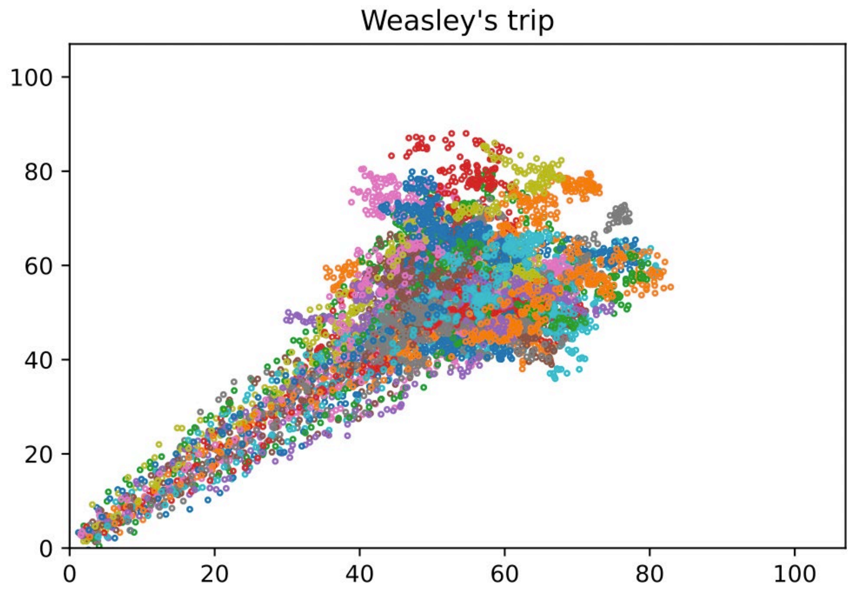
VE414 Project

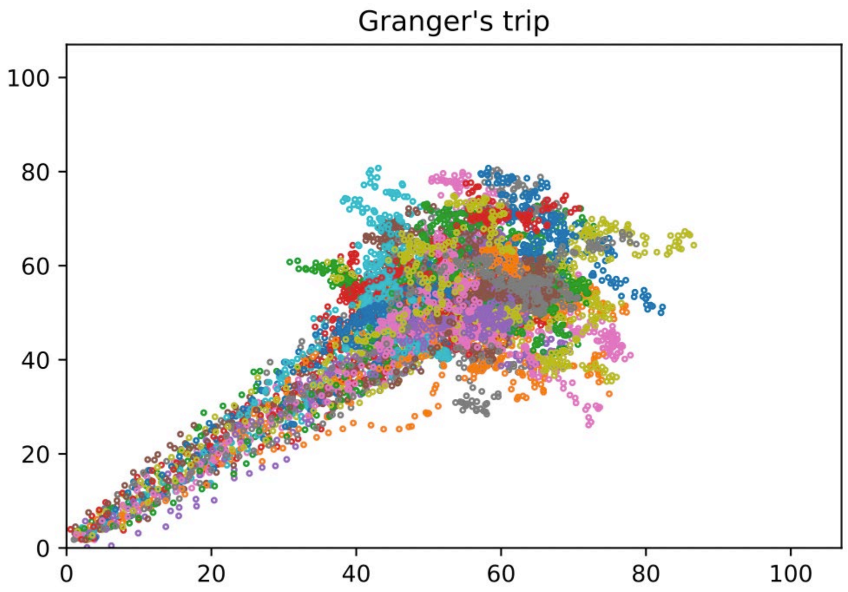
Shengyuan Xu 518370910200

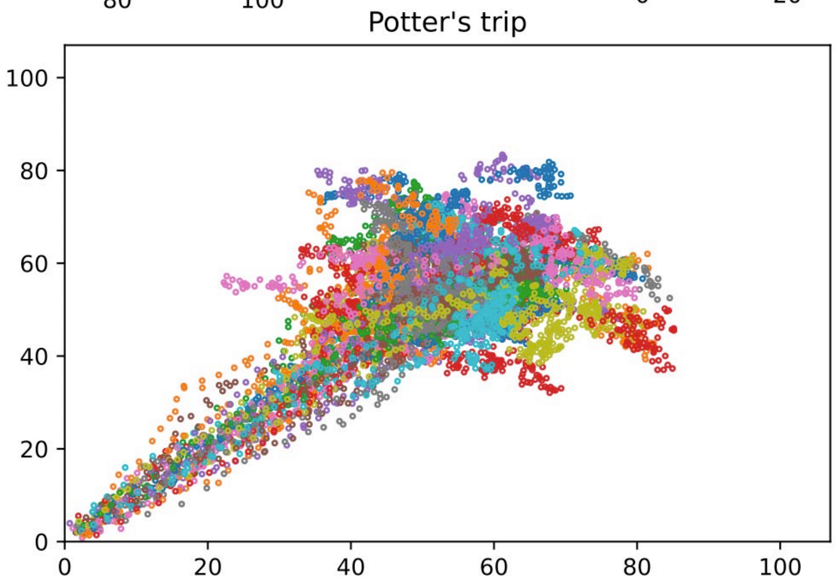
Fan Chen 518021910739

Qiansiqi Hu 519370910097

Magic Trip

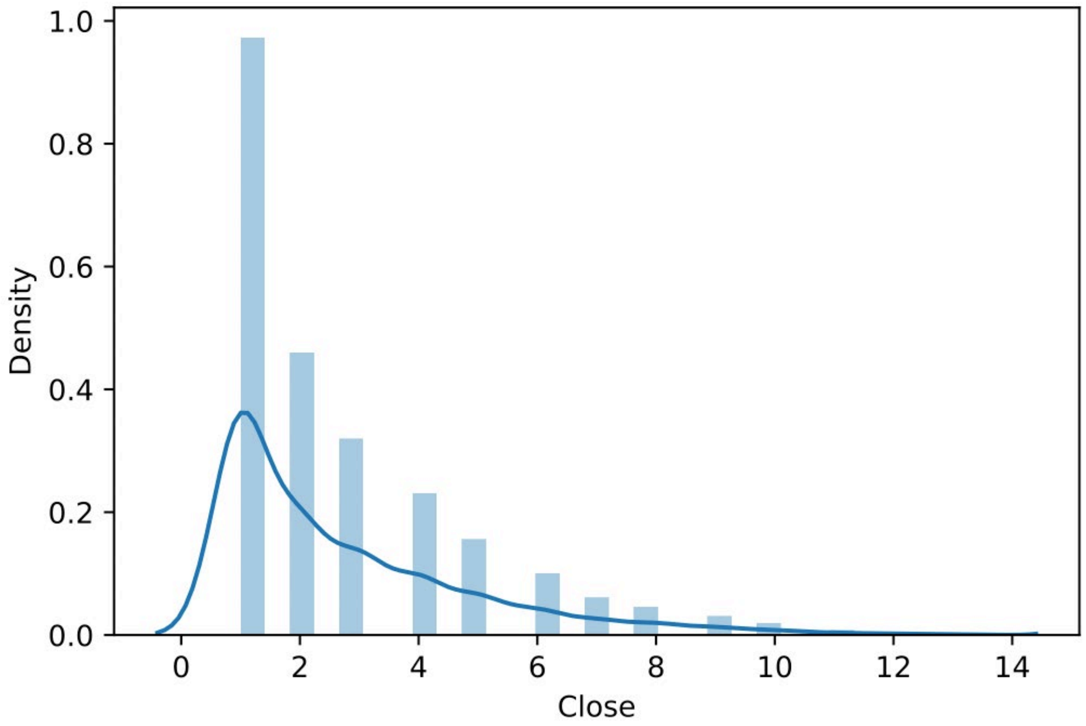

Weasley's trip
0
20
40
60
80
100
0
20
40
60
80
100

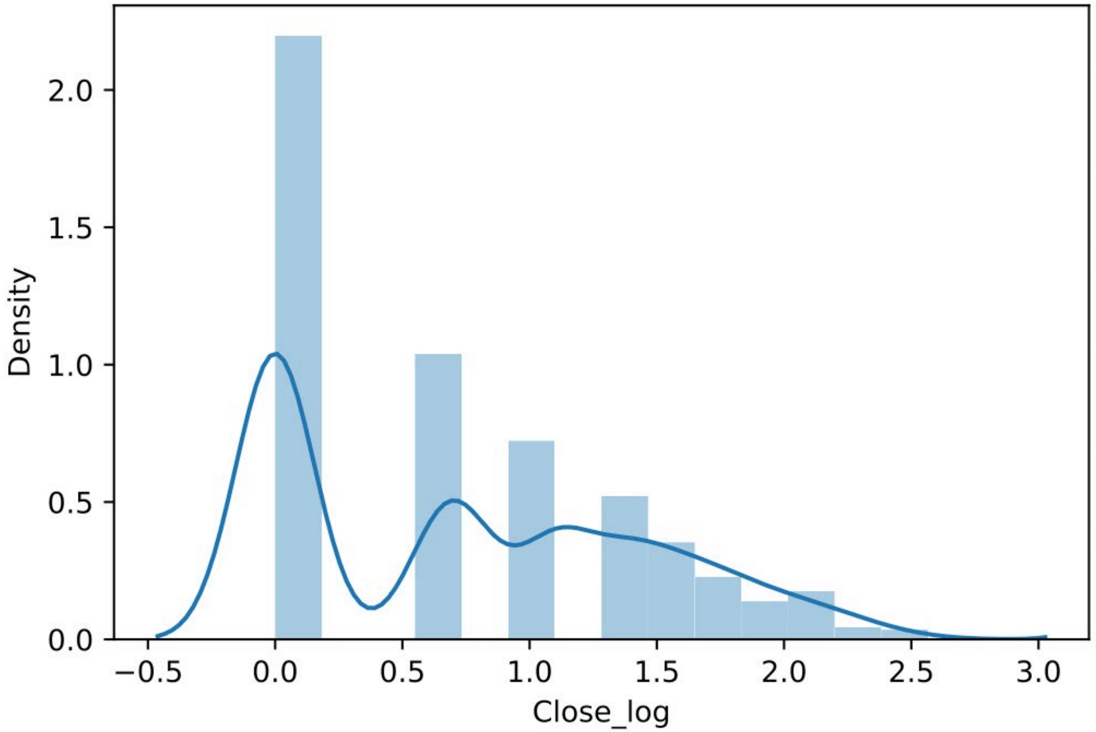

Granger's trip
0
20
40
60
80
100
0
20
40
60
80
100


Potter's trip
0
20
40
60
80
100
0
20
40
60
80
100

Distribution of Tayes

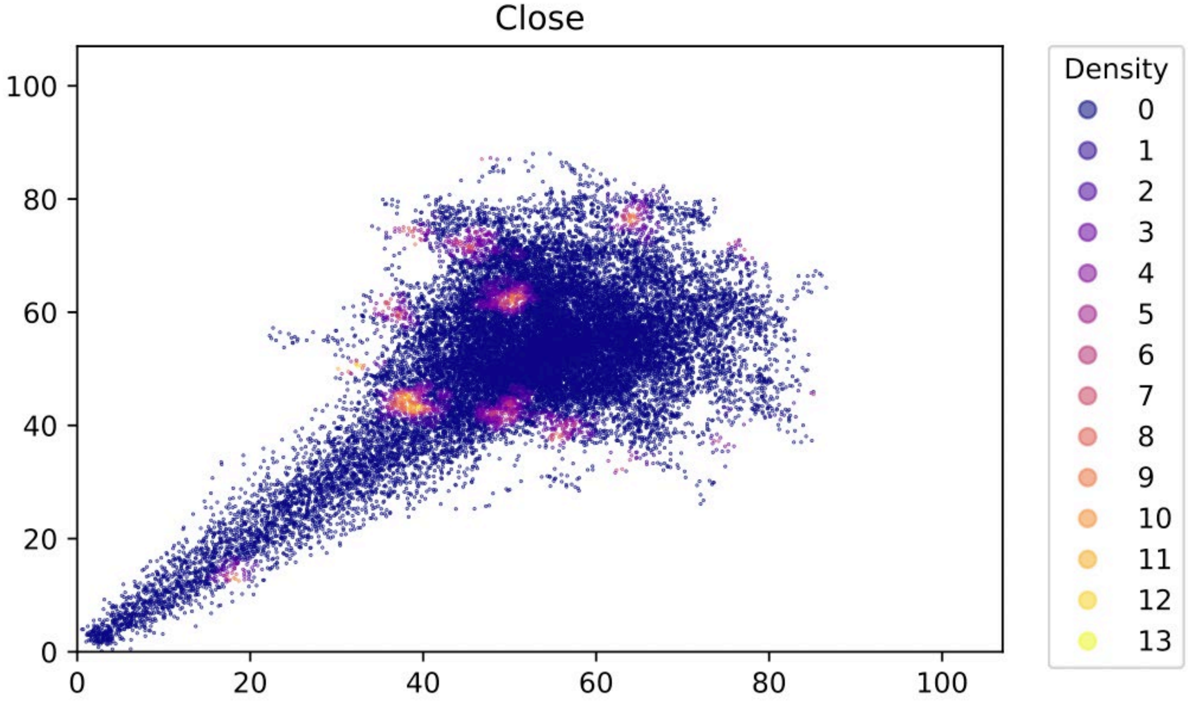
✧ Data Transformation by log function

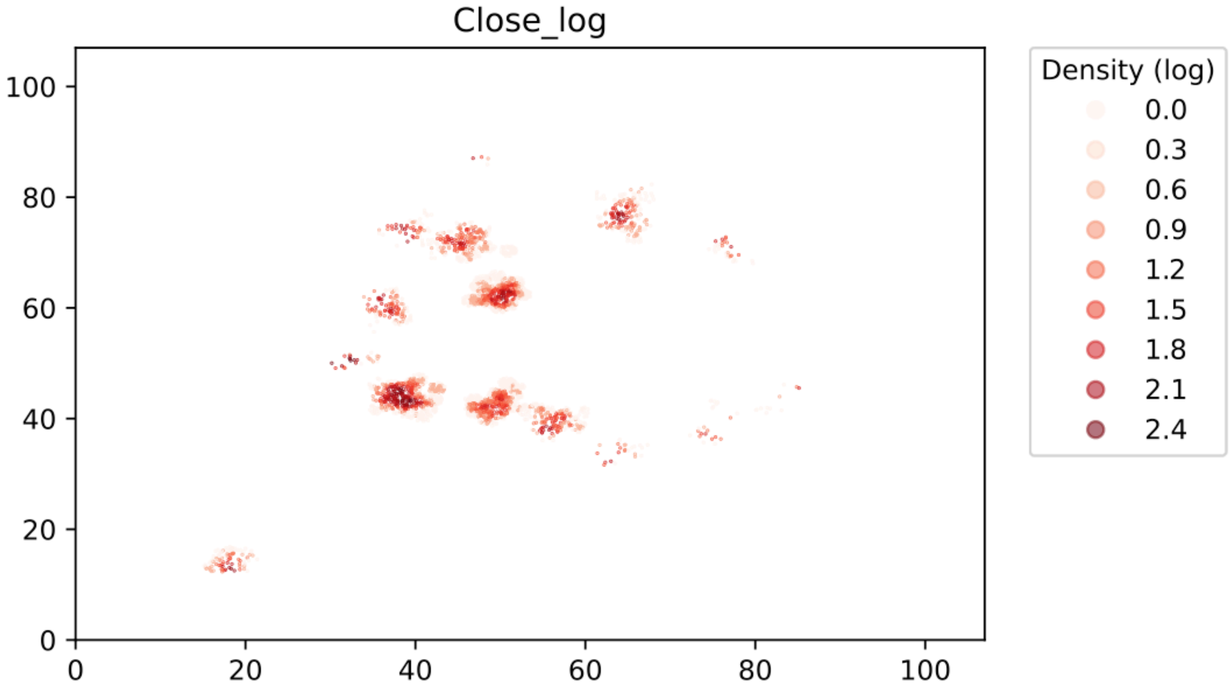




Distribution of Tayes

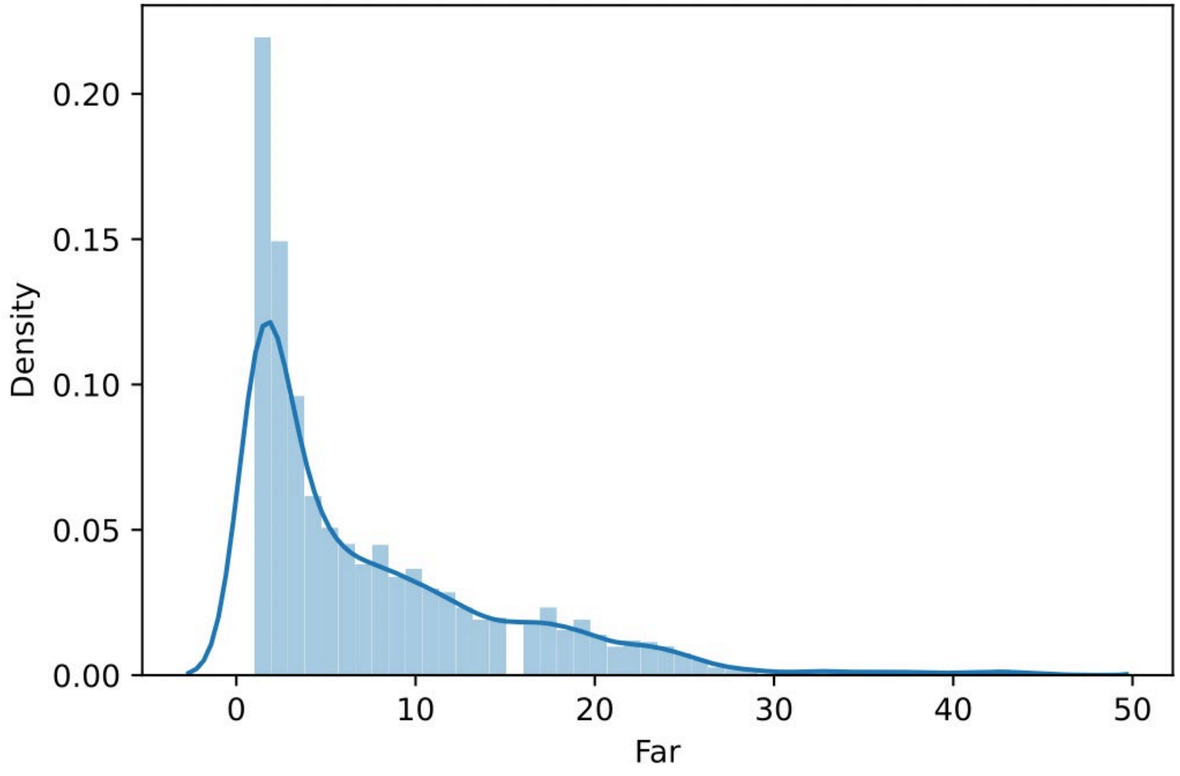
✧ Data Transformation by log function

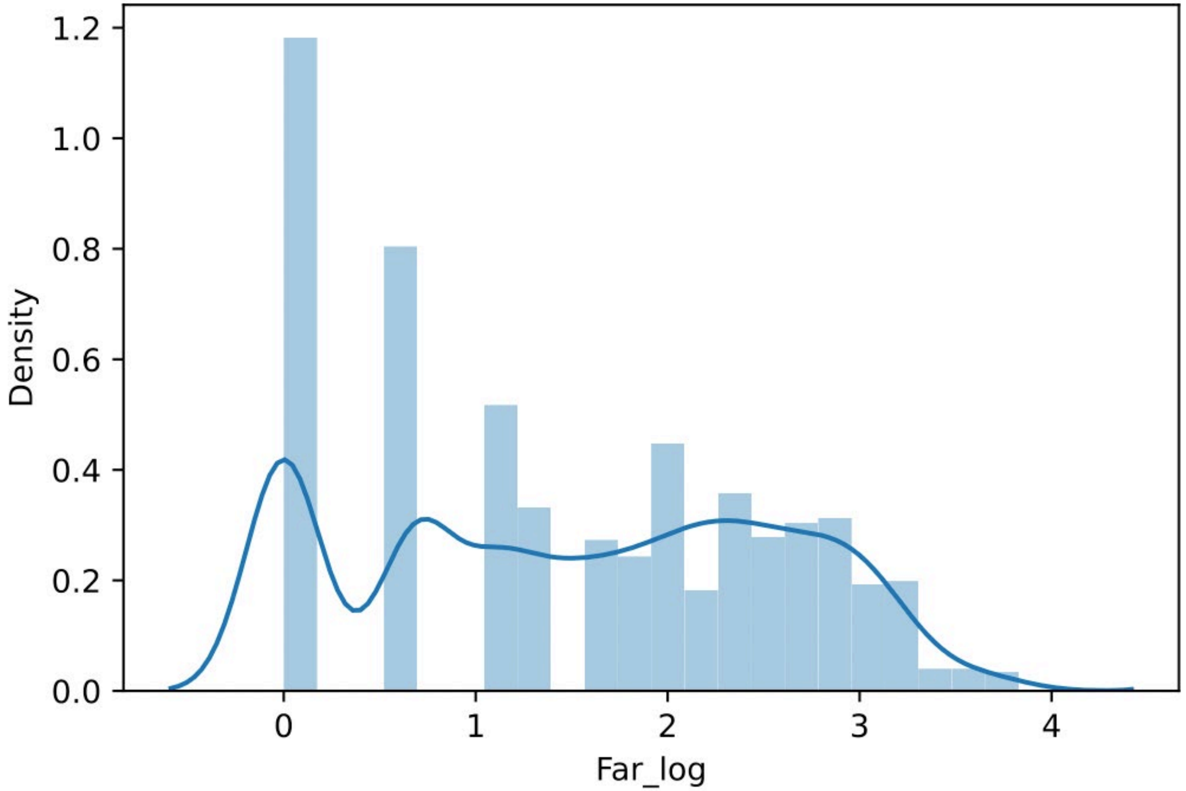




Distribution of Tayes

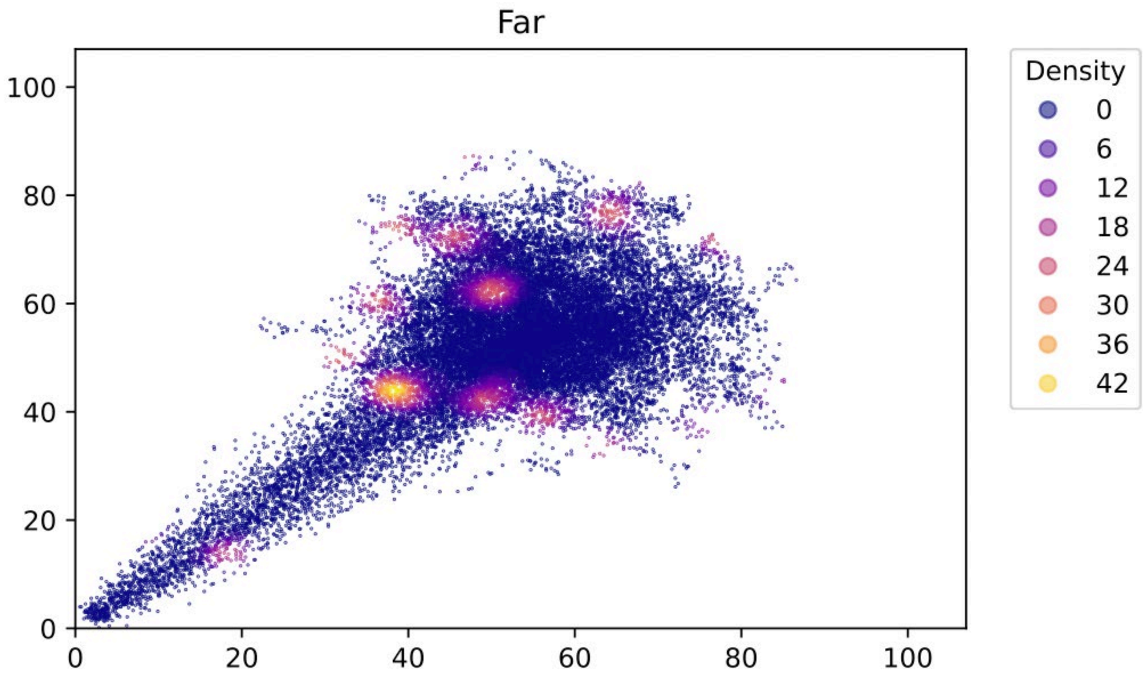
✧ Data Transformation by log function

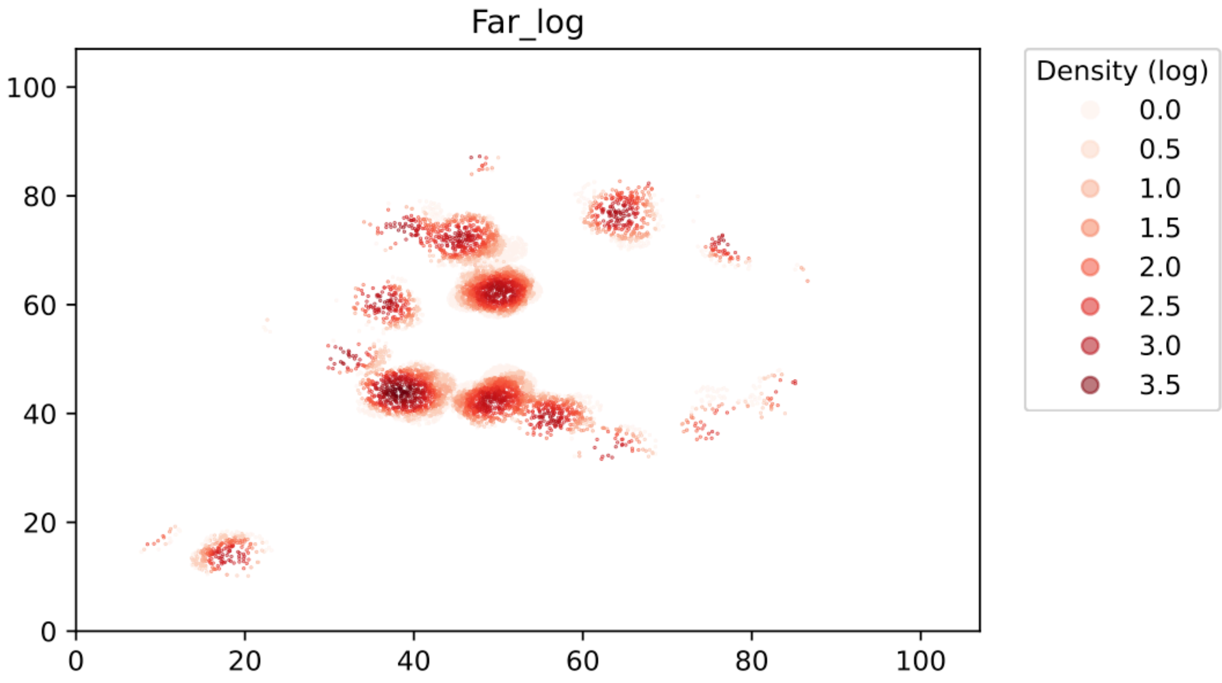




Distribution of Tayes

✧ Data Transformation by log function



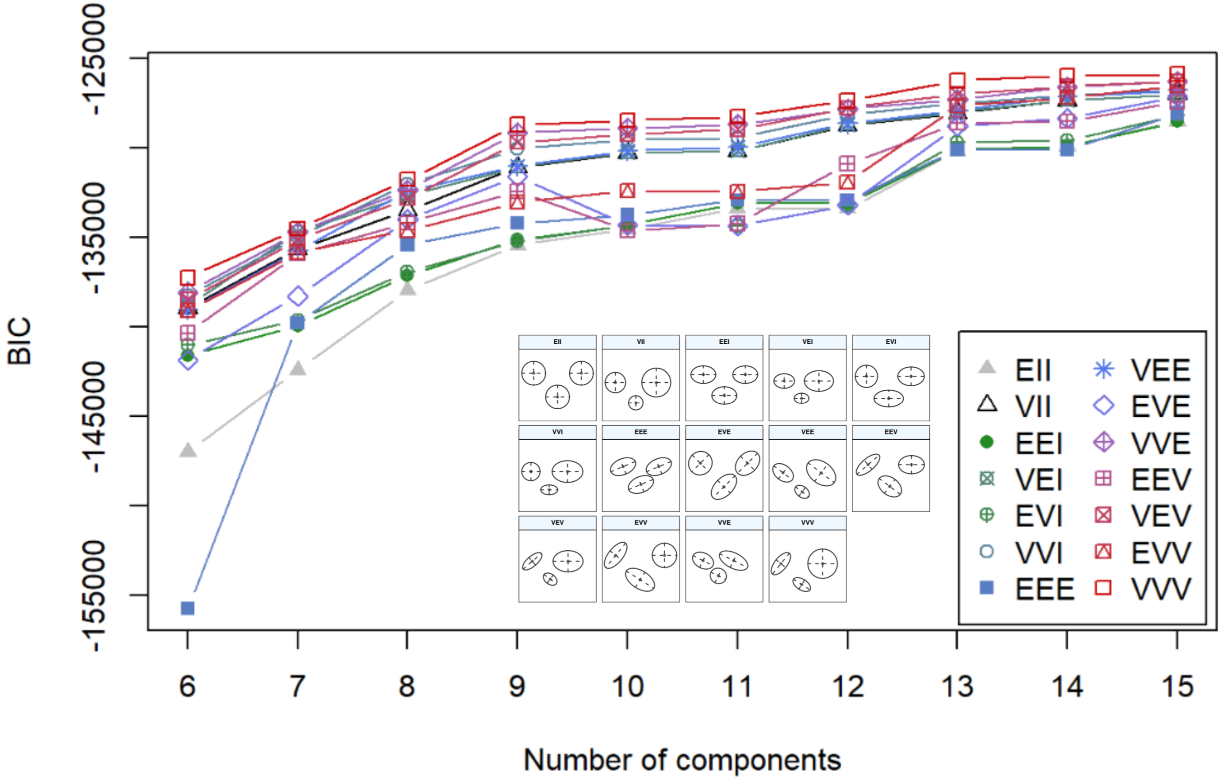


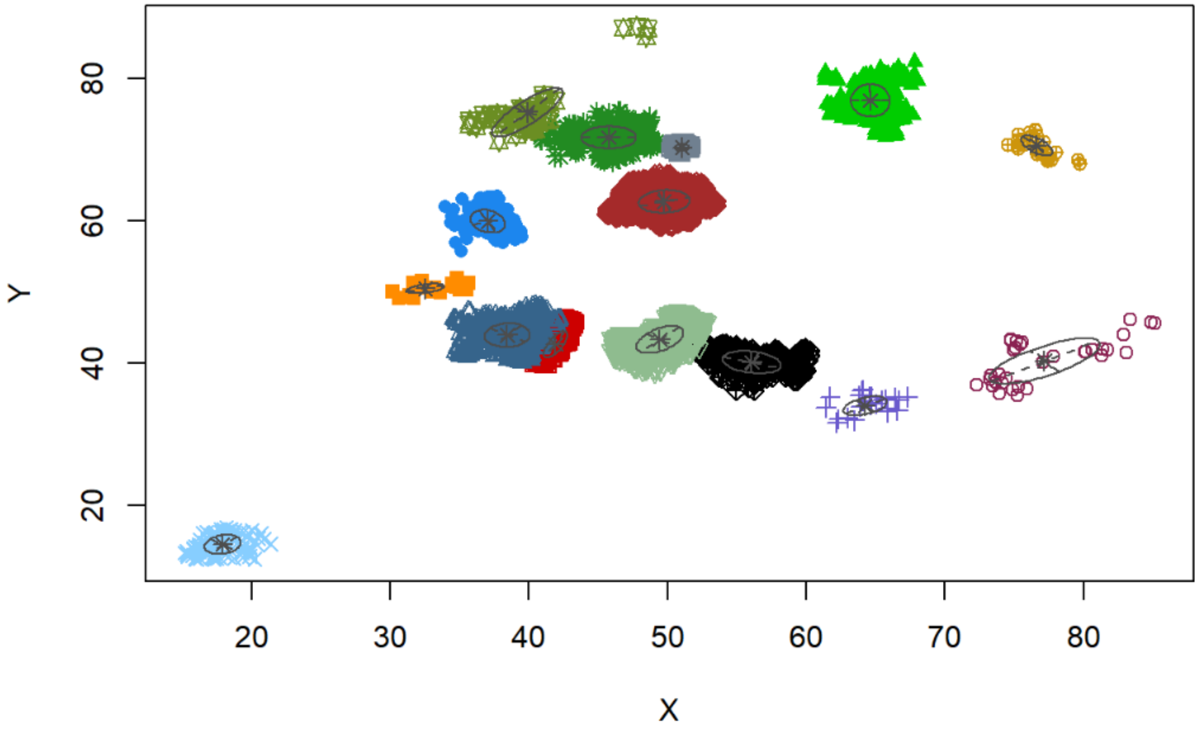
EM Clustering

- Assumption: Gaussian mixture model
- Goal:
  - ✧ estimate the number of clusters (Jiuling)
  - ✧ find out the approximate position of each Jiuling
  - ✧ look for possible improvements based on the result

Analysis

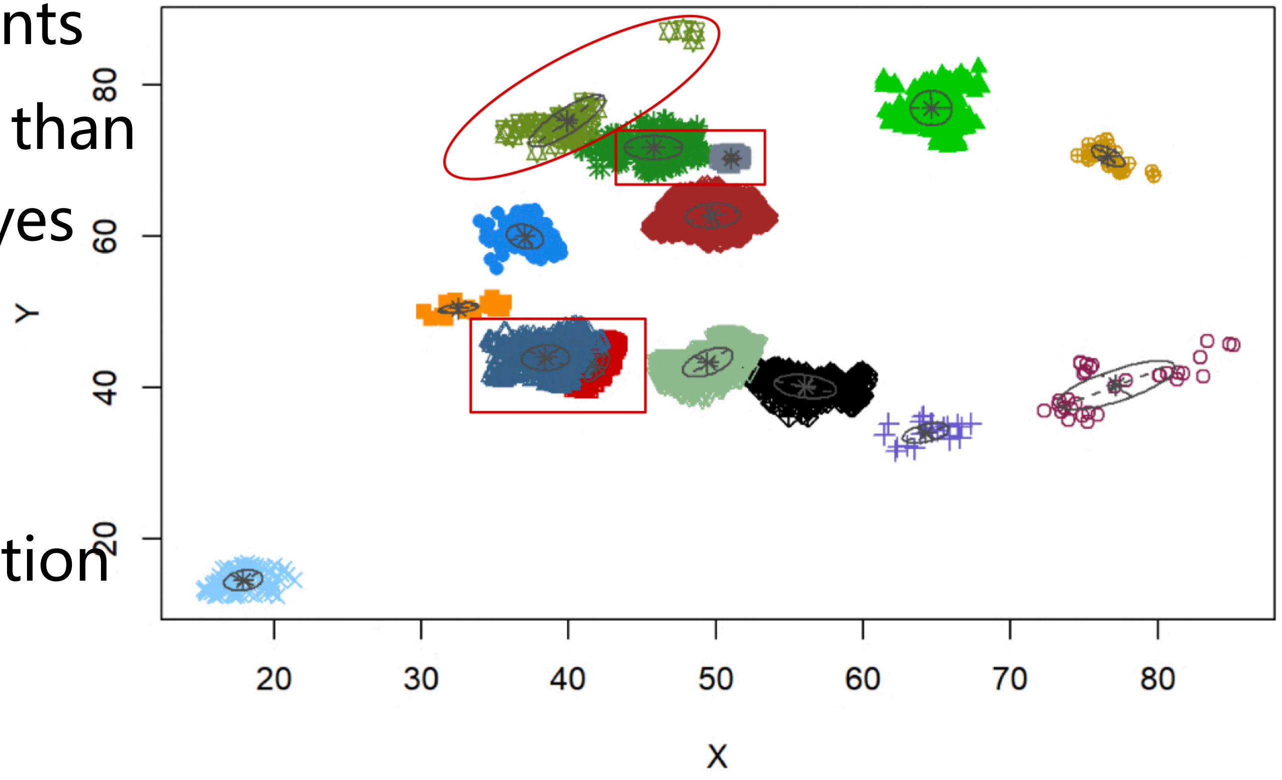
- BIC——Bayesian Information Criterion $BIC = k\ln(n) - 2\ln(L)$





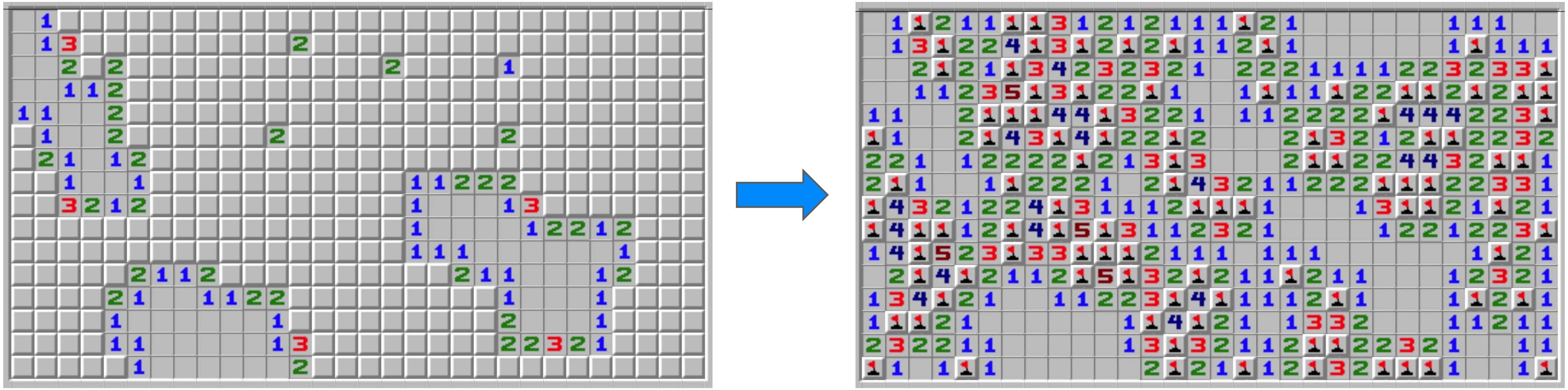
Analysis

- Limitation
  - Concentrated distribution of data points
  - Based on nearby Tayes counts, rather than the actual number and location of Tayes
- Possible solution
  - Aggregation of the data
  - Estimation of actual number and location of Tayes



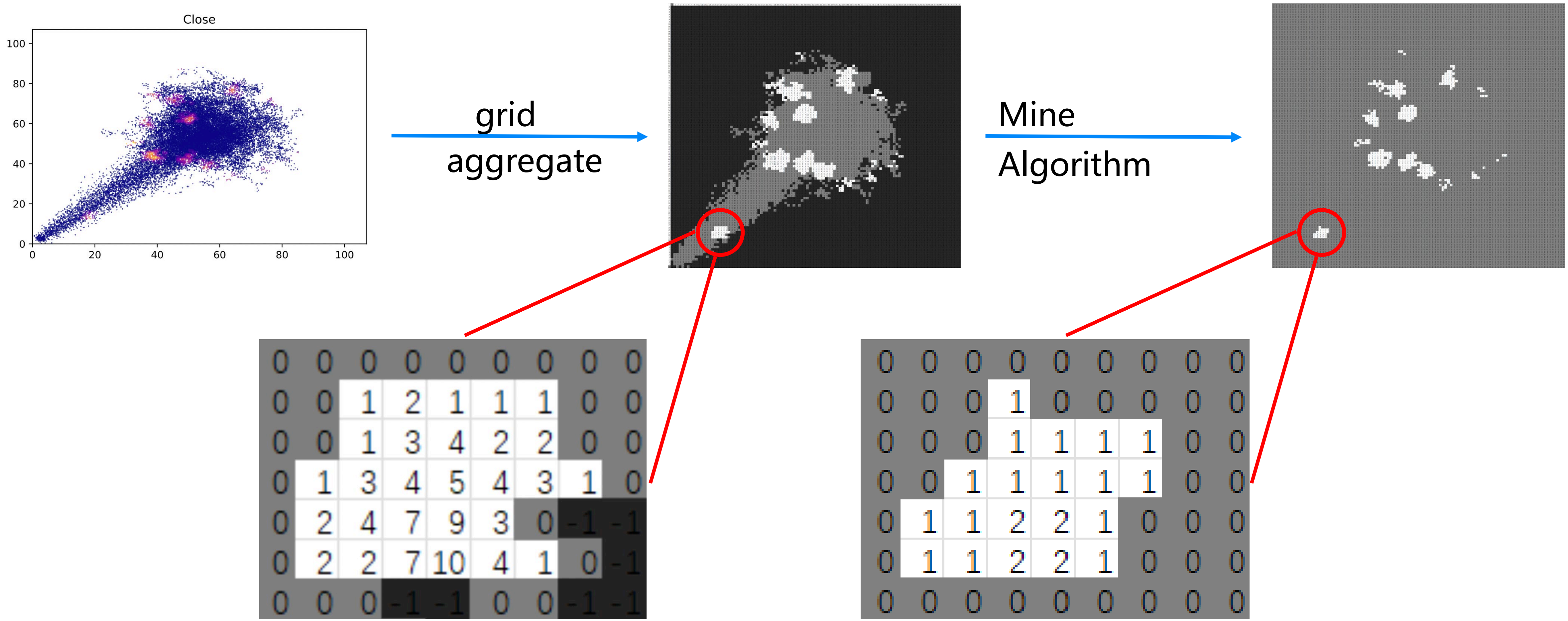
Tayes Distribution Estimation

- Key Problem: Estimate the number and location of Tayes



Maybe we can act as minesweepers

Tayes Distribution Estimation



Mine Algorithm

Algorithm 1: Generate the prediction on fruit distribution

Input: Nearby Fruit Distribution $Target$

Output: Prediction of fruit number and location $Prediction$

function predict(Target)

Initialize $Prediction$, $loss$, tot_{loss}, $\Delta_{tot_{loss}}$,

while $\Delta_{tot_{loss}} > 3$ *or max term in loss* > 2 *or min term in loss* < -2 **do**

- newLoss, MSE, $loss$ = getLoss(target, prediction)
- update $\Delta_{tot_{loss}}$, $\text{tot}_{Loss}, loss_{max}, loss_{min}$
- **if** $loss_{max} + loss_{min} > 0$ *or* $loss_{max} + loss_{min} == 0$ *and* $tot_{loss} > 0$ **then**
  - **while** $loss_{max}$ *doesn't change* **do**
    - x, y is the position of $loss_{max}$
    - increase $Prediction[x, y]$
    - update $\Delta_{tot_{loss}}$, $\text{tot}_{Loss}, loss_{max}, loss_{min}$
- **else if** $loss_{max} + loss_{min} < 0$ *or* $loss_{max} + loss_{min} == 0$ *and* $tot_{loss} < 0$ **then**
  - **while** $loss_{min}$ *doesn't change* **do**
    - x, y is the position of $loss_{min}$
    - decrease $Prediction[x, y]$
    - update $\Delta_{tot_{loss}}$, $\text{tot}_{Loss}, loss_{max}, loss_{min}$
- **else**
  - **return** $Prediction$

Algorithm 2: Sum the quantities in a cross area of $Prediction$

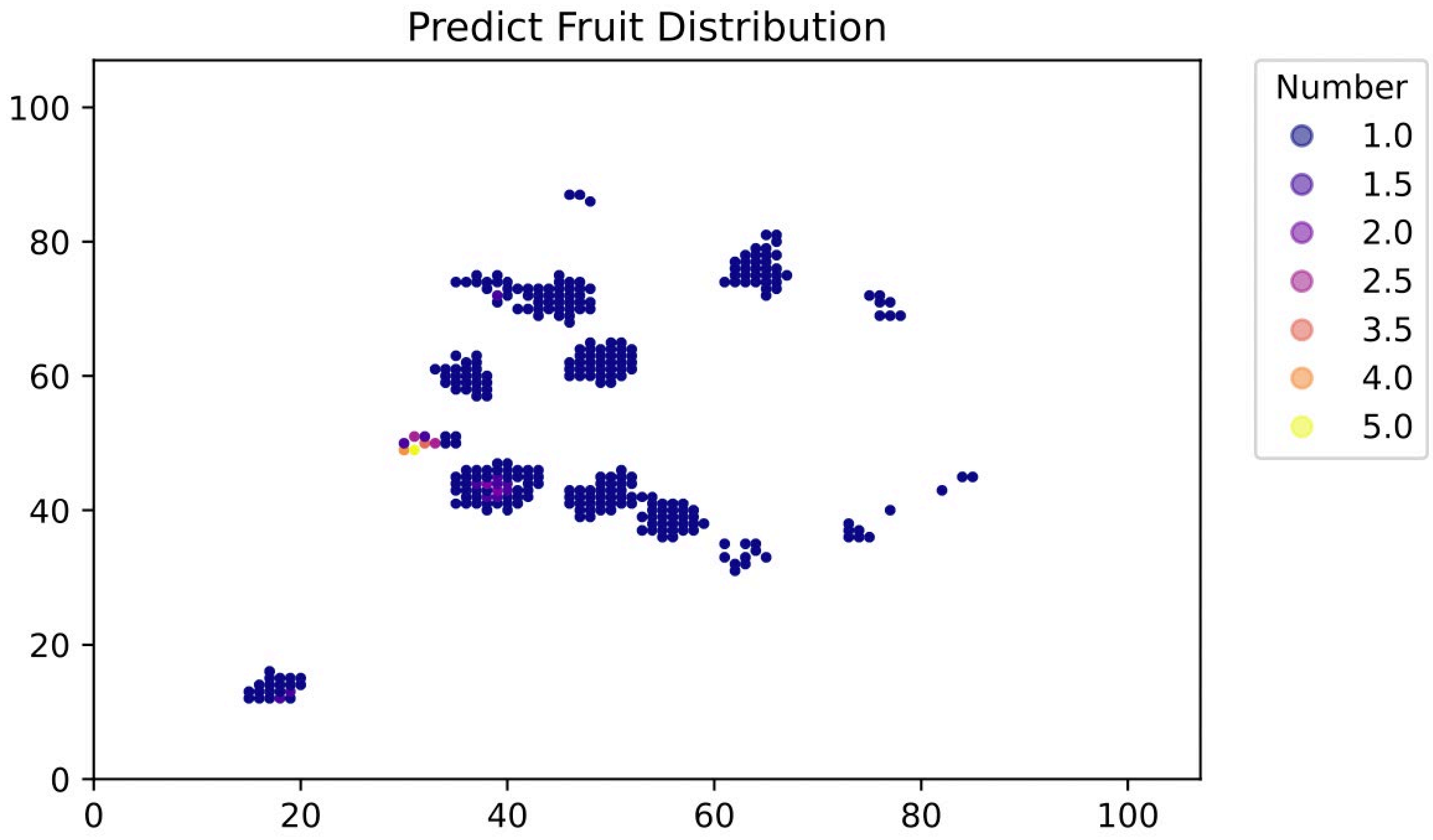
Input: Prediction of fruit number and location $Prediction$, Center Position (x, y)

Output: local sum of the cross area centering at (x, y)

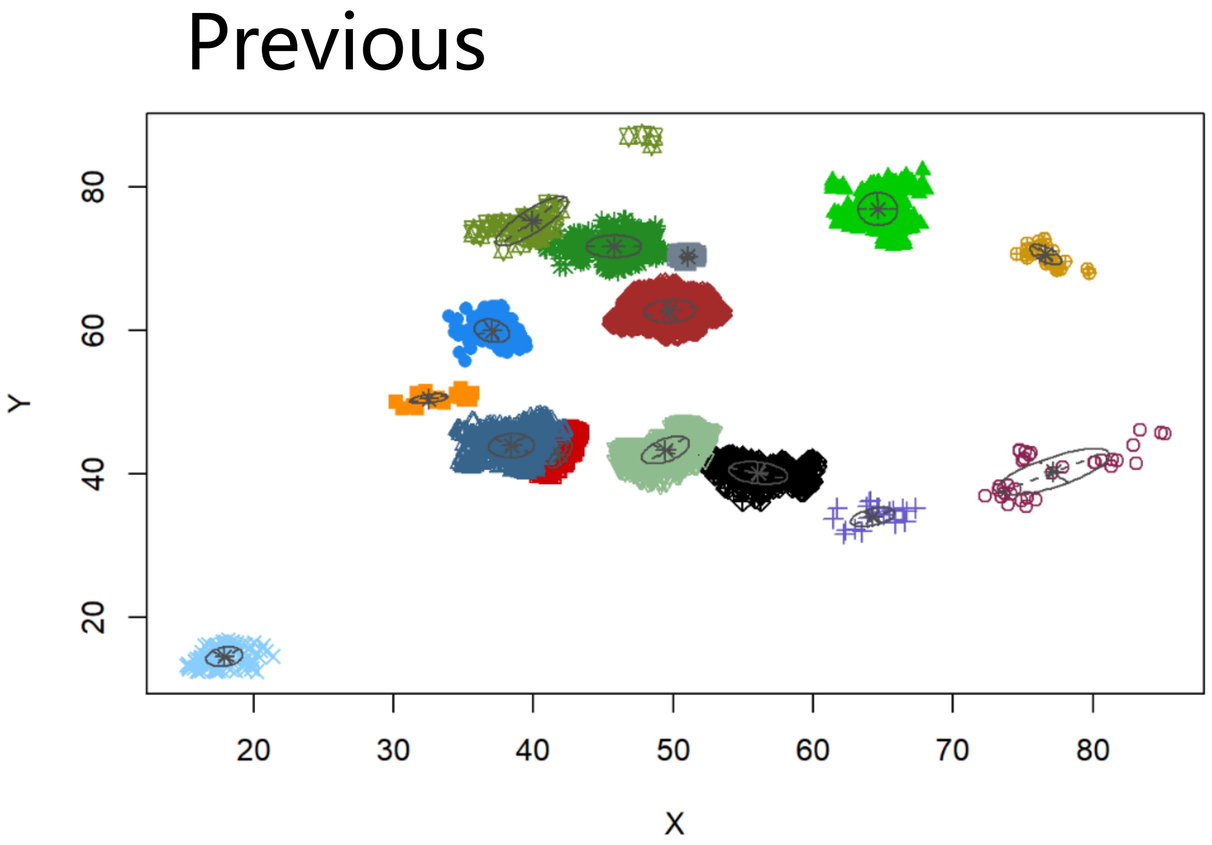
function LocalSum($Prediction$, x, y)

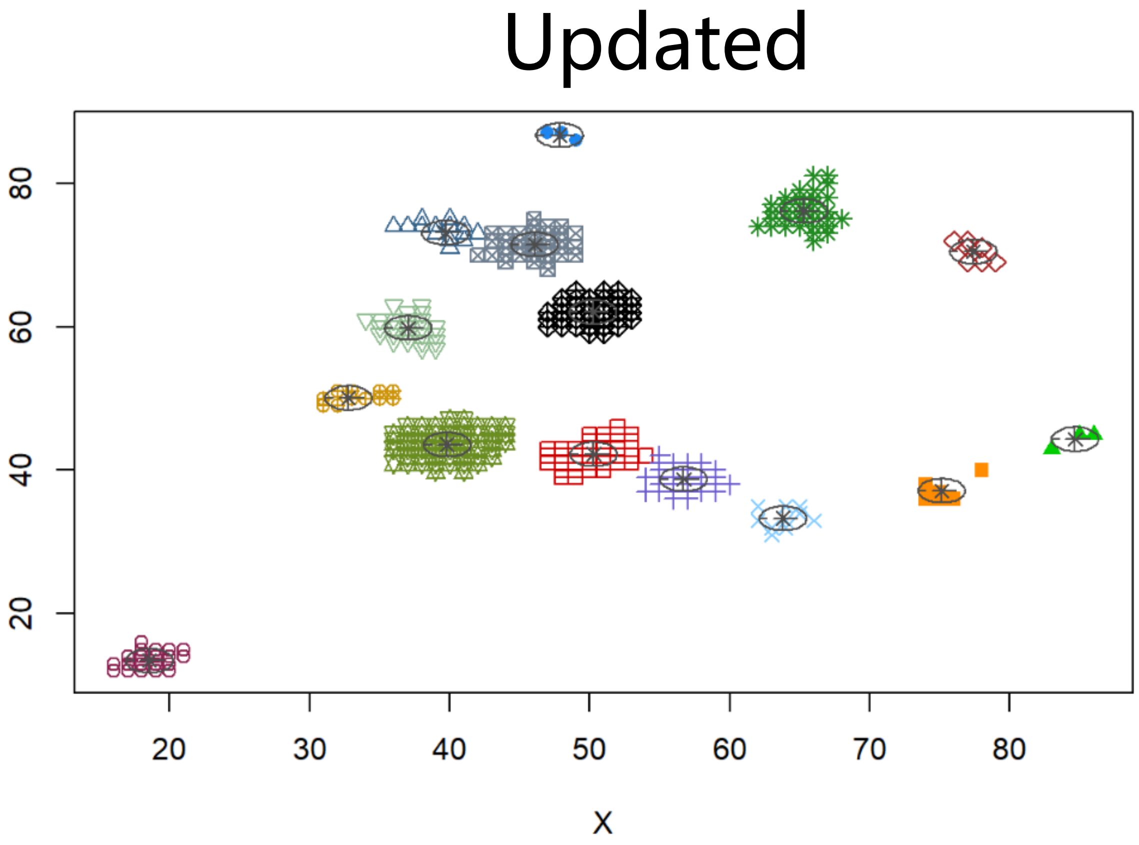
return $Prediction[i, j-1 : j+2] + prediction[i-1 : i+2, j] - prediction[i, j]$

Tayes Distribution Estimation

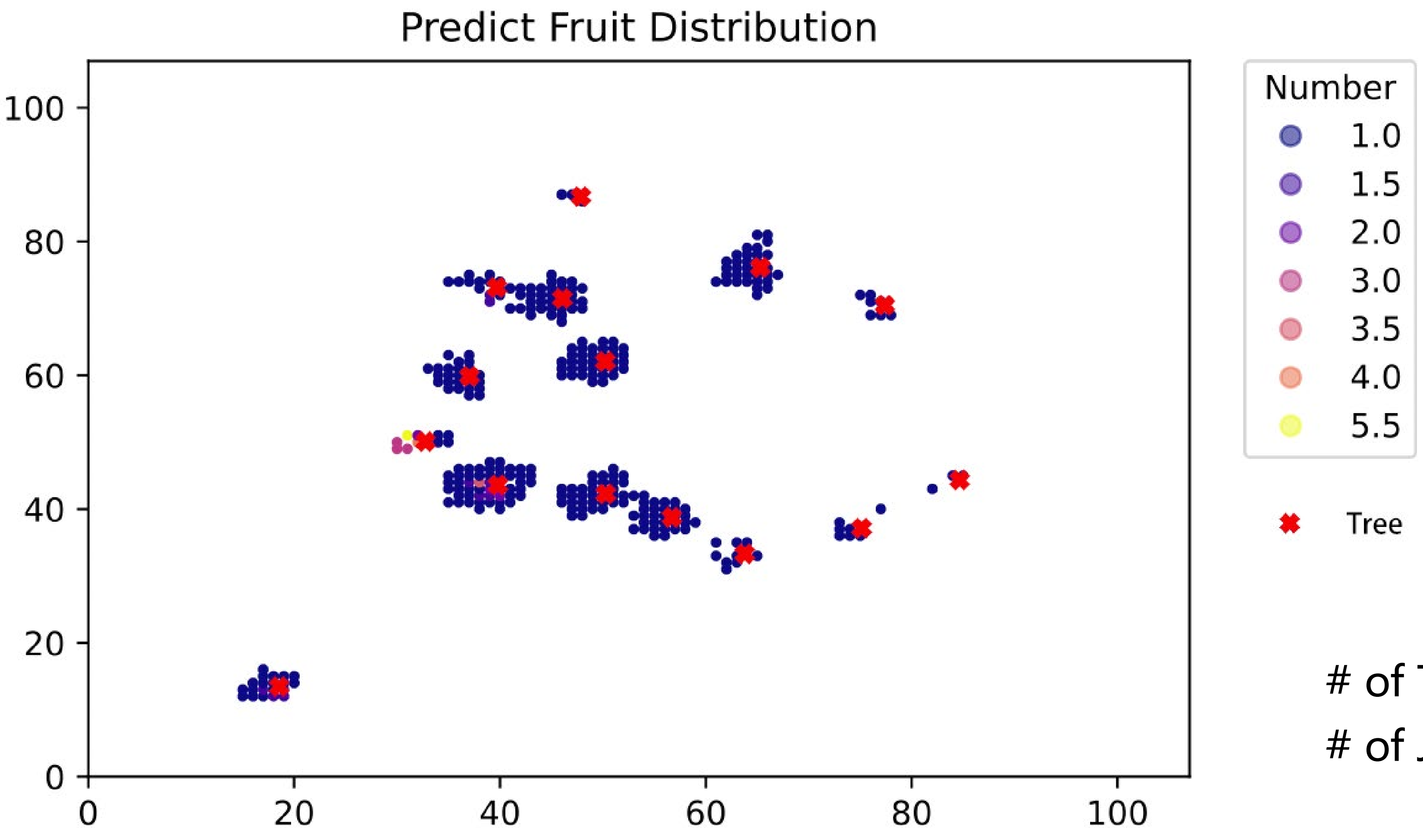


Result Comparison





Tayes Distribution Estimation

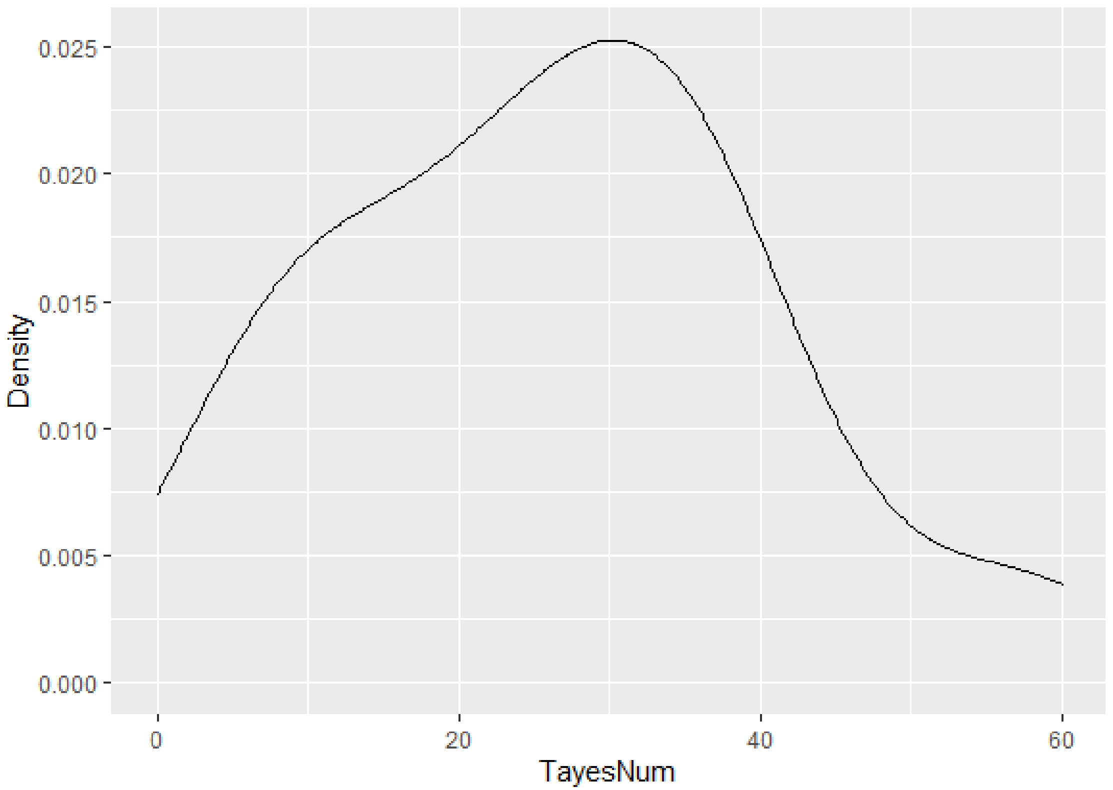


# of Taye: 351

# of Jiuling: 15

Tayes Number pre Jiuling

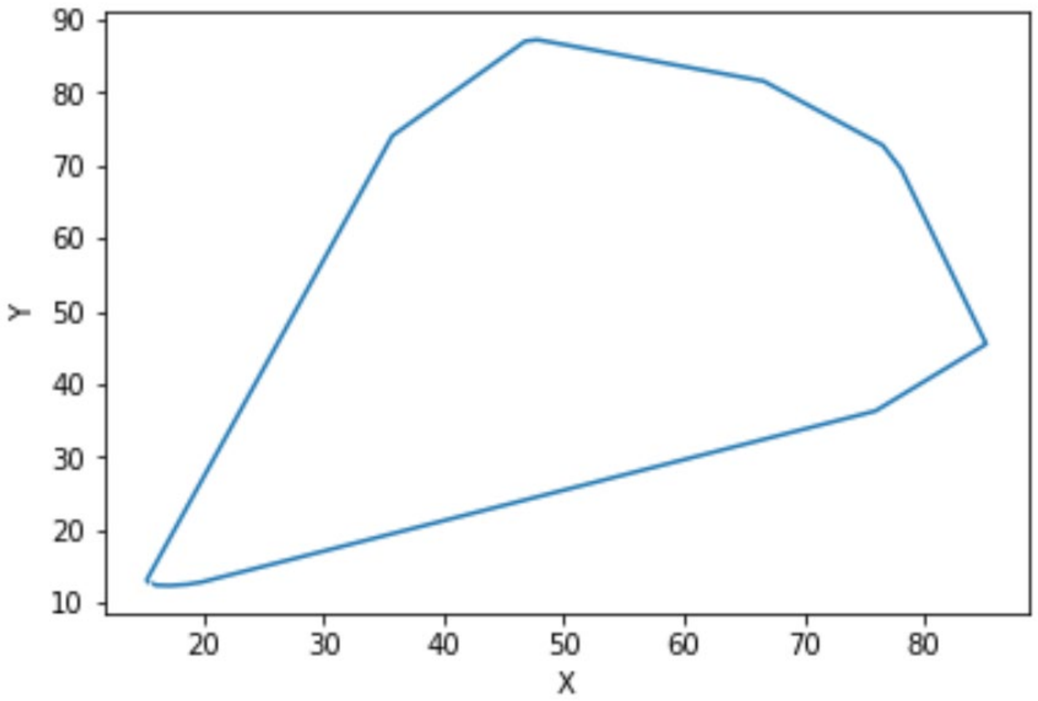
- Roughly a Gaussian Distribution with μ = 30
- Left part slightly higher than Gaussian Distribution
  - little information on clusters at the edge of explored area



What about the total number of Jiuling?

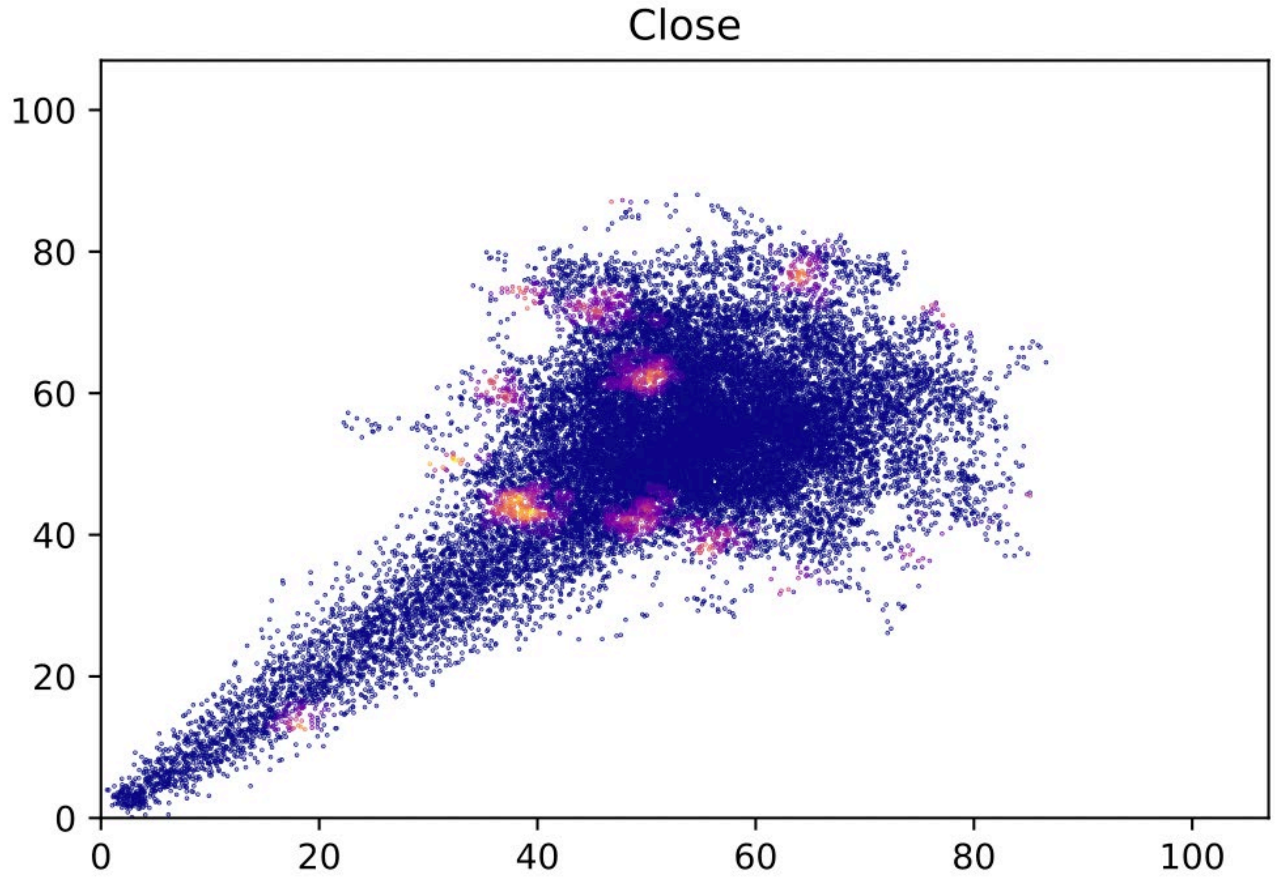
- Assumption
  - Jiuling are uniformly randomly distributed
- Quantity predicted >> Total quantity
- Area detected >> Total area
- Estimation of area detected
  - Modified Graham-scan Algorithm
  - Sum of detected grid

Regular Graham-scan Algorithm

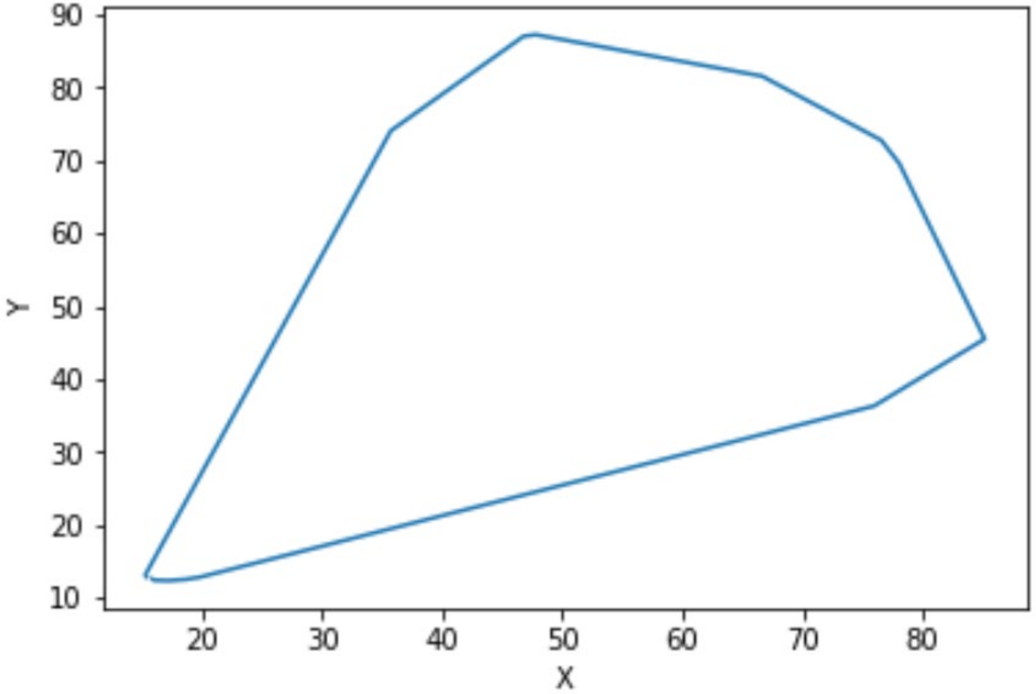


Regular Graham-scan algorithm

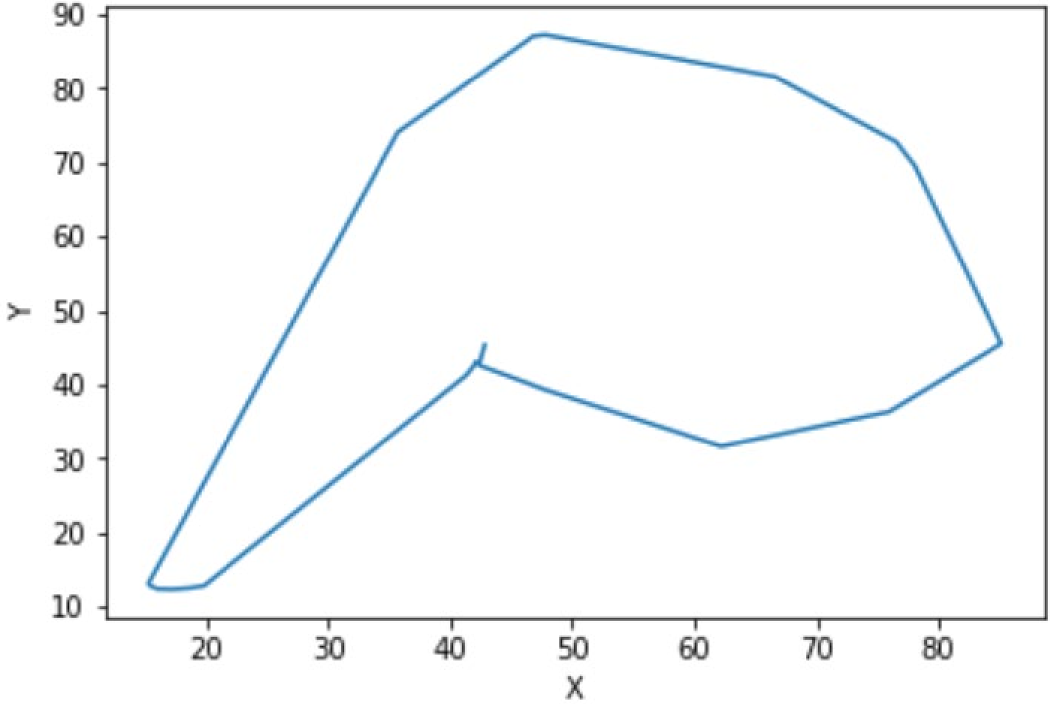
Size : 2927.89



Modified Graham-scan Algorithm



Regular Graham-scan algorithm
Size : 2927.89
Proportion : 25.57%



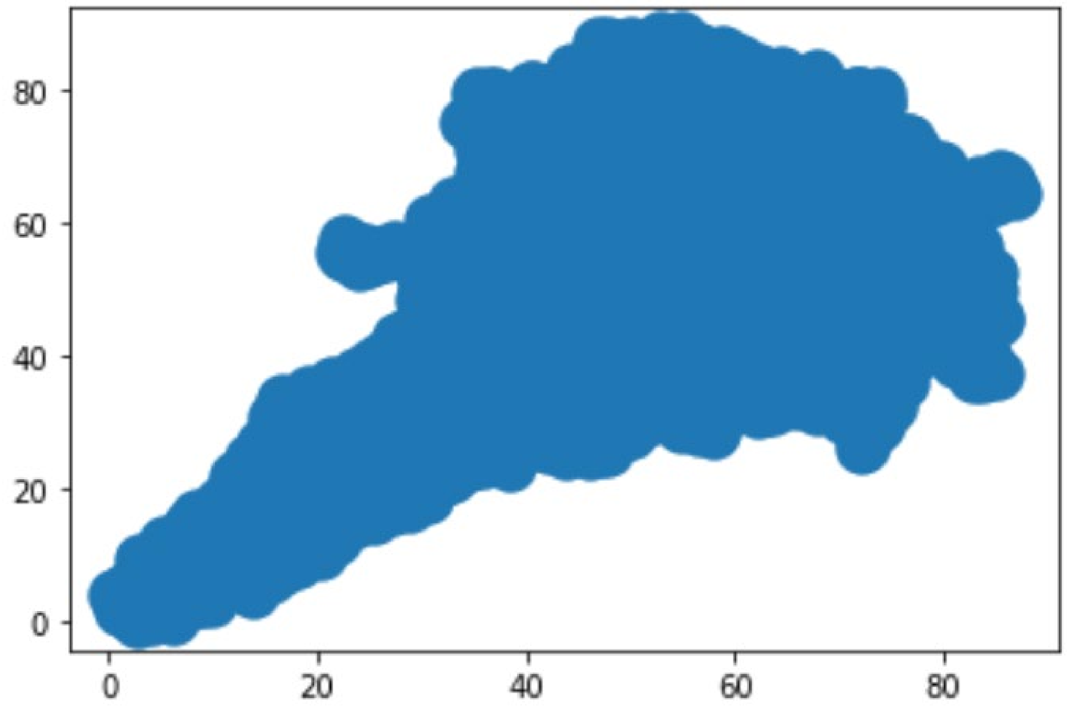
Modified Graham-scan algorithm
Size : 2480.89
Proportion : 21.66%

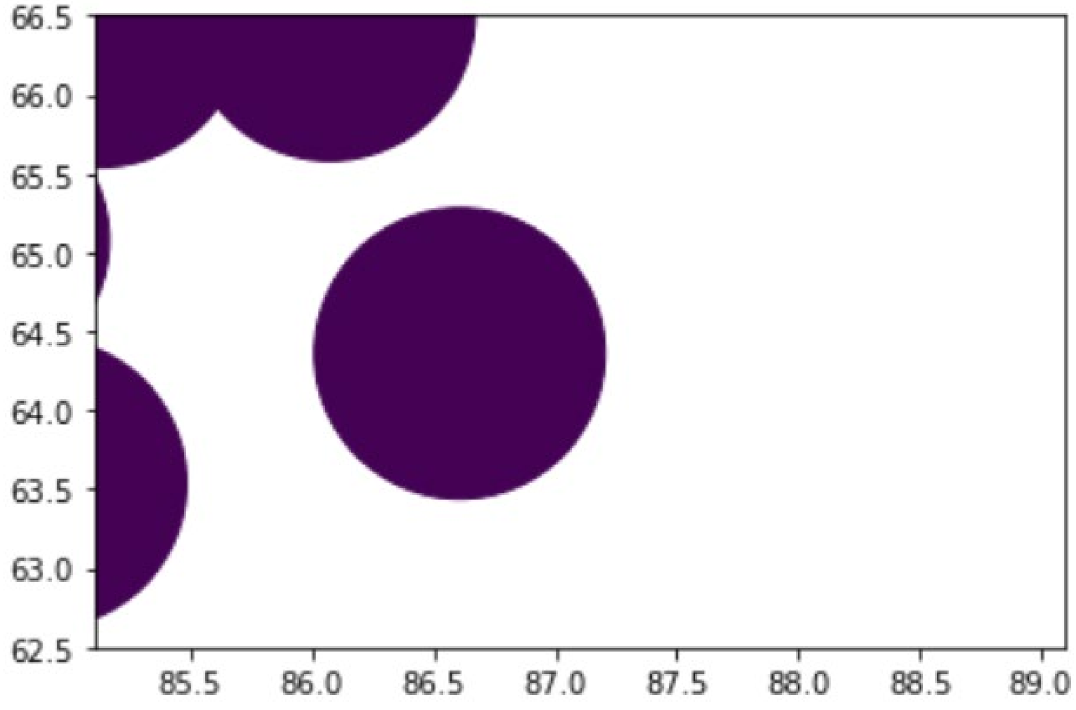
Area Estimation

General idea:

- Divide the forest into 107 *107 **tiny** grids
- For each grid, if there exists any record of trip within one meter, mark this grid as **detected**.
- Calculated the **total size** of grids detected.

Area Estimation





Total Size : 3228

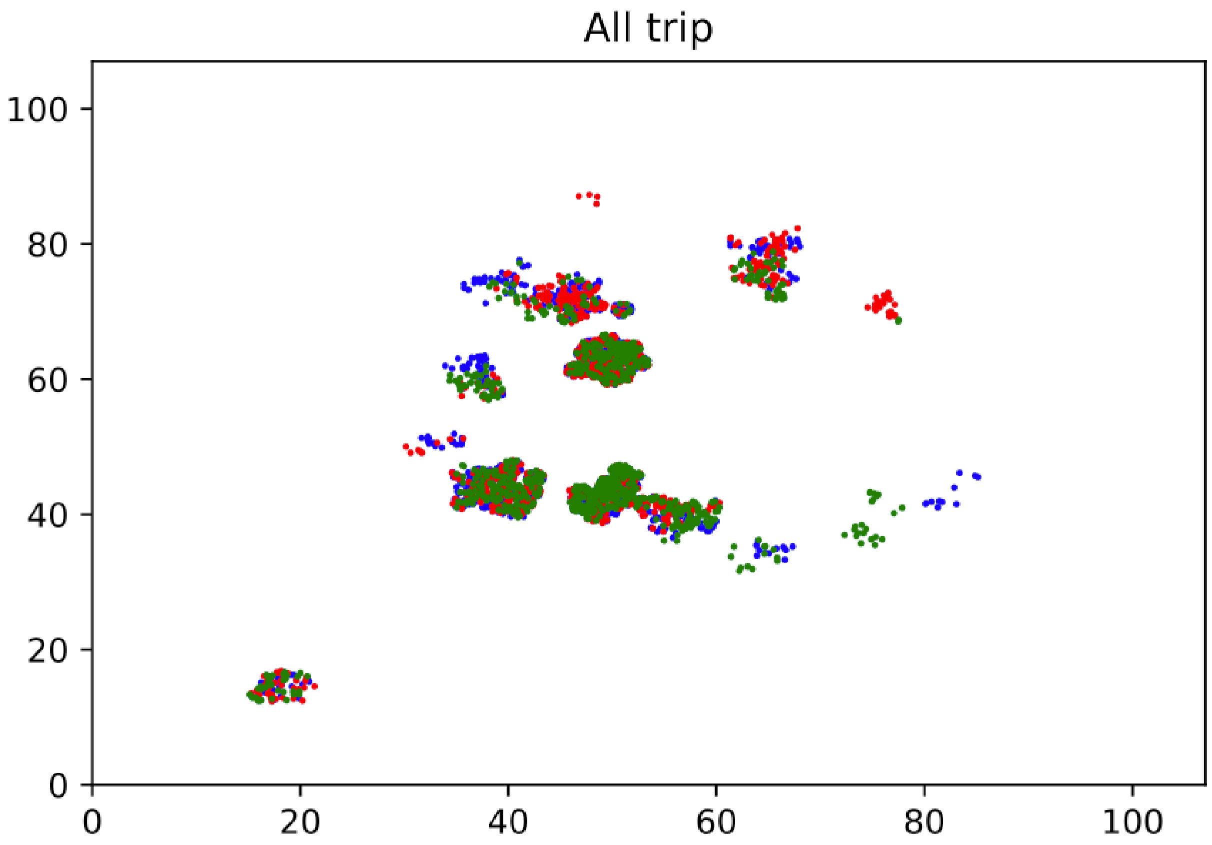
Proportion : 28.19%

Area Estimation

Method	Value	Proportion	Tree cnt
Regular Graham scan	2927.89	25.57%	58.66
Modified Graham scan	2480.89	21.66%	69.25
Circular Coverage	3228	28.19%	53.21
Loose Grid	2743	23.96%	62.60
Strict Grid	2283	19.94%	75.23

What if Jiuling can move?

- Two key questions:
  - ✧ Existence of Tayes (permanent or not)?
  - ✧ How Tayes fall from Jiuling (continuous or periodic)? (frequent or rare)?



And we'll Need:

- Starting time of each trip
- More groups of data covering the whole forest
- More people travelling at the same time

Work Distribution

Name	Aka	Contribution
Fan Chen	PowerPointer	Area Estimation, K-means
Qiansiqi Hu	EMaster	EM Clustering, Problem Analysis
Shengyuan Xu	MachineLearner	Taye num&loc Estimation, Data Visualization

Thanks!