CRIME LOCATION PREDICTION

Crime Alert

- Min Zhang
- Dong An

- Introduction
- Basic Theory
- Assumption
- Data
- Method
- Solution
- Conclusion

INTRODUCTION

- Define the problem

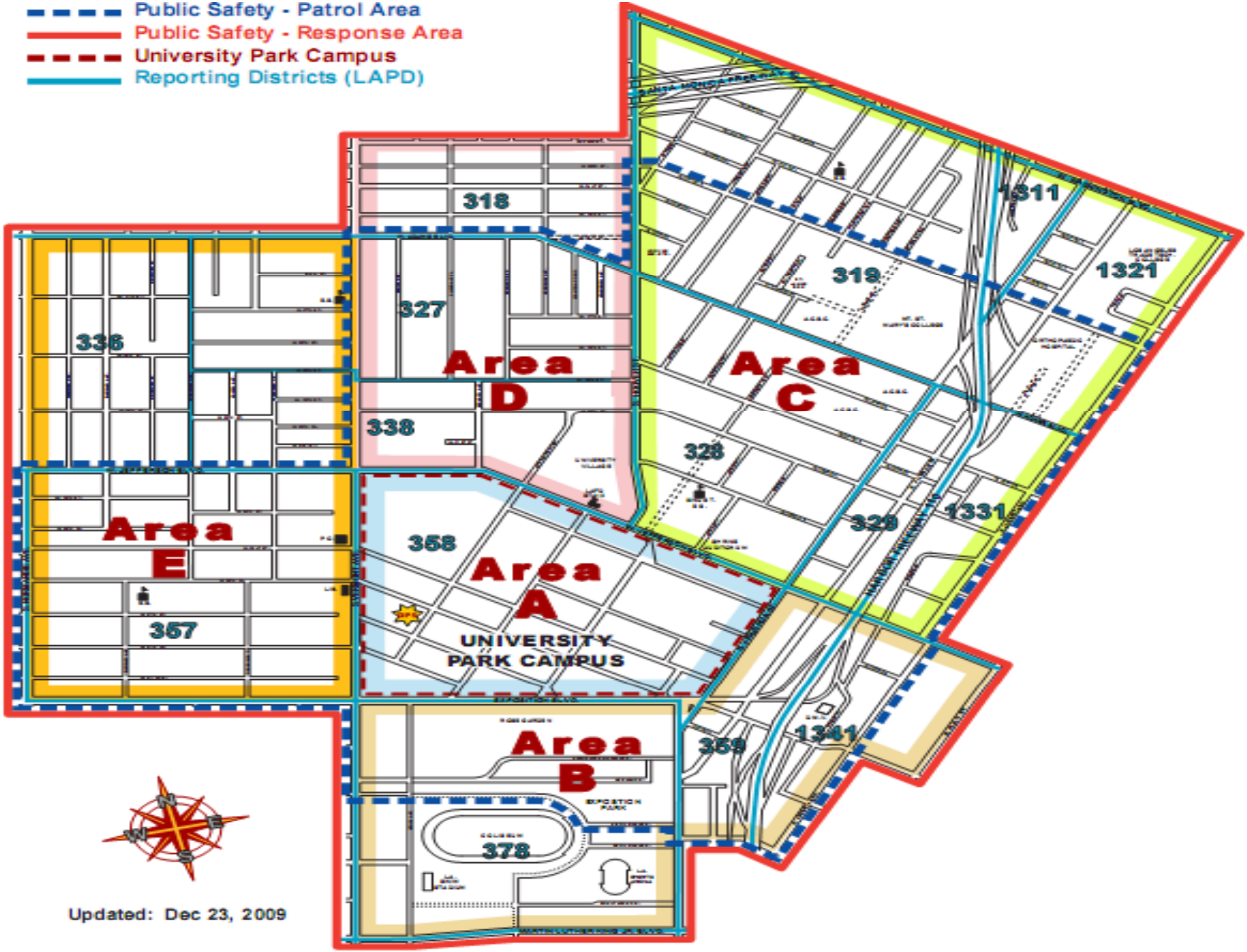
  Predict

  a. probability of each specific area that crime will happen accounts for the whole area.

  b. which area will have the highest density of crime and its probability of crime accounts for the whole area in a specific time period.

INTRODUCTION

- Prediction area


Public Safety - Patrol Area
Public Safety - Response Area
University Park Campus
Reporting Districts (LAPD)
318
1311
319
1321
327
336
Area
D
Area
C
338
328
329
1331
Area
E
358
Area
A
357
UNIVERSITY
PARK CAMPUS
1341
359
Area
B
378
Updated: Dec 23, 2009

BASIC THEORY

- Crime pattern theory

  supposes that crime is not random. The location of a crime is likely near a criminal's normal activity space. The normal activity space is the collection of areas where the individual most frequently comes into contact with others.

- Routine activity theory

  supposes that for a crime to take place three requirements needed to be present: a motivated offender, a suitable target, and absence of capable guardians.

ASSUMPTION

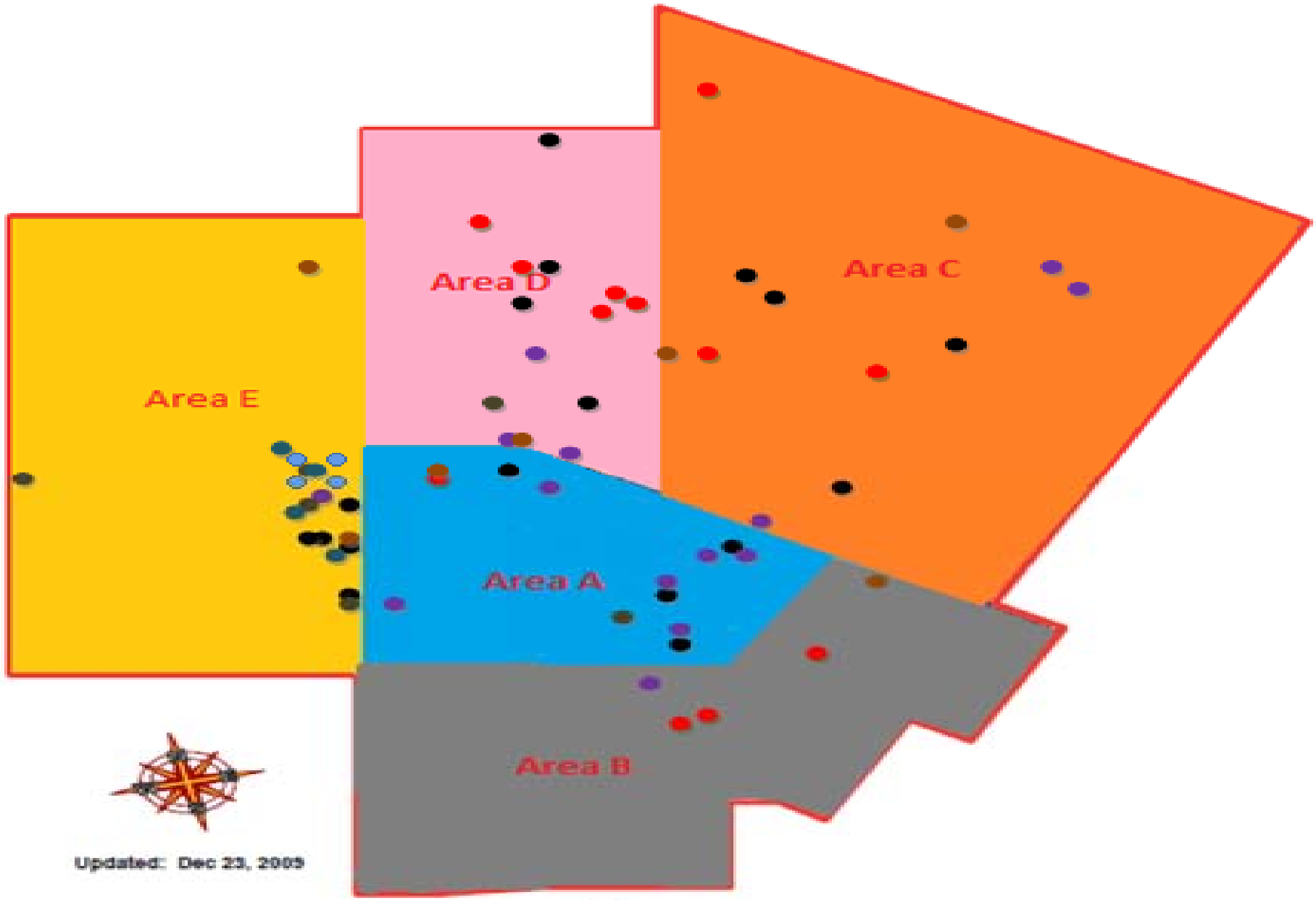
- Criminal is more likely to choose the same area for next crime.
- The security situation of each area in the prediction period has not changed during times.
- Criminals' motivation will not be considered in models.

DATA

- Crime alert archive for 2009 and 2010. It includes 6 semesters' crime information in our prediction area.

UNIVERSITY PARK CAMPUS

Patrol and Response Boundaries

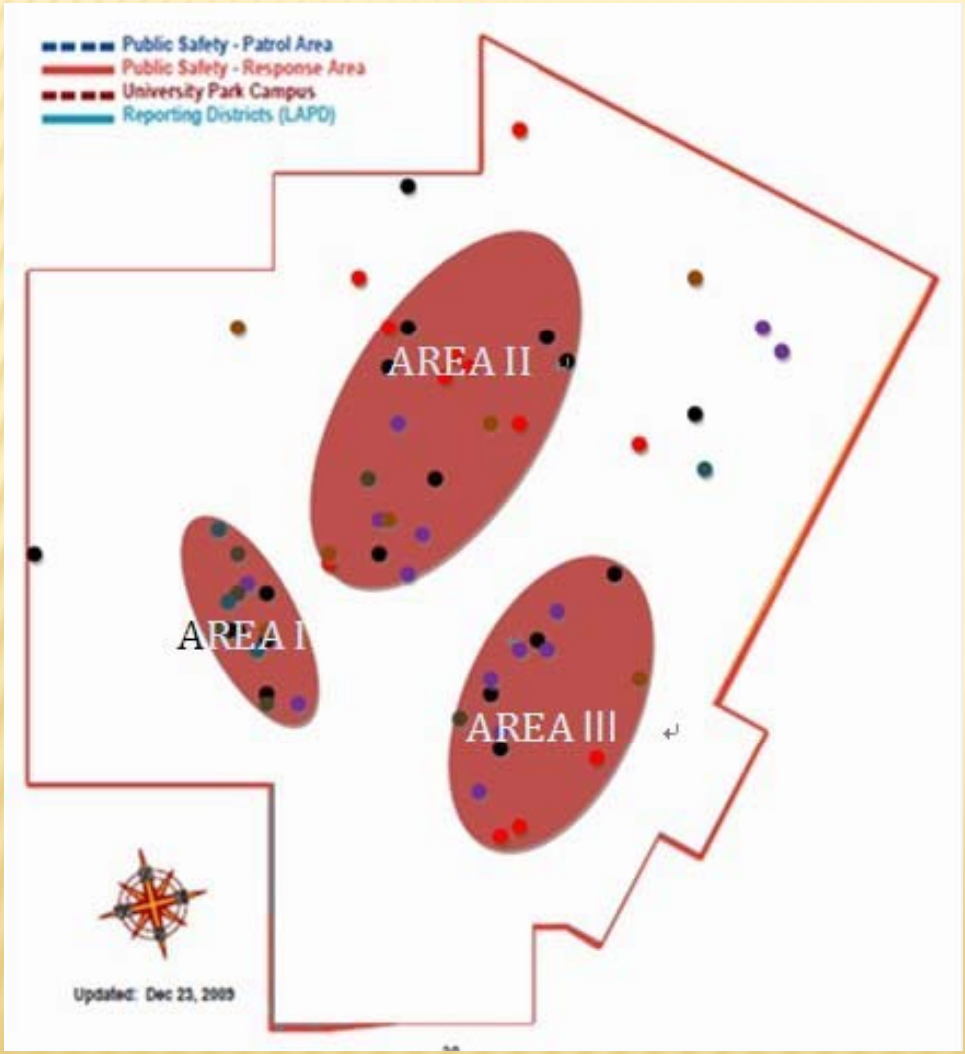


AREA	Spring 2009	Summer 2009	Fall 2009	Spring 2010	Summer 2010	Fall 2010
A	4	2	8	1	1	
B		3	1	1		
C	3	2	2	1		1
D	4	5	4	2		
E	5		1	2	4	3

METHOD

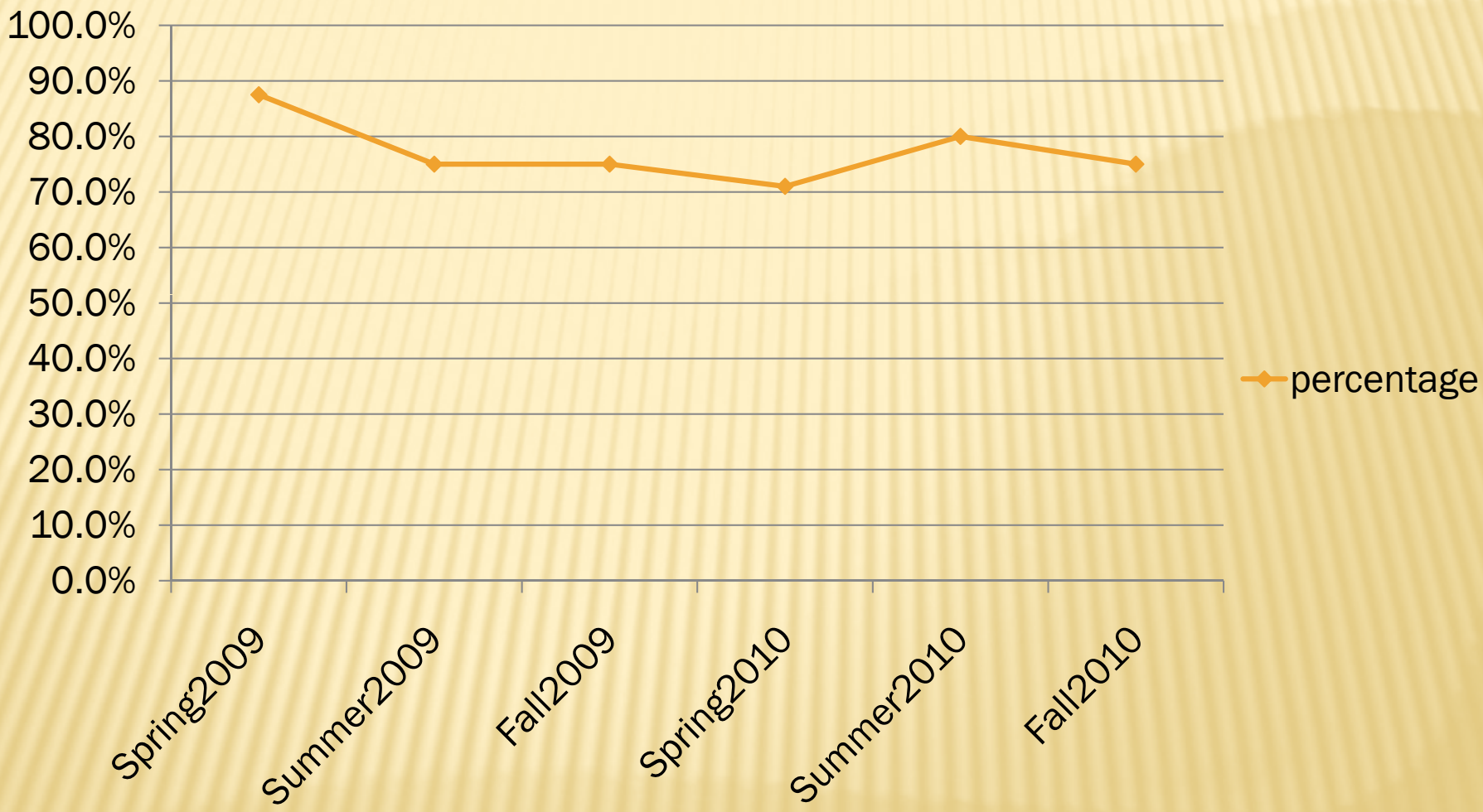
- Markov Chin
- Clustering √
- Liner regression √

CLUSTERING



- Three clustering (most dense area)
- Area Ⅰ
- Area Ⅱ
- Area Ⅲ

AREA	Spring 2009	Summer 2009	Fall 2009	Spring 2010	Summer 2010	Fall 2010	Total
I	4	0	3	1	3	3	14
II	7	6	4	3	0	0	20
III	3	3	5	1	1	0	15
Total	16	12	16	7	5	4	60
Percentage	87.5%	75%	75%	71%	80%	75%	82%


100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Spring2009
Summer2009
Fall2009
Spring2010
Summer2010
Fall2010
percentage

LINER REGRESSION

SUMMARY OUTPUT

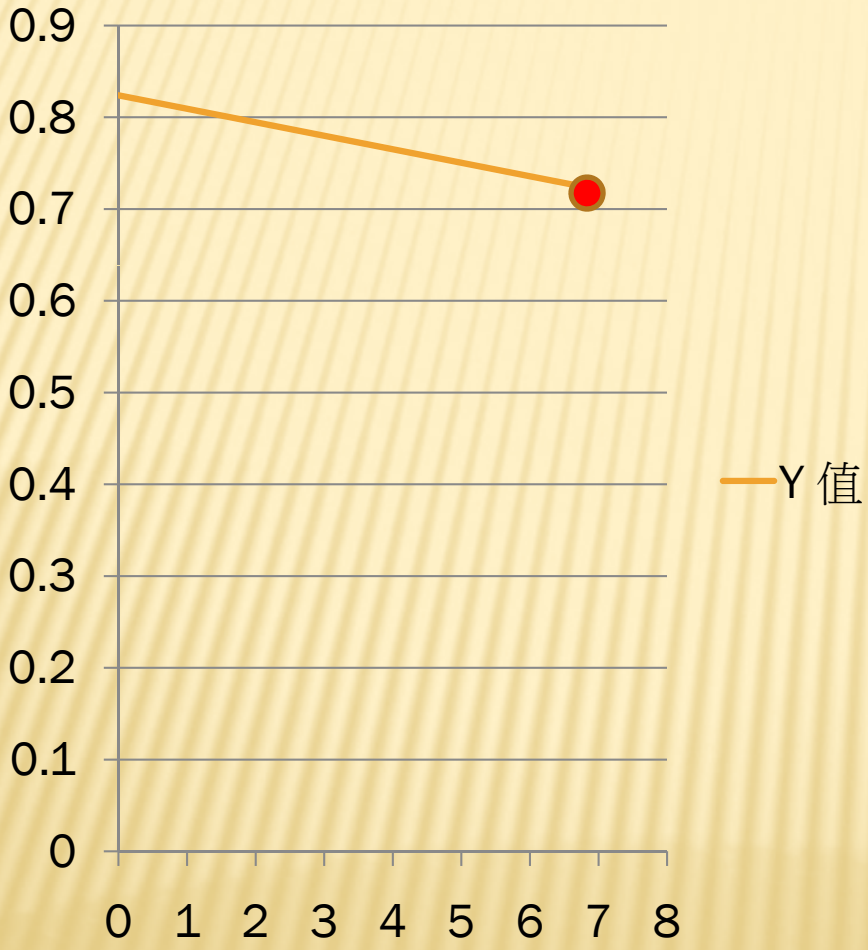
Regression Statistics	
Multiple R	0.4765
R Square	0.227052
Adjusted R Square	0.033815
Standard Error	0.056786
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.003789	0.003789	1.174992	0.339346
Residual	4	0.012899	0.003225		
Total	5	0.016688			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.824	0.052865	15.58692	9.89E-05	0.677224	0.970776	0.677224	0.970776
X Variable 1	-0.01471	0.013574	-1.08397	0.339346	-0.052403	0.022974	-0.0524	0.022974

SOLUTION



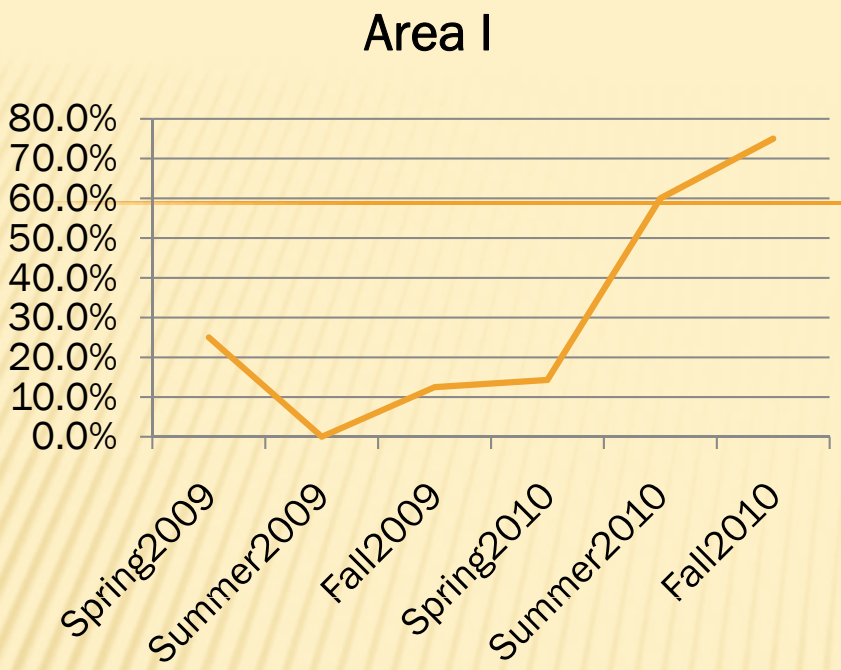
Y=0.824 – 0.01471X

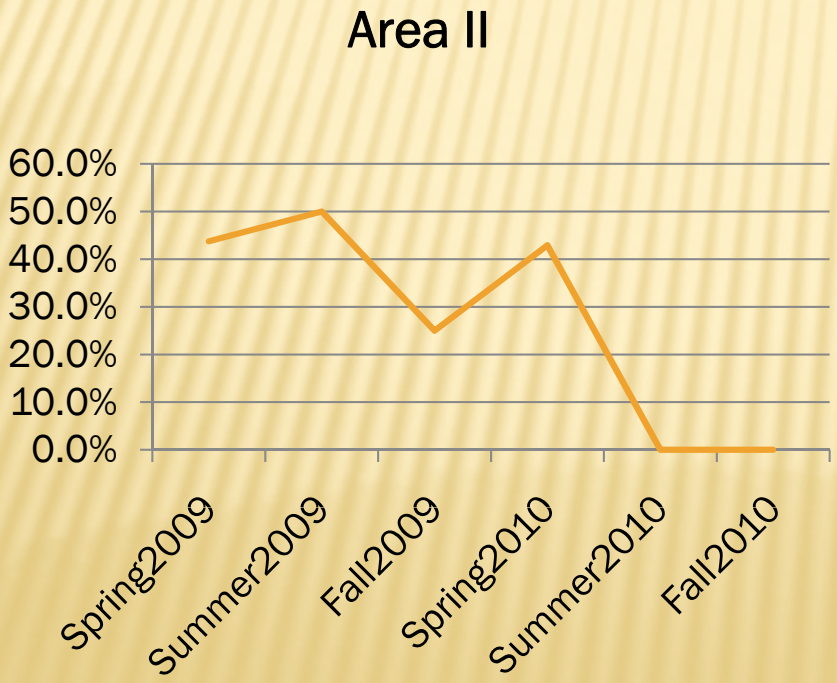
X=7

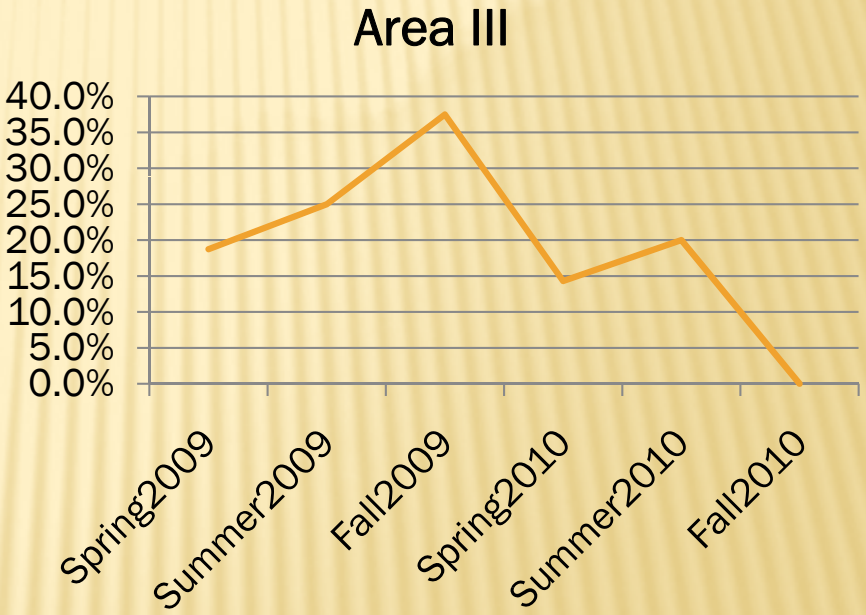
Y=72.1%

Next semester, in these three areas, the probability of crime accounts for the whole prediction area is 72.1%

AREA	Spring 2009	Summer 2009	Fall 2009	Spring 2010	Summer 2010	Fall 2010
I	25%	0%	12.5%	14.3%	60%	75%
II	43.75%	50%	25%	42.9%	0%	0%
III	18.75%	25%	37.5%	14.3%	20%	0%


Area I
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Spring2009
Summer2009
Fall2009
Spring2010
Summer2010
Fall2010


Area II
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Spring2009
Summer2009
Fall2009
Spring2010
Summer2010
Fall2010


Area III
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
Spring2009
Summer2009
Fall2009
Spring2010
Summer2010
Fall2010

AREA I

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.778489
R Square	0.606045
Adjusted R Square	0.507557
Standard Error	0.208053
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.266359	0.266359	6.153451	0.068166
Residual	4	0.173144	0.043286		
Total	5	0.439503			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.12047	0.193687	-0.62197	0.567661	-0.65823	0.417294	-0.65823	0.417294
X Variable 1	0.123371	0.049734	2.480615	0.068166	-0.01471	0.261456	-0.01471	0.261456

- Y= -0.12047+0.123371X

  X=7

  Y=74.31%

AREA II

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.834637
R Square	0.696618
Adjusted R Square	0.620773
Standard Error	0.138369
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.175851	0.175851	9.184719	0.038757
Residual	4	0.076584	0.019146		
Total	5	0.252435			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.620267	0.128815	4.815184	0.008553	0.26262	0.977914	0.26262	0.977914
X Variable 1	-0.10024	0.033077	-3.03063	0.038757	-0.19208	-0.00841	-0.19208	-0.00841

- Y= 0.620267-0.10024X

  X=7

  Y= - 8.14% ←

AREA III

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.571324
R Square	0.326411
Adjusted R Square	0.158014
Standard Error	0.113278
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.024873	0.024873	1.938338	0.236257
Residual	4	0.051328	0.012832		
Total	5	0.0762			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.324533	0.105456	3.077426	0.037024	0.03174	0.617326	0.03174	0.617326
X Variable 1	-0.0377	0.027079	-1.39224	0.236257	-0.11288	0.037482	-0.11288	0.037482

- Y= 0.324533-0.0377X

  X=7

  Y=6.06%

CONCLUSION

- To improve the accuracy of the prediction
    1. Data (the more, the better)
    2. Adding more factors into the model

- Population density and degree of urbanization.
- Variations in composition of the population, particularly youth concentration.
- Modes of transportation and highway system.
- Economic conditions, including median income, poverty level, and job availability.
- Cultural factors and educational, recreational, and religious characteristics.
- Family conditions with respect to divorce and family cohesiveness.
- Climate.
- Policies of other components of the criminal justice system
- Citizens' attitudes toward crime.

Questions?

Thank you.