.58   299.6300   299.7643         295.8514          302.4
##   min_air_temp_k precip_amt_kg_per_m2 relative_humidity_percent
## 1          295.9                32.00                  73.36571
## 2          296.4                17.94                  77.36857
## 3          297.3                26.10                  82.05286
## 4          297.0                13.90                  80.33714
## 5          297.5                12.20                  80.46000
## 6          298.1                26.49                  79.89143
##   specific_humidity   tdtr_k
## 1          14.01286 2.628571
## 2          15.37286 2.371429
## 3          16.84857 2.300000
## 4          16.67286 2.428571
## 5          17.21000 3.014286
## 6          17.21286 2.100000
```

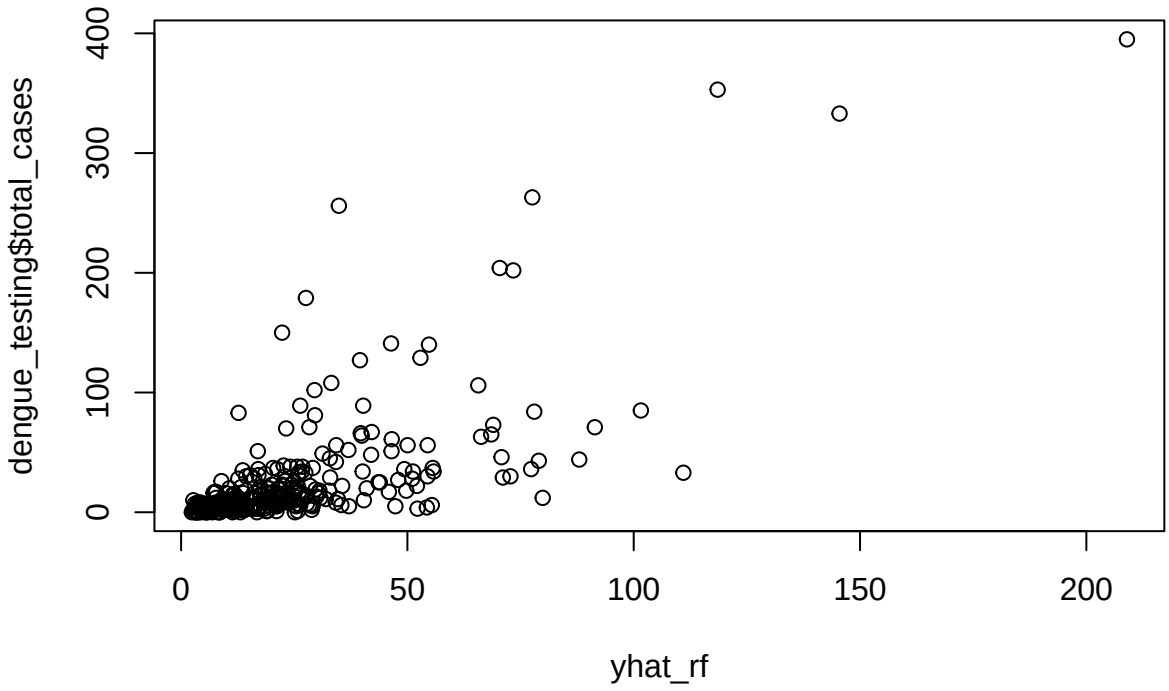
```
##   city season total_cases   ndvi_ne   ndvi_nw   ndvi_se   ndvi_sw
## 1   sj spring           4 0.1226000 0.1037250 0.1984833 0.1776167
## 2   sj spring           5 0.1699000 0.1421750 0.1623571 0.1554857
## 3   sj spring           4 0.0322500 0.1729667 0.1572000 0.1708429
## 4   sj spring           3 0.1286333 0.2450667 0.2275571 0.2358857
## 5   sj spring           6 0.1962000 0.2622000 0.2512000 0.2473400
## 6   sj summer           2 0.1468334 0.1748500 0.2543143 0.1817429
##   precipitation_amt air_temp_k avg_temp_k dew_point_temp_k max_air_temp_k
## 1             12.42   297.5729   297.7429         292.4143          299.8
## 2             22.82   298.2114   298.4429         293.9514          300.9
## 3             34.54   298.7814   298.8786         295.4343          300.5
## 4             15.36   298.9871   299.2286         295.3100          301.4
## 5              7.52   299.5186   299.6643         295.8214          301.9
## 6              9.58   299.6300   299.7643         295.8514          302.4
##   min_air_temp_k precip_amt_kg_per_m2 relative_humidity_percent
## 1          295.9                32.00                  73.36571
## 2          296.4                17.94                  77.36857
## 3          297.3                26.10                  82.05286
## 4          297.0                13.90                  80.33714
## 5          297.5                12.20                  80.46000
## 6          298.1                26.49                  79.89143
##   specific_humidity   tdtr_k
## 1          14.01286 2.628571
## 2          15.37286 2.371429
## 3          16.84857 2.300000
## 4          16.67286 2.428571
## 5          17.21000 3.014286
## 6          17.21286 2.100000
```

```
## Distribution not specified, assuming gaussian ...
```

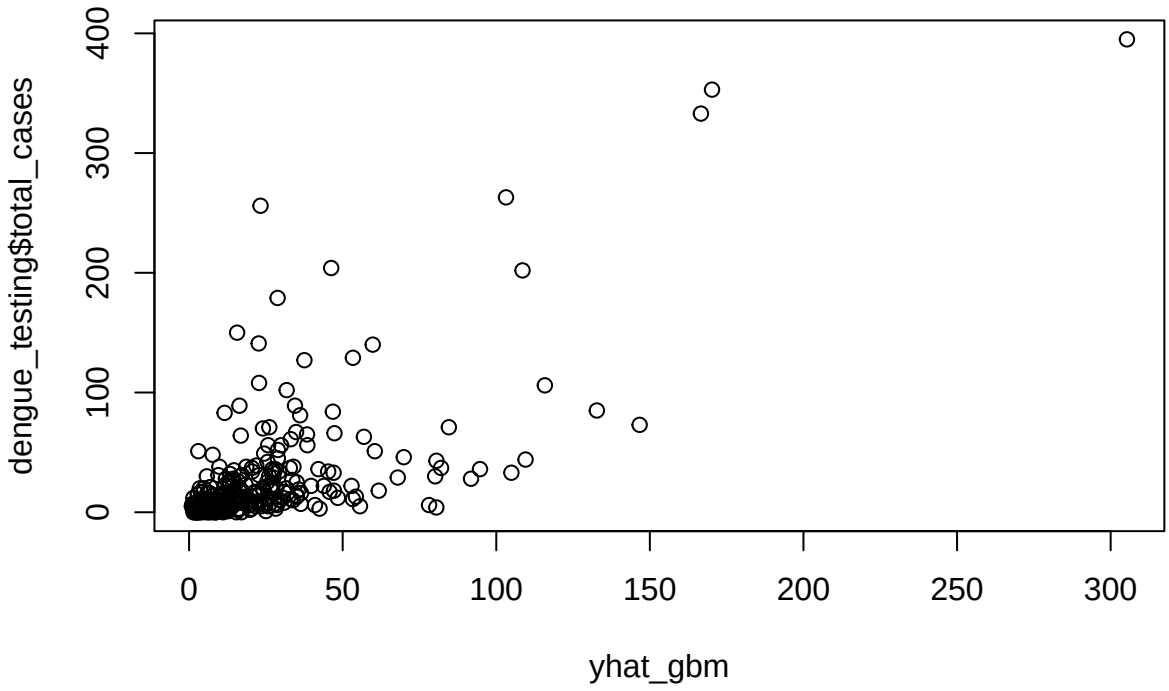
Predict the dengue cases with CART model, Random Forest model and Gradient-boosted model


dengue_testing$total_cases
0
100
200
300
400
0
50
100
150
200
yhat_cart

```
## [1] 37.76668
```


dengue_testing$total_cases
0
100
200
300
400
0
50
100
150
200
yhat_rf

```
## [1] 37.48847
```




```
## [1] 36.8411
```

let's compare RMSE on the test set

```
## [1] 37.76668
```

```
## [1] 37.48847
```

```
## [1] 36.8411
```

A K-fold test is conducted to further evaluate the performance

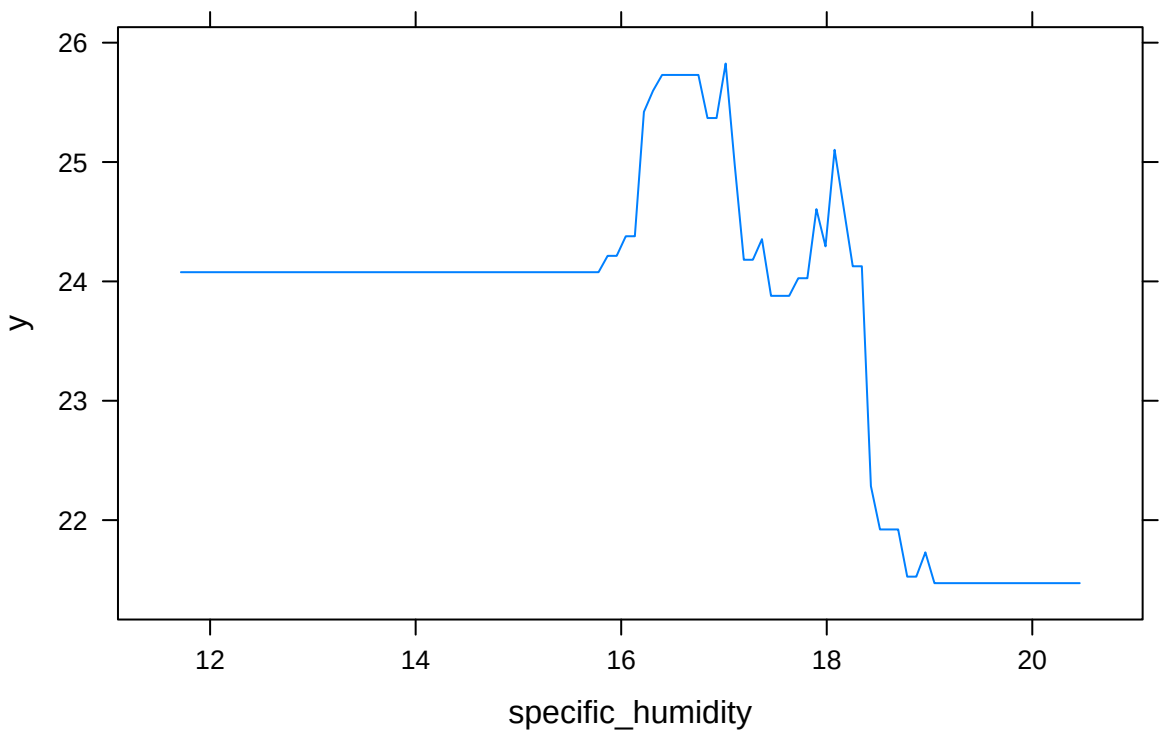
```
## Distribution not specified, assuming gaussian ...
```

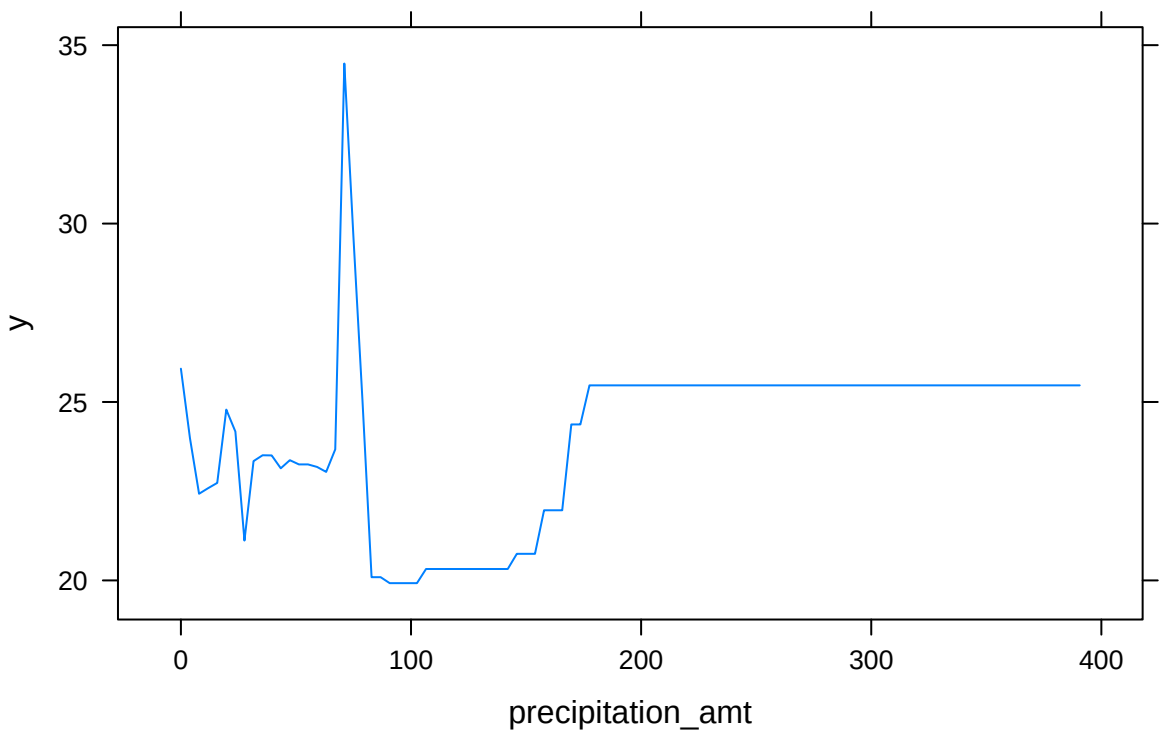
```
## CART accuracy: 1421.238
```

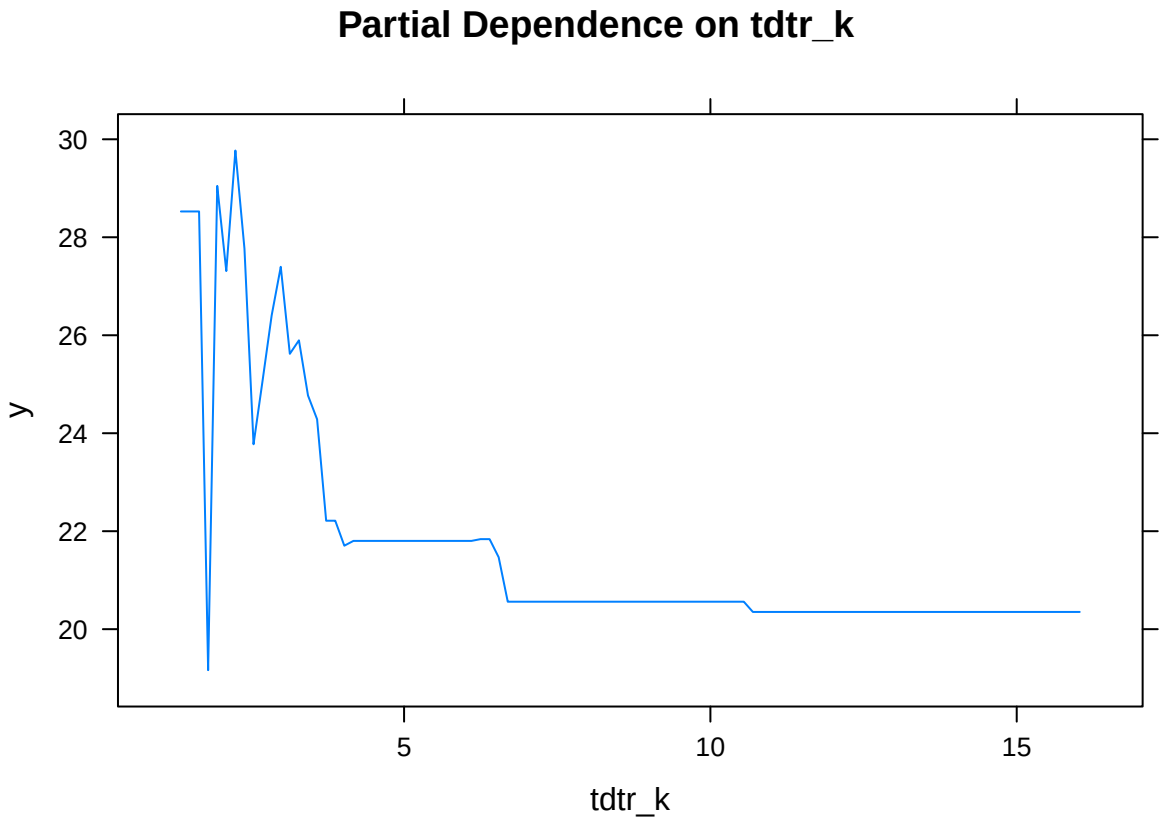
```
## Random Forest accuracy: 1414.238
```

```
## Gradient Boosting accuracy: 1279.522
```

Because the gradient-boosted model has the smallest out-of-sample RMSE and MSE, we decided to choose it as the model to make partial dependence plots.


Partial Dependence on specific humidity
26
25
24
23
22
y
12
14
16
18
20
specific_humidity


Partial Dependence on precipitation_amt
35
30
25
20
y
0
100
200
300
400
precipitation_amt



We choose "tdtr_k" to make a partial dependence plots because we think that if the DTR is bigger, it's more difficult for mosquito to live as a result, we want to know whether DTR affect the total dengue fever cases and we can see that it has big influence on the infection cases.