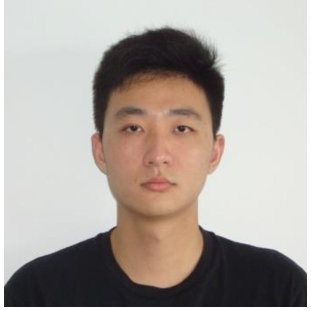
Crime Analysis

of Chicago in 2015

Team member

Lan Chang



Jiwei Wei

Zhizhou Chen

Overview

City of Chicago Data Portal

```
262893 obs. of 22 variables
```

ID, Case.Number, **Date**, Block, IUCR, **Primary.Type**, **Description**, **Location.Description**, **Arrest**, Domestic, Beat, District, Ward, Community.Area, **FBI.Code**, X.Coordinate, Y.Coordinate, Year, Updated.On, **Latitude**, **Longitude**, Location

```
259807 obs. of 22 variables
```

```
01/18/2015 04:00:00 PM
```

```
date_month date_day date_year time_hour time_minute time_second
         1       18      2015        16           0           0
```

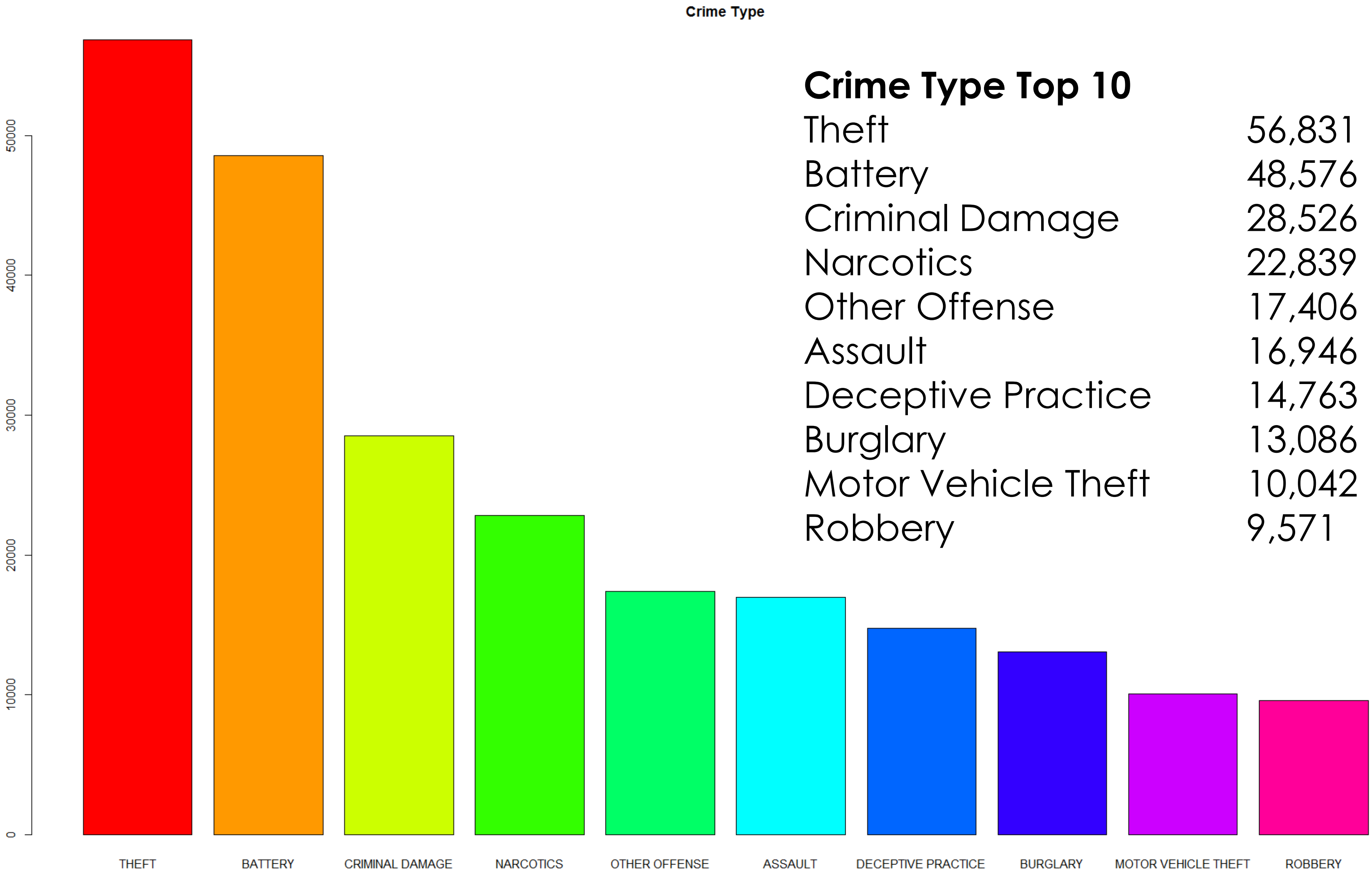
As citizens or tourists, what should they care for?

- Which kind of crime has the most frequency to happen?
- Theft or Robbery…?

- Which area of Chicago has the most crime rate?
- Downtown or countryside?

- What time has the most crime rate?
- Day or night?

Crime Type



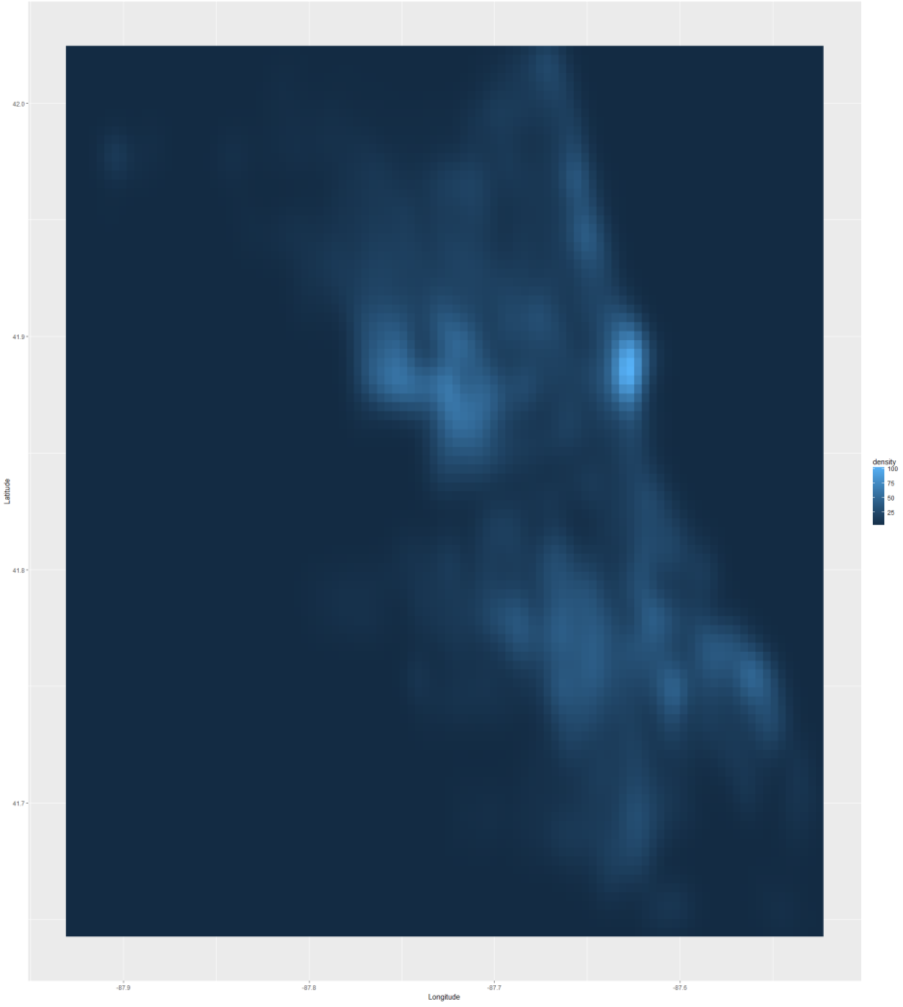
Crime Type Top 10

Theft	56,831
Battery	48,576
Criminal Damage	28,526
Narcotics	22,839
Other Offense	17,406
Assault	16,946
Deceptive Practice	14,763
Burglary	13,086
Motor Vehicle Theft	10,042
Robbery	9,571

Crime

Point Map

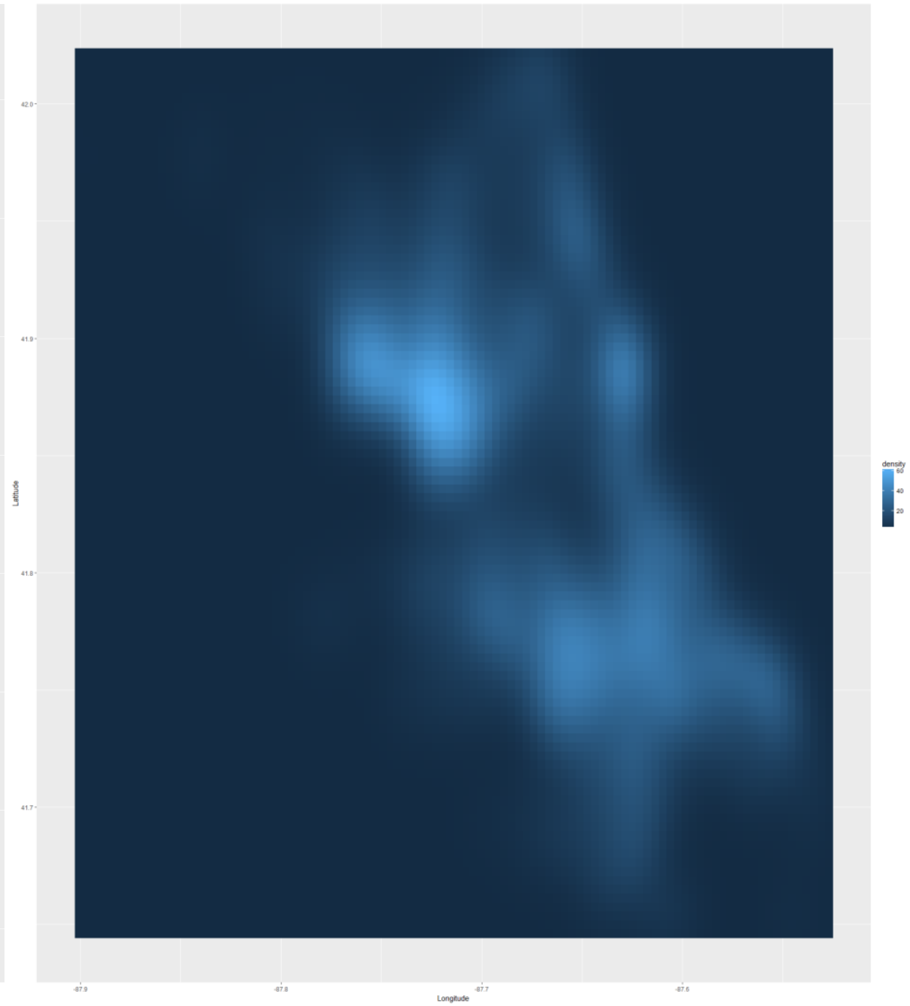
Raster Map



Robbery

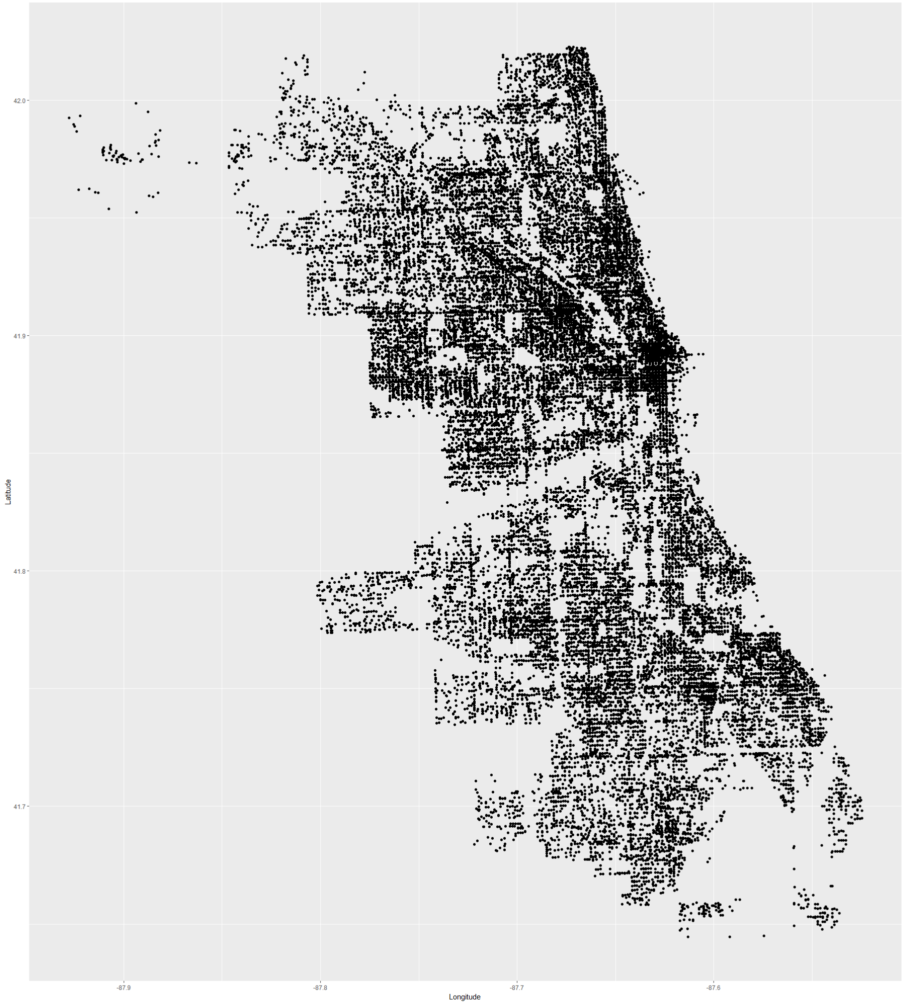
Point Map

Raster Map

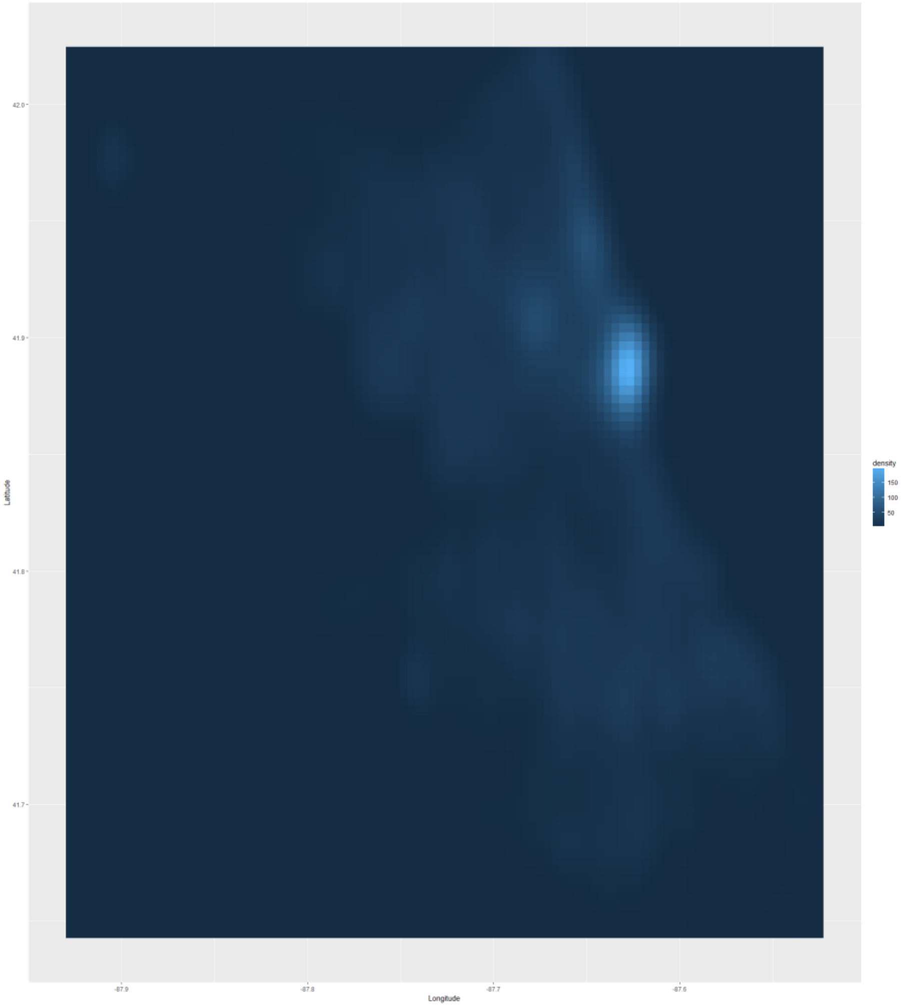


Theft

Point Map



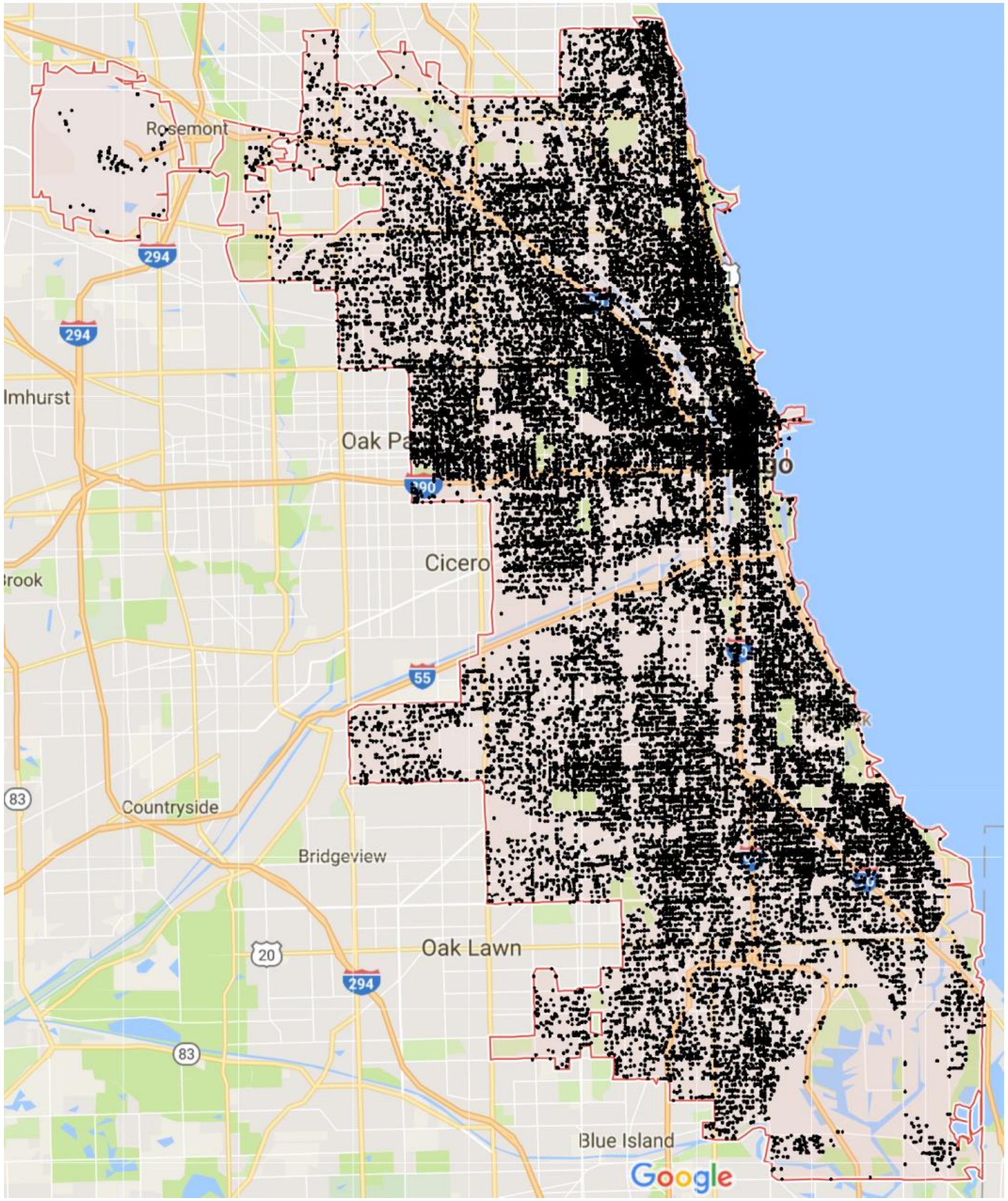
Raster Map



Theft

We want to pay more attention on **theft**

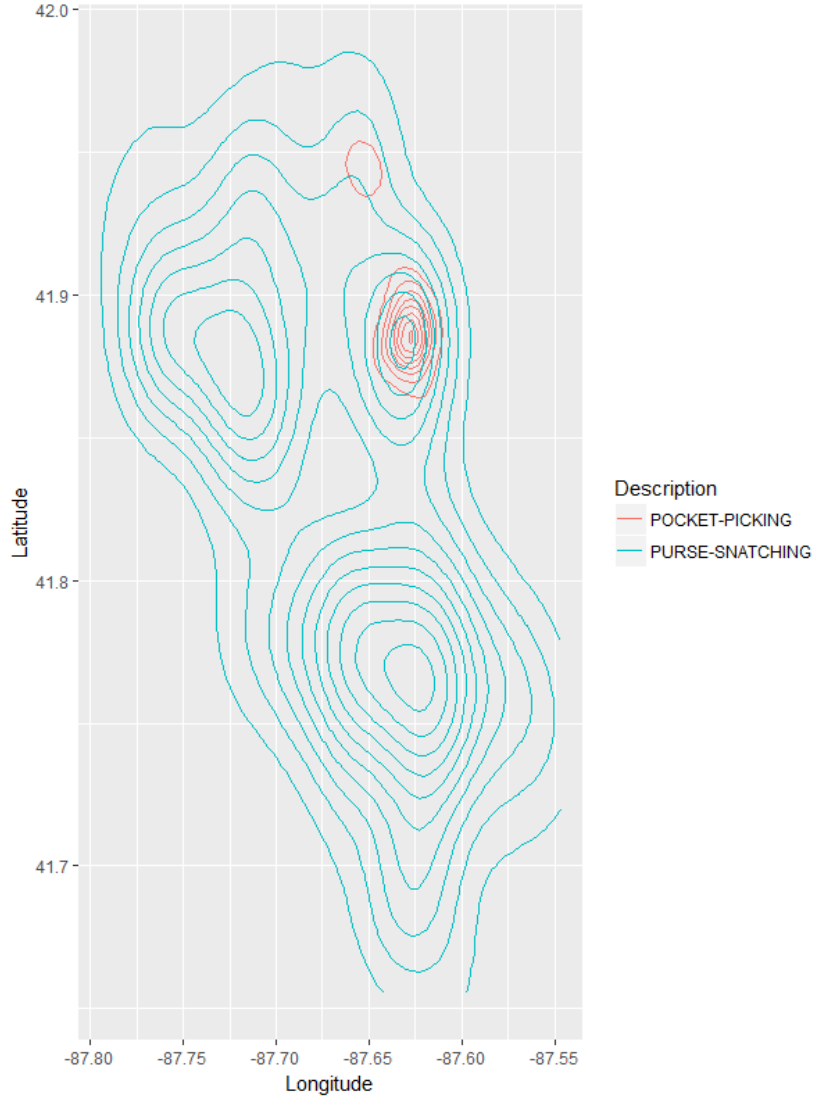
Theft Point on Google Map



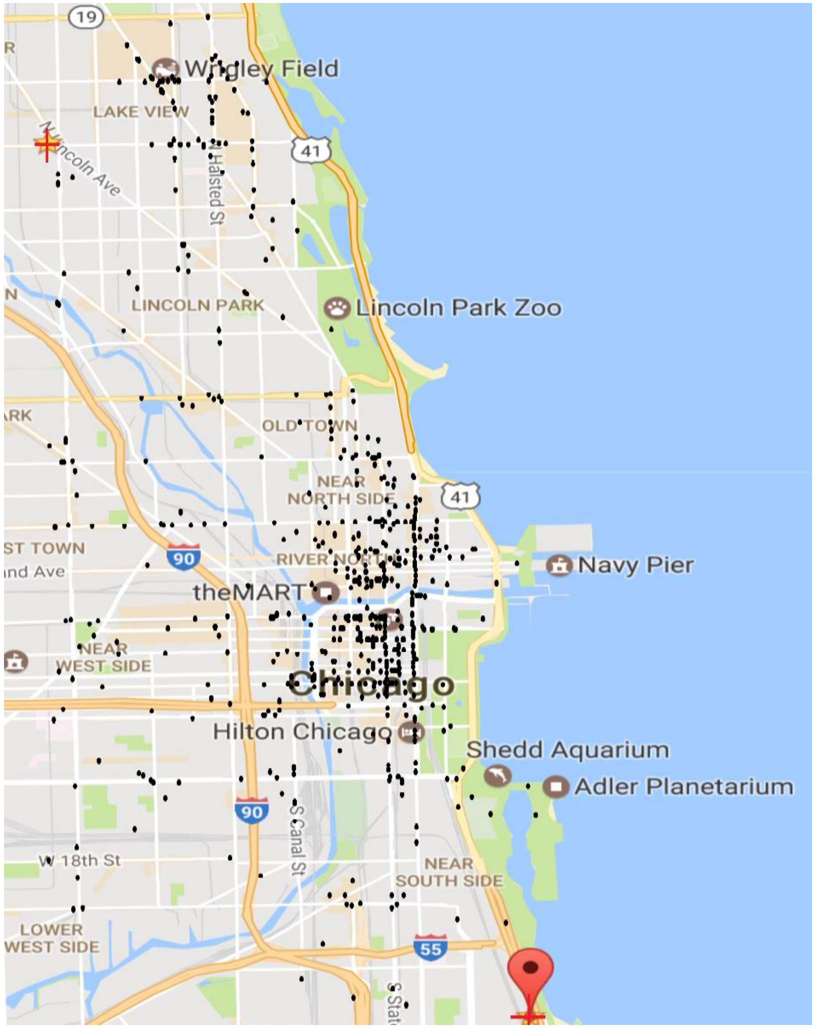
More:
Downtown Area

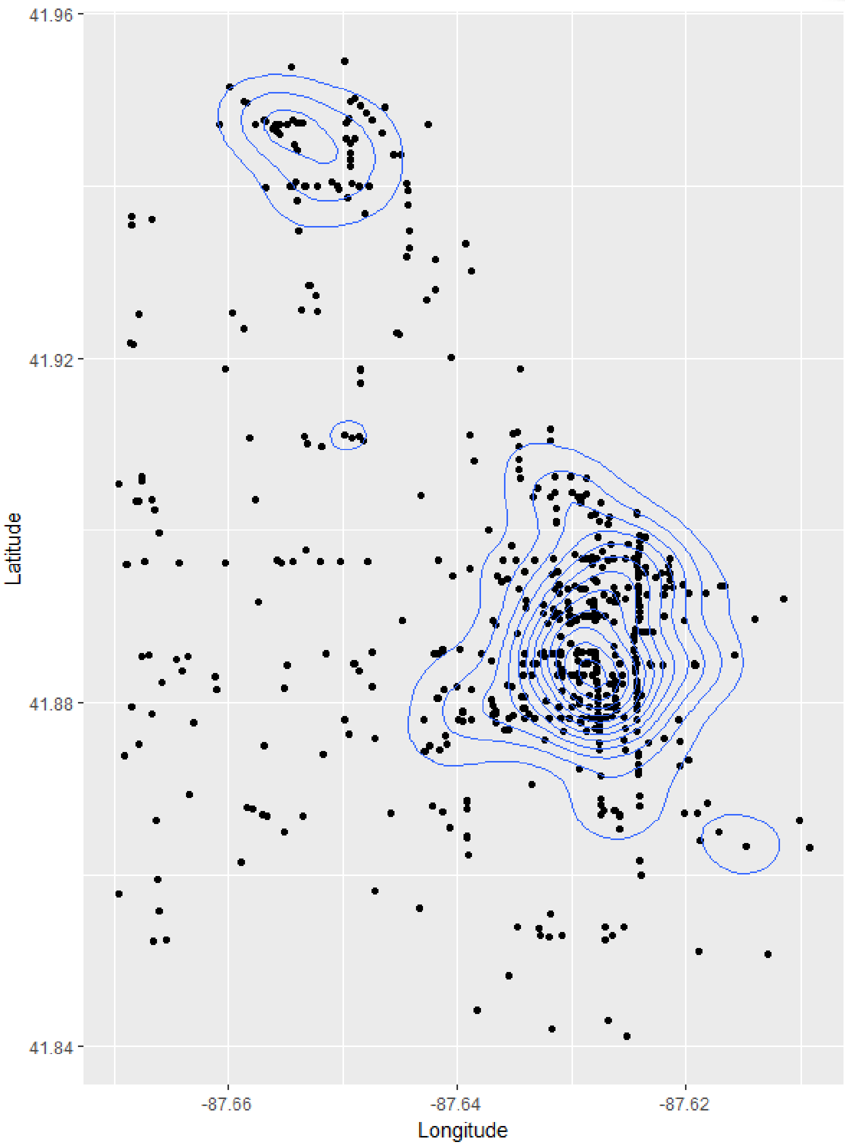
Less:
Highway,
Park,
River,
Lake,
Airport

Theft: Pocket-Picking & Purse-Snatching

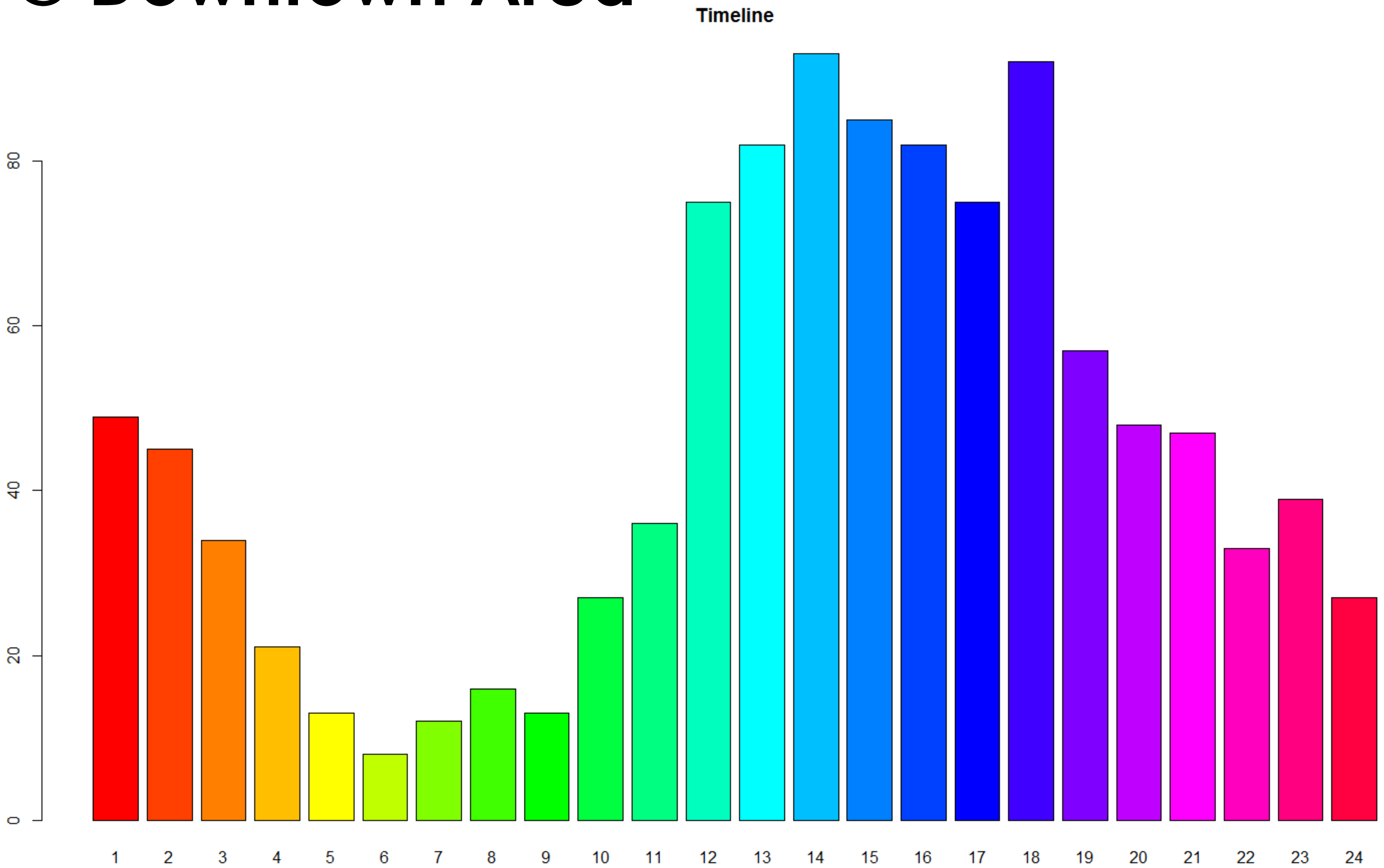


Theft: Pocket-Picking & Purse-Snatching @ Downtown Area

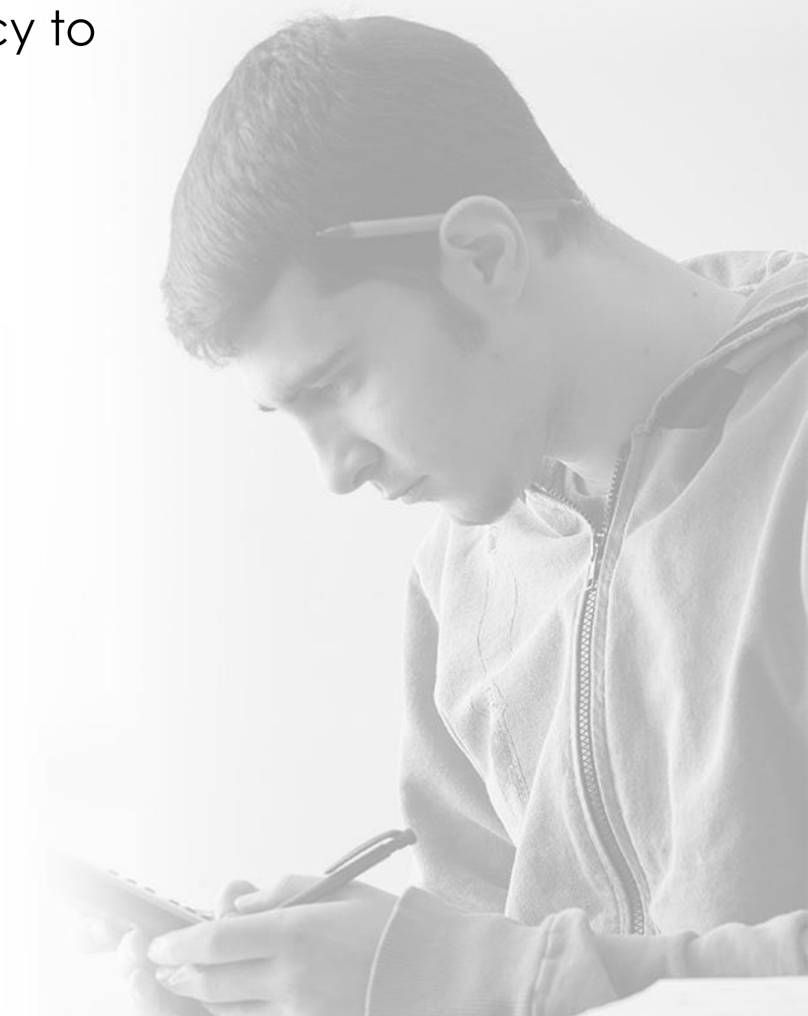




Theft: Pocket-Picking & Purse-Snatching @ Downtown Area



As police department, what should they care for?

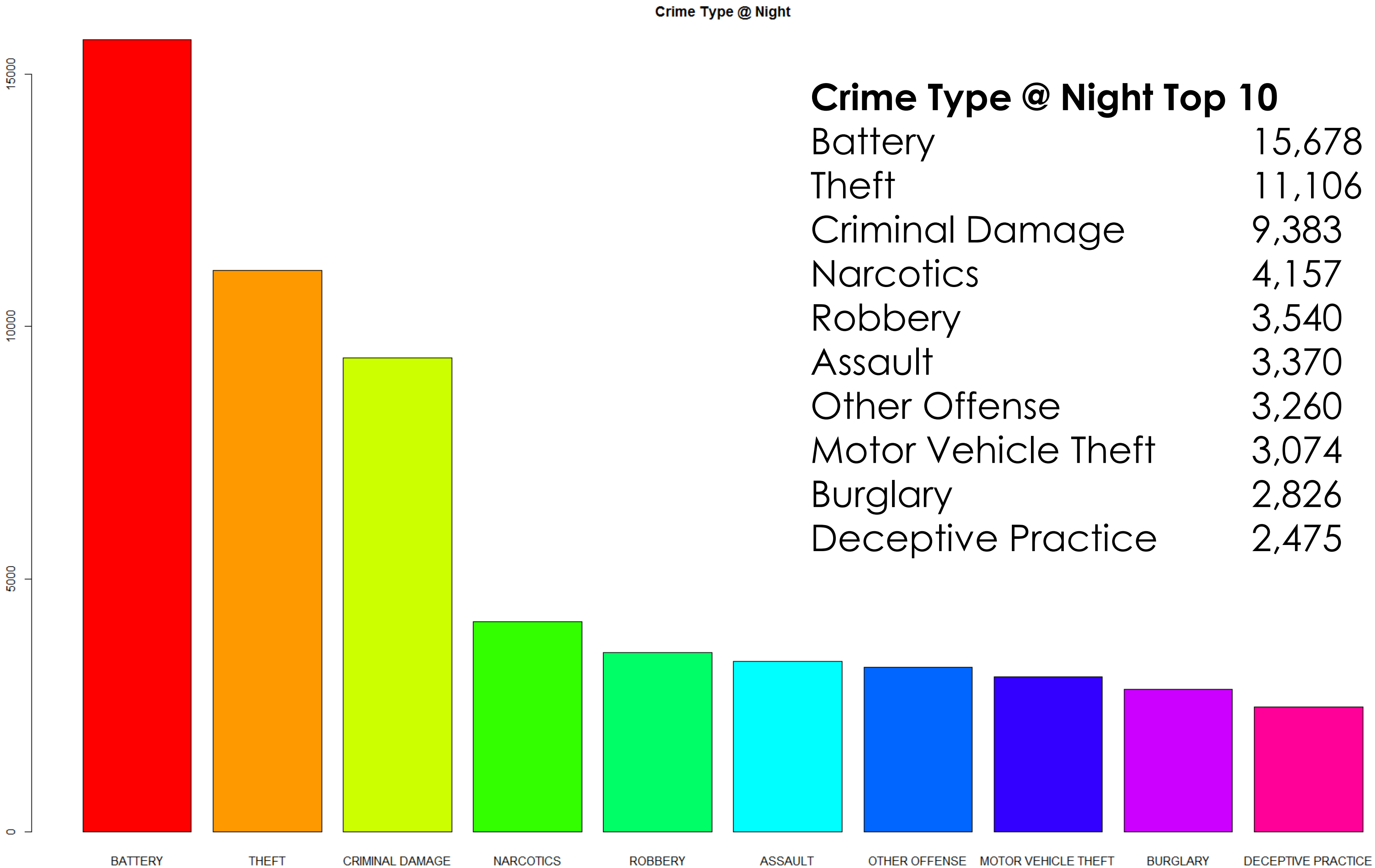


- Which kind of crime has the most frequency to happen at night?
- Battery…?

- Has gun?
- Yes or no?

- How should they patrol?
- Which area and what is the center point?

Crime Type @ Night (10pm-6am)

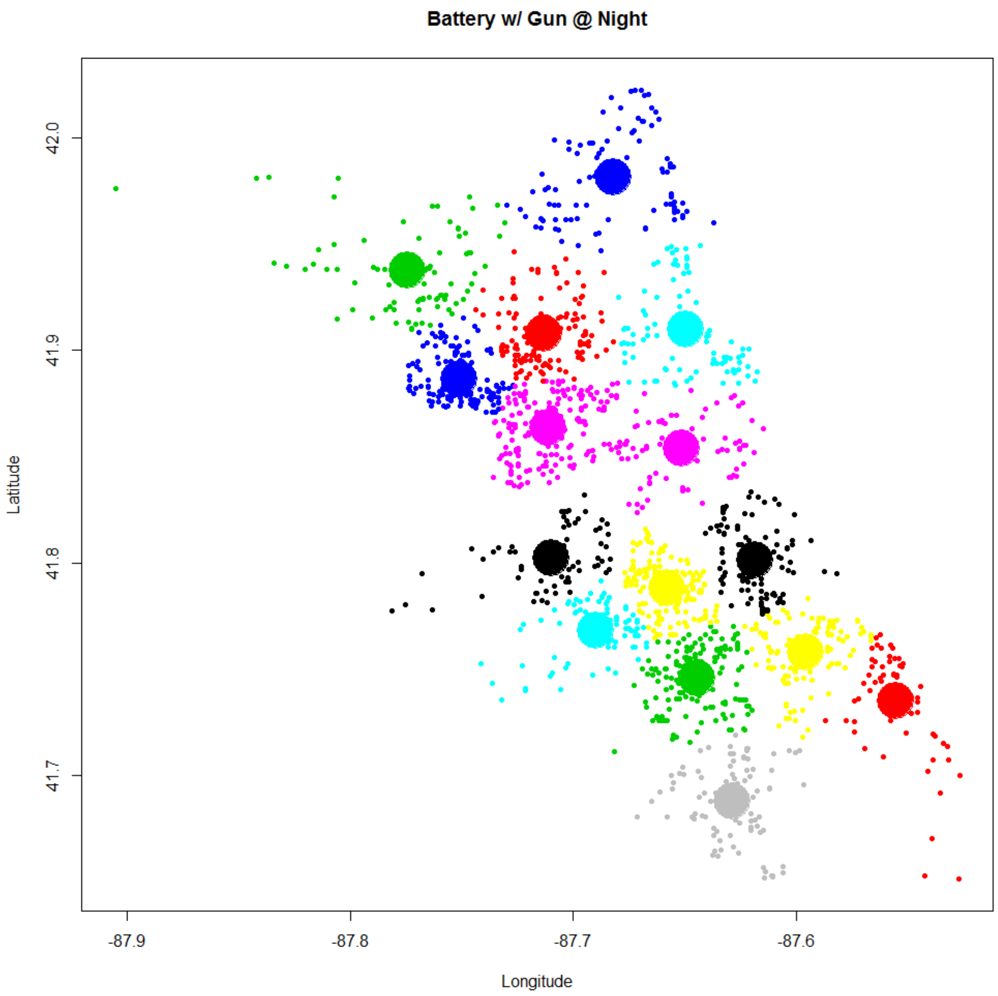


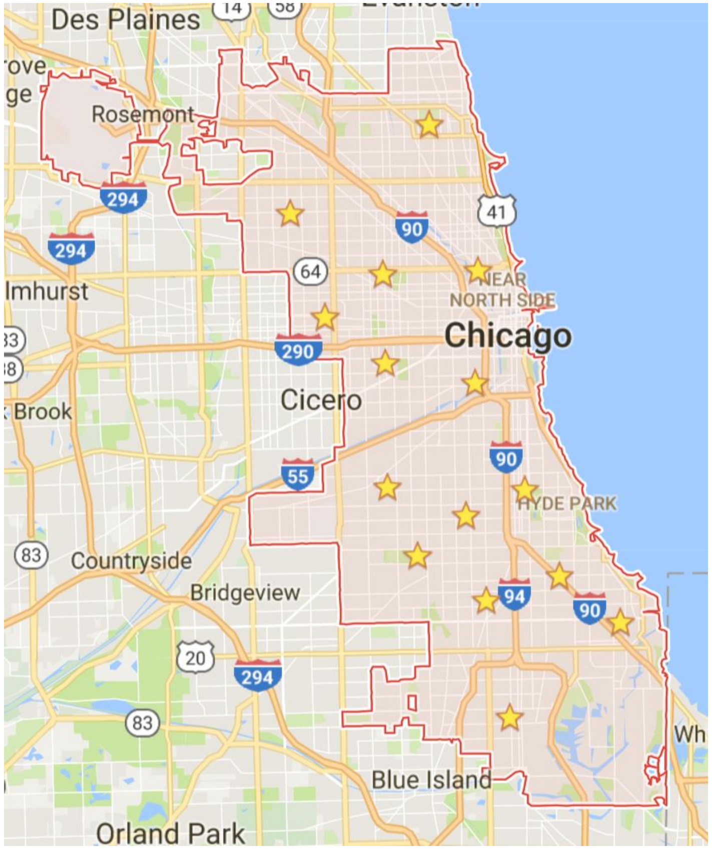
Crime Type @ Night Top 10

Battery	15,678
Theft	11,106
Criminal Damage	9,383
Narcotics	4,157
Robbery	3,540
Assault	3,370
Other Offense	3,260
Motor Vehicle Theft	3,074
Burglary	2,826
Deceptive Practice	2,475

Battery w/ Gun @ Night

Use k-means to find the center points with Latitude and Longitude (k=15) .





Information for police from this data

Can we predict the crimes type?

- If we can predict the crime type before the police get to the scene of crime, the officer can be prepared which can reduce the danger.
- We have Latitude, Longitude and date, looks like we can use *k*-Nearest Neighbor Algorithm to predict the crimes types.

Use k-nearest neighbor to predict the crime type with Latitude and Longitude (k=50) .

						Actual							
Predict	ARSON	ASSAULT	BATTERY	CRIM SEXUAL ASSAULT	HOMICIDE	INTERFERENCE WITH PUBLIC OFFICER	KIDNAPPING	OBSCENITY	OFFENSE INVOLVING CHILDREN	PROSTITUTION	PUBLIC INDECENCY	THEFT	WEAPONS VIOLATION
ARSON	0	0	0	0	0	0	0	0	0	0	0	0	0
ASSAULT	0	5	14	0	0	1	0	0	0	0	0	14	1
BATTERY	54	2226	6569	141	71	198	28	2	286	69	2	3820	507
CRIM SEXUAL ASSAULT	0	0	0	0	0	0	0	0	0	0	0	0	0
HOMICIDE	0	0	0	0	0	0	0	0	0	0	0	0	0
INTERFERENCE WITH PUBLIC OFFICER	0	0	0	0	0	0	0	0	0	0	0	0	0
KIDNAPPING	0	0	0	0	0	0	0	0	0	0	0	0	0
OBSCENITY	0	0	0	0	0	0	0	0	0	0	0	0	0
OFFENSE INVOLVING CHILDREN	0	0	0	0	0	0	0	0	0	0	0	0	0
PROSTITUTION	1	12	46	2	0	2	2	0	2	159	0	30	2
PUBLIC INDECENCY	0	0	0	0	0	0	0	0	0	0	0	0	0
THEFT	21	1150	3042	97	23	84	14	3	133	43	1	7554	156
WEAPONS VIOLATION	0	0	0	0	0	0	0	0	0	0	0	0	0

Because the amount of **Battery** and **Theft** type are much more than other crime types, no matter where the crimes happen, it will be highly likely predicted as **Battery** or **Theft**.

Use k-nearest neighbor to predict the crime type with Latitude, Longitude and date (k=50).

						Actual							
Predict	ARSON	ASSAULT	BATTERY	CRIM SEXUAL ASSAULT	HOMICIDE	INTERFERENCE WITH PUBLIC OFFICER	KIDNAPPING	OBSCENITY	OFFENSE INVOLVING CHILDREN	PROSTITUTION	PUBLIC INDECENCY	THEFT	WEAPONS VIOLATION
ARSON	0	0	0	0	0	0	0	0	0	0	0	0	0
ASSAULT	0	10	14	1	0	0	0	0	0	1	0	12	0
BATTERY	65	2029	6173	154	68	216	29	1	241	162	1	3792	513
CRIM SEXUAL ASSAULT	0	0	0	0	0	0	0	0	0	0	0	0	0
HOMICIDE	0	0	0	0	0	0	0	0	0	0	0	0	0
INTERFERENCE WITH PUBLIC OFFICER	0	0	0	0	0	0	0	0	0	0	0	0	0
KIDNAPPING	0	0	0	0	0	0	0	0	0	0	0	0	0
OBSCENITY	0	0	0	0	0	0	0	0	0	0	0	0	0
OFFENSE INVOLVING CHILDREN	0	0	0	0	0	0	0	0	0	0	0	0	0
PROSTITUTION	0	10	23	1	2	3	0	0	0	66	0	22	3
PUBLIC INDECENCY	0	0	0	0	0	0	0	0	0	0	0	0	0
THEFT	11	1344	3461	84	24	66	15	4	180	42	2	7592	150
WEAPONS VIOLATION	0	0	0	0	0	0	0	0	0	0	0	0	0

This time I use normalized Latitude, Longitude and date to predict the crime type. But the **Battery** and **Theft** type still dominate the prediction, no matter where and when the crimes happen, it still will be highly likely predicted as **Battery** or **Theft**.

Use weighted k-nearest neighbor to predict the crime type with Latitude, Longitude and date (k=50).

						Actual							
Predict	ARSON	ASSAULT	BATTERY	CRIM SEXUAL ASSAULT	HOMICIDE	INTERFERENCE WITH PUBLIC OFFICER	KIDNAPPING	OBSCENITY	OFFENSE INVOLVING CHILDREN	PROSTITUTION	PUBLIC INDECENCY	THEFT	WEAPONS VIOLATION
ARSON	0	0	57	0	0	0	0	0	0	0	0	19	0
ASSAULT	0	90	1932	0	0	0	0	0	0	15	0	1355	1
BATTERY	0	193	6004	0	0	0	0	0	0	40	0	3430	4
CRIM SEXUAL ASSAULT	0	7	159	0	0	0	0	0	0	1	0	73	0
HOMICIDE	0	2	65	0	0	0	0	0	0	0	0	27	0
INTERFERENCE WITH PUBLIC OFFICER	0	7	202	0	0	0	0	0	0	3	0	72	1
KIDNAPPING	0	0	26	0	0	0	0	0	0	0	0	18	0
OBSCENITY	0	1	1	0	0	0	0	0	0	0	0	3	0
OFFENSE INVOLVING CHILDREN	0	9	244	0	0	0	0	0	7	0	0	161	0
PROSTITUTION	0	3	98	0	0	0	0	0	0	131	0	39	0
PUBLIC INDECENCY	0	0	1	0	0	0	0	0	0	0	0	2	0
THEFT	0	159	3672	0	0	0	0	0	0	31	0	7549	7
WEAPONS VIOLATION	0	17	476	0	0	0	0	0	0	3	0	169	1

The weighted k-nearest neighbor Algorithm still can’t make any improvement. The **Battery** and **Theft** type still dominate the prediction.

What if we remove Battery and Theft type.

- Since Battery and theft are not serious crimes, what if we remove these two type and choose some serious crimes with close amount?
- I chose **ASSAULT**, **BURGLARY**, **DECEPTIVE PRACTICE**, **MOTOR VEHICLE THEFT**, **OTHER OFFENSE** and **ROBBERY**. These six types have similar amount.

Use weighted k-nearest neighbor to predict the crime type among six different types (remove Battery and Theft) with Latitude, Longitude and date(k=50).

	Actual					
Predict	ASSAULT	BURGLARY	DECEPTIVE PRACTICE	MOTOR VEHICLE THEFT	OTHER OFFENSE	ROBBERY
ASSAULT	1112	317	586	186	954	173
BURGLARY	675	536	521	197	590	125
DECEPTIVE PRACTICE	485	238	1438	130	544	88
MOTOR VEHICLE THEFT	515	236	309	290	462	151
OTHER OFFENSE	978	343	757	194	1122	168
ROBBERY	568	226	264	166	437	282

This one is much better than the former results.

Measure the performance and decide *K* value

Use average correct rate to measure the performance.

$$\text{Correct rate} = \frac{n_{correct\ prediction\ in\ one\ type}}{n_{all\ prediction\ in\ one\ type}}$$

	Actual					
Predict	ASSAULT	BURGLARY	DECEPTIVE PRACTICE	MOTOR VEHICLE THEFT	OTHER OFFENSE	ROBBERY
ASSAULT	1112	317	586	186	954	173
BURGLARY	675	536	521	197	590	125
DECEPTIVE PRACTICE	485	238	1438	130	544	88
MOTOR VEHICLE THEFT	515	236	309	290	462	151
OTHER OFFENSE	978	343	757	194	1122	168
ROBBERY	568	226	264	166	437	282

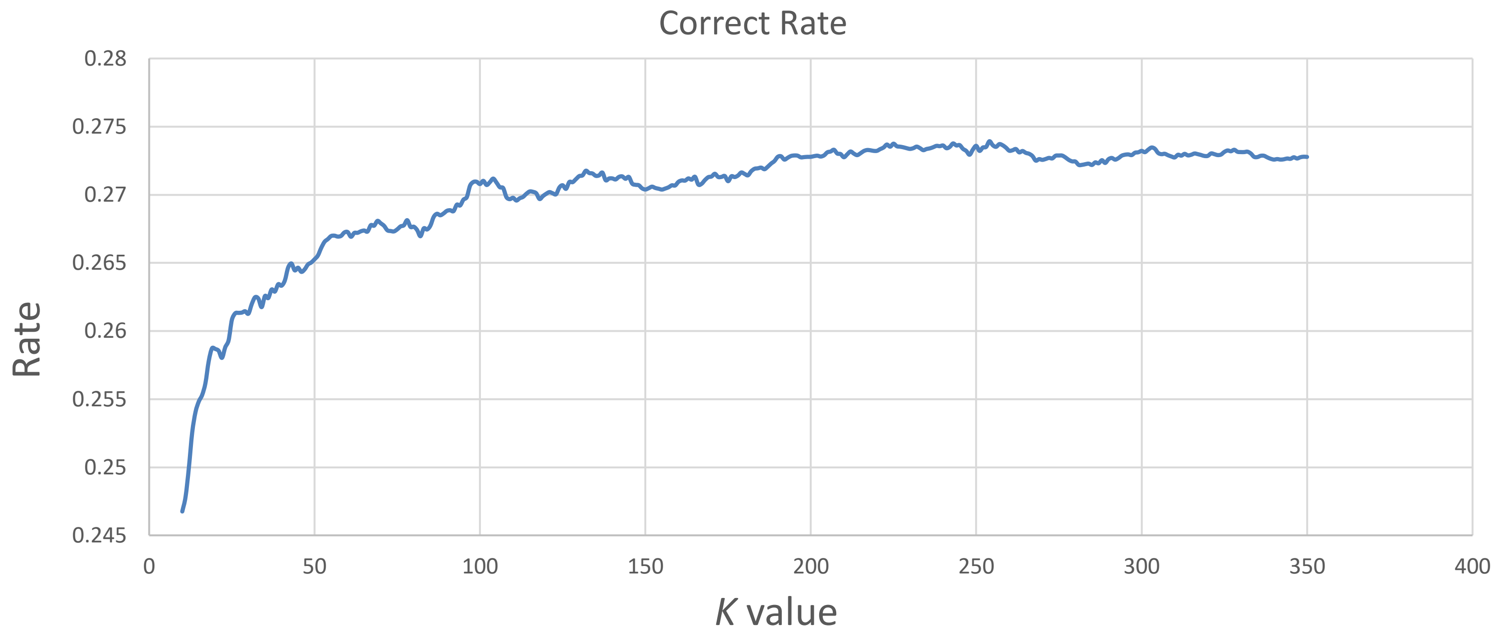
All predictions in **Deceptive Parctice** type

Correct *predictions* in **Deceptive Partice** type

Average correct rate=average(correct rate)

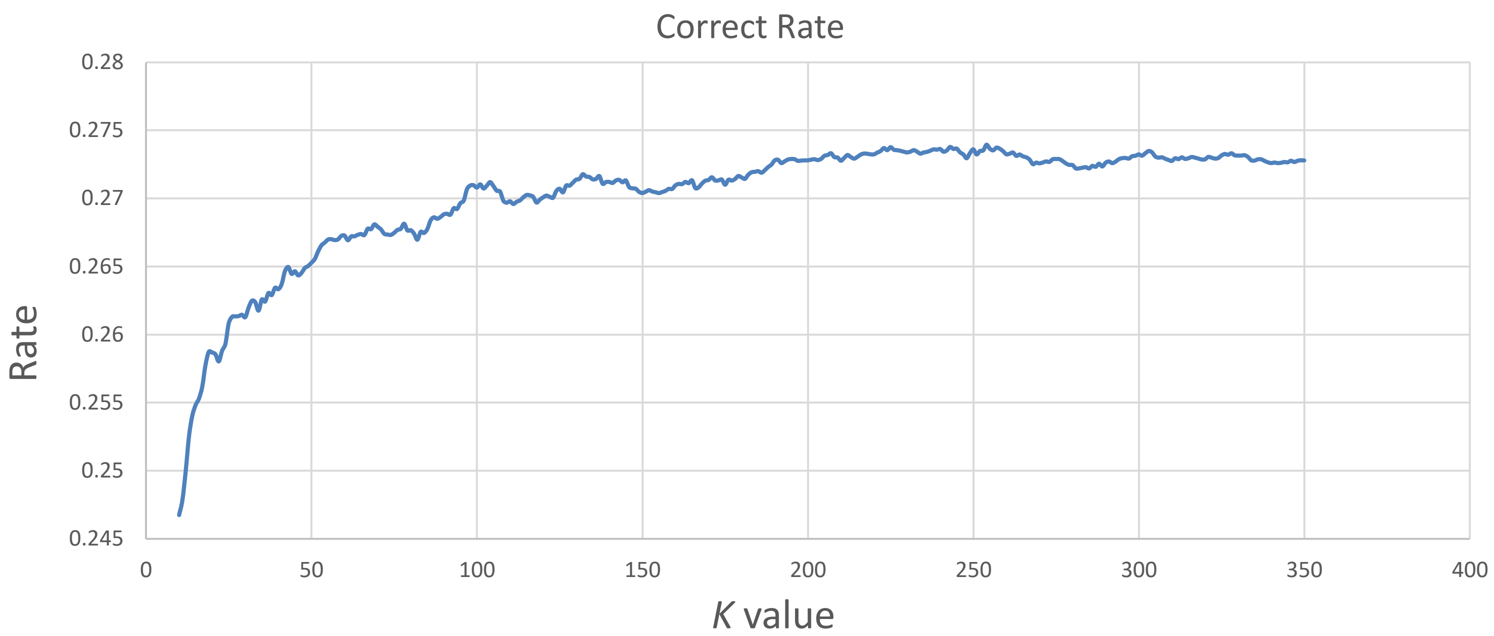
Measure the performance and decide *K* value

Decide the *K* value based on the performance.



Calculate the correct rate when *K* vary from 10 to 350.

Good model or bad model

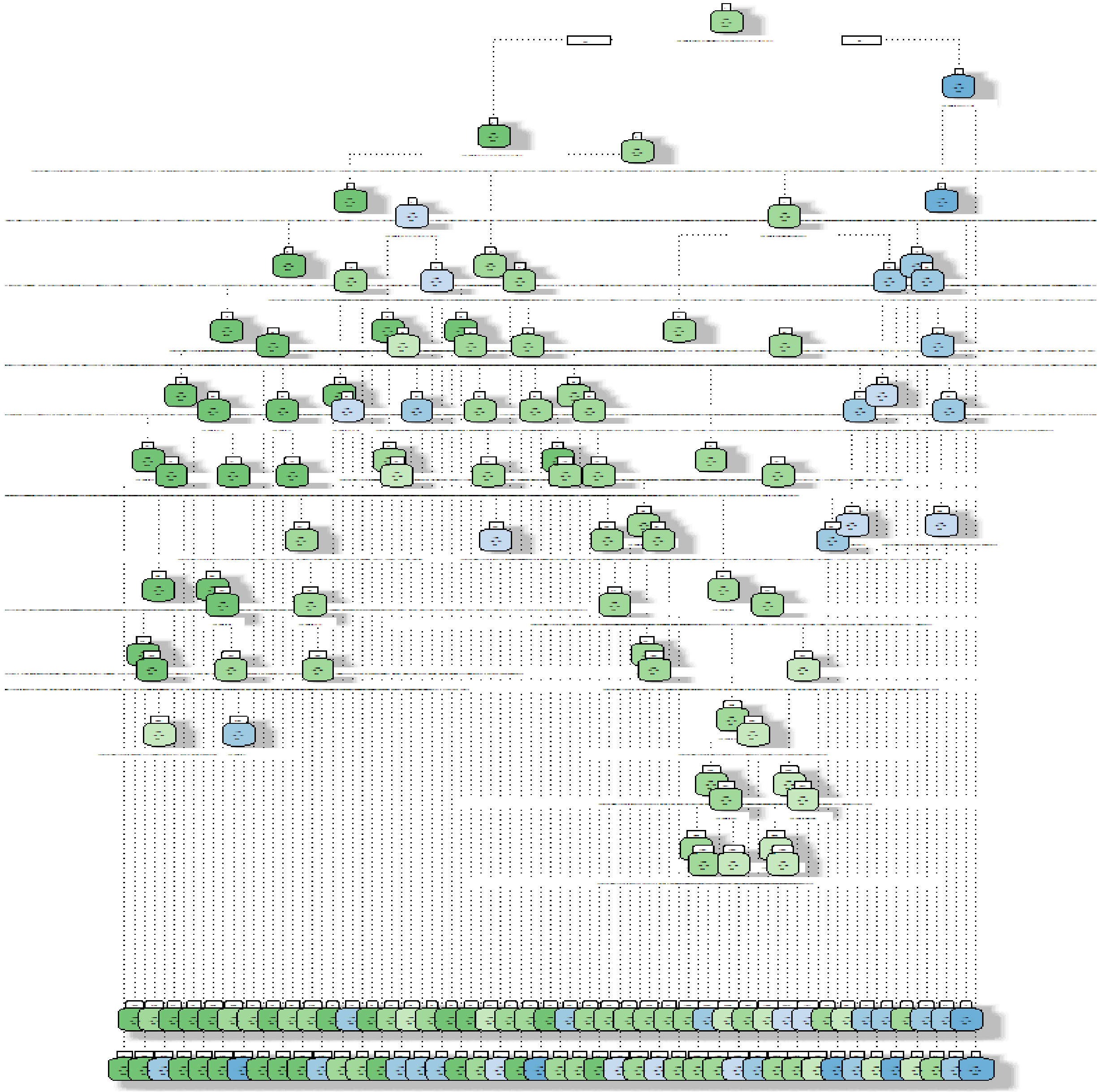


The max correct rate is around 27.5%. Which means this model only has about one fourth chance to get a correct prediction.

Unfortunately, this model is not a good model and may be *K*-nn is not useful on this dataset.

If a new case is reported, can police catch the suspect?

- The data contains primary type and location of the crimes, when a new case is reported, the police can use this model to know the arrested rate of the specified crime type. If the chance of catching the suspect is high, the police can use more force to catch the suspect in the short time.



Filtered Data

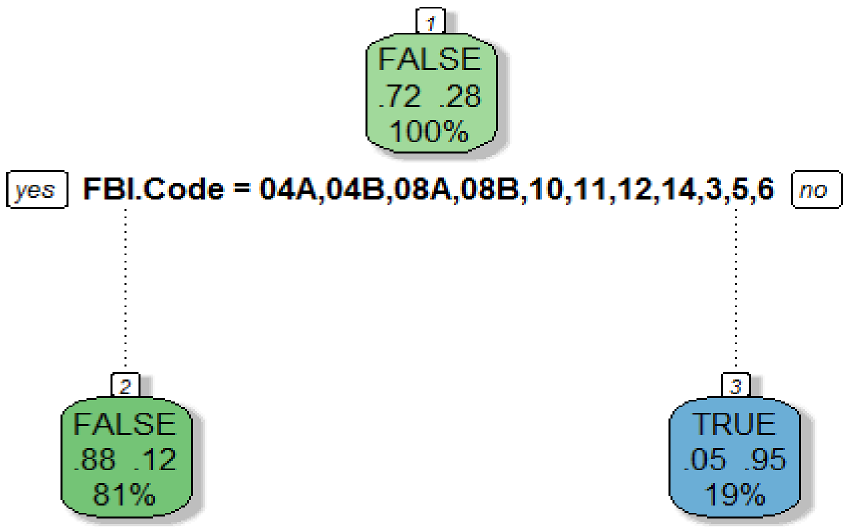
118,626 records

```
> table(dataset$FBI.Code)

  04A   04B   08A   08B    10    11    12    13    14    15    16    18    26     3     5     6 
 2867  4669  6073 23962   414  3054     4    24 16757  2104  1161 17250  2151  6649  4855 26632 
> table(dataset$Primary.Type)

           ASSAULT            BATTERY           BURGLARY    CRIMINAL DAMAGE  CRIMINAL TRESPASS DECEPTIVE PRACTICE 
              8940              28631               4855              16757               1816               3496 
         NARCOTICS       PROSTITUTION            ROBBERY              THEFT  WEAPONS VIOLATION 
             17585               1161               6649              26632               2104 
> table(dataset$Location.Description)

                     APARTMENT PARKING LOT/GARAGE(NON.RESID.)                     RESTAURANT 
                         30482                           6476                           4755 
                      SIDEWALK                         STREET         VEHICLE NON-COMMERCIAL 
                         25298                          47606                           4009 
```

Crime Categories (NIBRS Code)	Uniform Crime Reporting (UCR) Code and Description
All Crime	Includes all Crime Categories
Index Crime Definition: More serious offenses.	Includes the following Crime Categories (see below) Homicide 1st & 2nd Degree (01A) (Index) Criminal Sexual Assault (02) (Index) Robbery (03) (Index) Aggravated Assault (04A) (Index) Aggravated Battery (04B) (Index) Burglary (05) (Index) Larceny (06) (Index) Motor Vehicle Theft (07) (Index) Arson (09) (Index)
Non-Index Crime Definition: Less serious offenses.	Includes the following Crime Categories (see below) Involuntary Manslaughter (01B) Simple Assault (08A) Simple Battery (08B) Forgery & Counterfeiting (10) Fraud (11) Embezzlement (12) Stolen Property (13) Vandalism (14) Weapons Violation (15) Prostitution (16) Criminal Sexual Abuse (17) Drug Abuse (18) Gambling (19) Offenses Against Family (20) Liquor License (22) Disorderly Conduct (24) Misc Non-Index Offense (26)
Violent Crime Definition: Crime related to Violence	Includes the following Crime Categories (see below) Homicide 1st & 2nd Degree (01A) (Index) Criminal Sexual Assault (02) (Index) Robbery (03) (Index) Aggravated Assault (04A) (Index) Aggravated Battery (04B) (Index)
Property Crime Definition: Crime related to Property	Includes the following Crime Categories (see below) Burglary (05) (Index) Larceny (06) (Index) Motor Vehicle Theft (07) (Index) Arson (09) (Index)

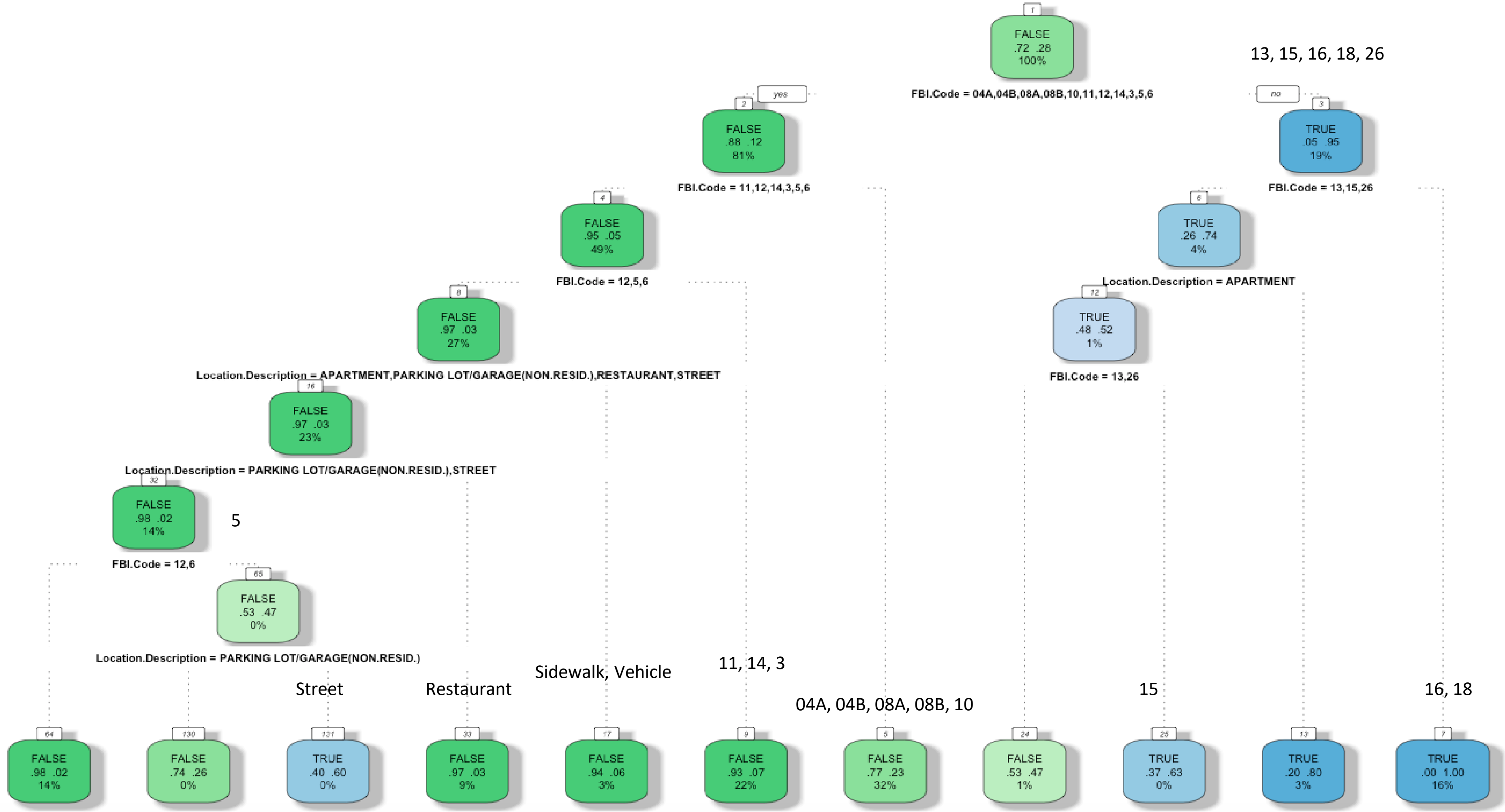
Rpart without using parameter

mytree <- rpart(Arrest~Location.Description+FBI.Code, data = dataset, method = "class")



```
mytree <- rpart(Arrest~Location.Description+FBI.Code, data = dataset, method = "class", control = rpart.control(minbucket = 9, cp=0.00001))
```

- minbucket: the minimum number of observations in any terminal <leaf> node.
- cp : complexity parameter.



FALSE .72 .28 100%
FBI.Code = 04A,04B,08A,08B,10,11,12,14,3,5,6
yes
no
13, 15, 16, 18, 26
FALSE .88 .12 81%
FBI.Code = 11,12,14,3,5,6
TRUE .05 .95 19%
FBI.Code = 13,15,26
FALSE .95 .05 49%
FBI.Code = 12,5,6
TRUE .26 .74 4%
Location.Description = APARTMENT
FALSE .97 .03 27%
Location.Description = APARTMENT,PARKING LOT/GARAGE(NON.RESID.),RESTAURANT,STREET
TRUE .48 .52 1%
FBI.Code = 13,26
FALSE .97 .03 23%
Location.Description = PARKING LOT/GARAGE(NON.RESID.),STREET
FALSE .98 .02 14%
FBI.Code = 12,6
5
FALSE .53 .47 0%
Location.Description = PARKING LOT/GARAGE(NON.RESID.)
Street
Restaurant
Sidewalk, Vehicle
11, 14, 3
04A, 04B, 08A, 08B, 10
15
16, 18
FALSE .98 .02 14%
FALSE .74 .26 0%
TRUE .40 .60 0%
FALSE .97 .03 9%
FALSE .94 .06 3%
FALSE .93 .07 22%
FALSE .77 .23 32%
FALSE .53 .47 1%
TRUE .37 .63 0%
TRUE .20 .80 3%
TRUE .00 1.00 16%

Statistical Results

- Arrest rate is low, only 19% are caught in most common type.
- Only 2% of Theft & Embezzlement are caught.
- The crimes happened in restaurant mostly can’t catch the criminal suspect.
- Nearly all suspects of the Narcotics & Prostitution are caught.

Summary

Thank you!