CHICAGO CRIME DATA

KUBIG
박소현 김효익 조송현 조규선 이영신

INDEX

Data description

Data description

‹Chicago crime data›

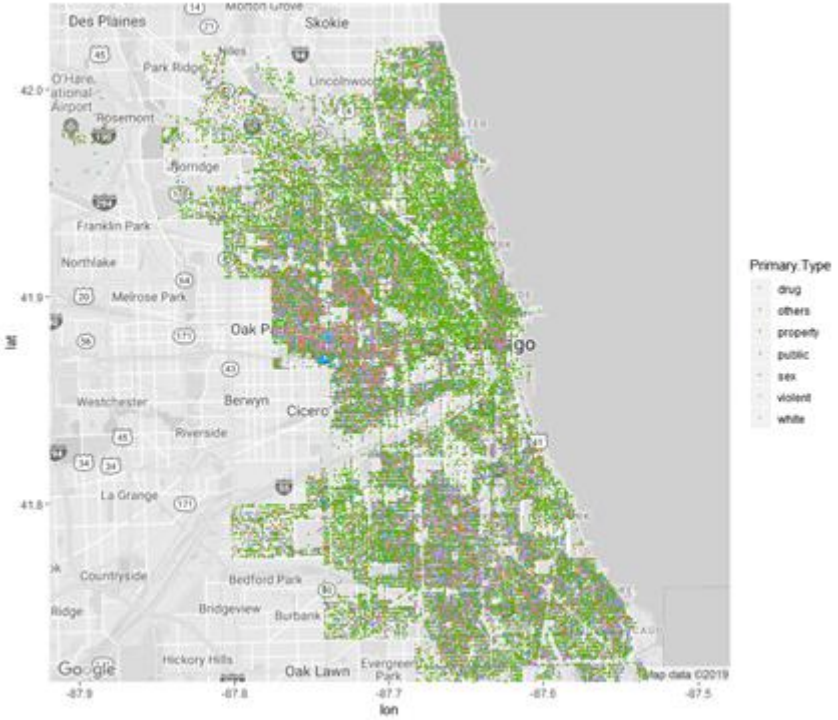
- Incidents of crime that occurred in the City of Chicago from 2012 to 2017

- Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system and includes unverified reports.

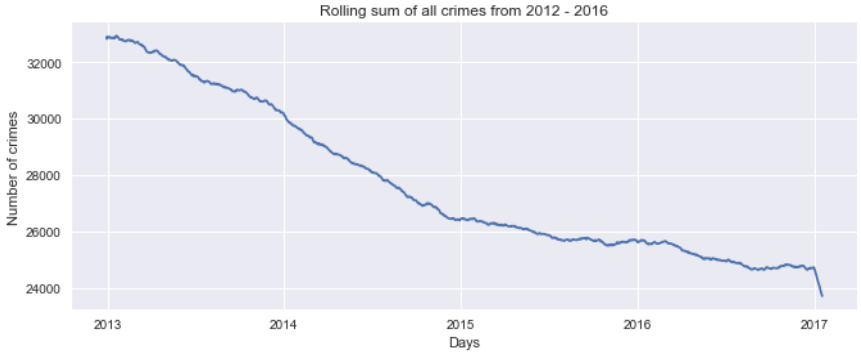
- https://www.kaggle.com/currie32/crimes-in-chicago

Visualization(EDA)

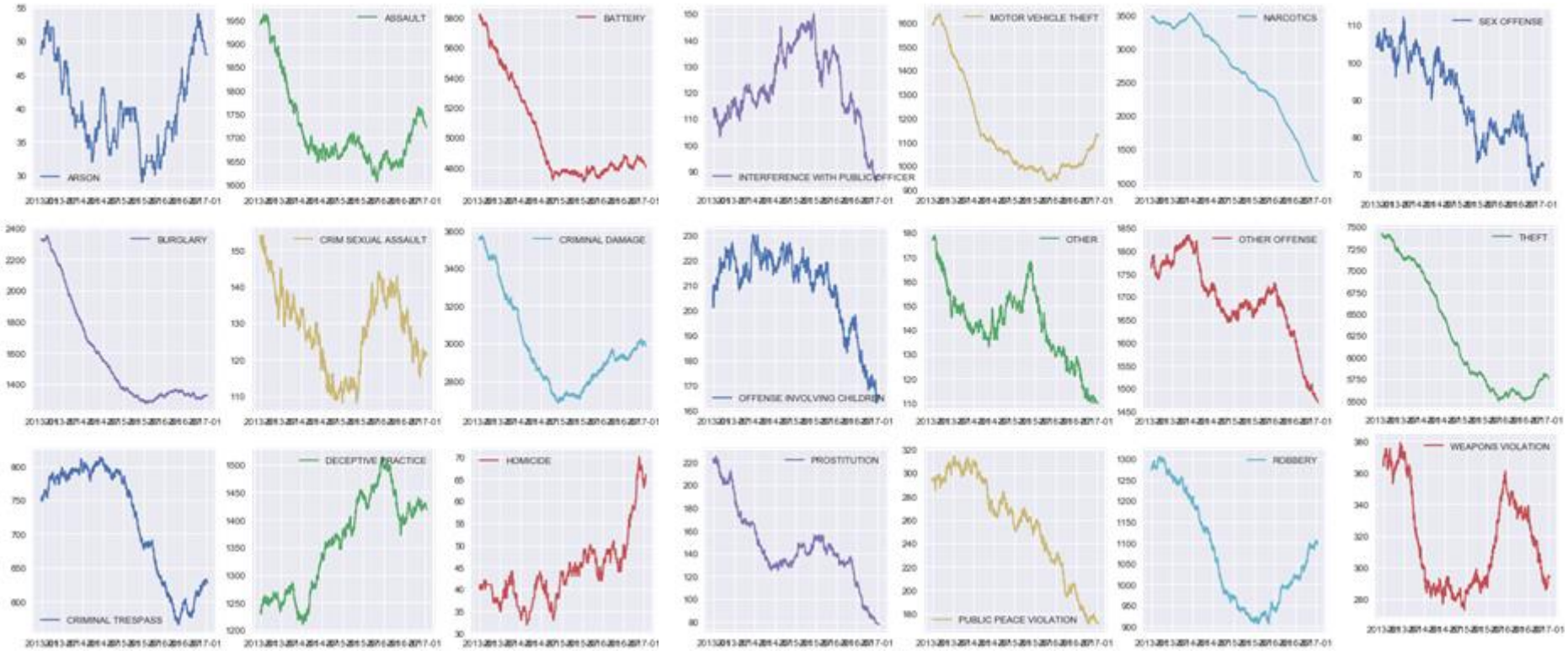
Visualization(EDA)



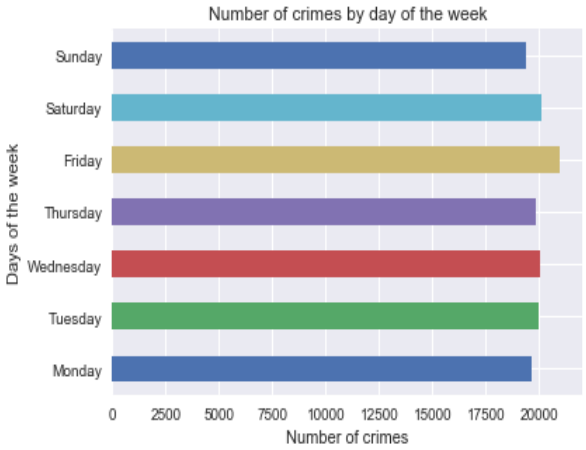
Visualization(EDA)

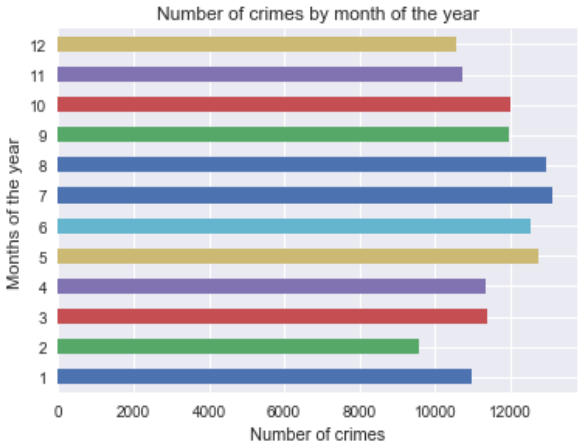


Visualization(EDA)

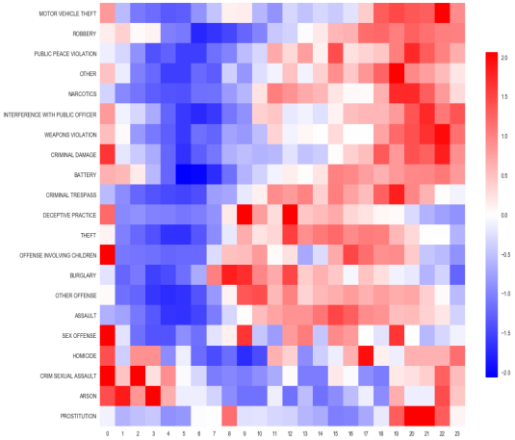


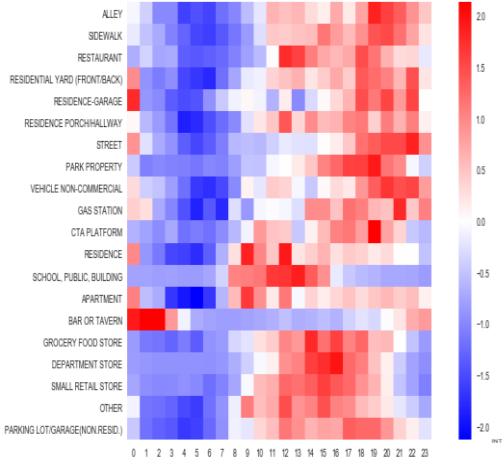
Visualization(EDA)

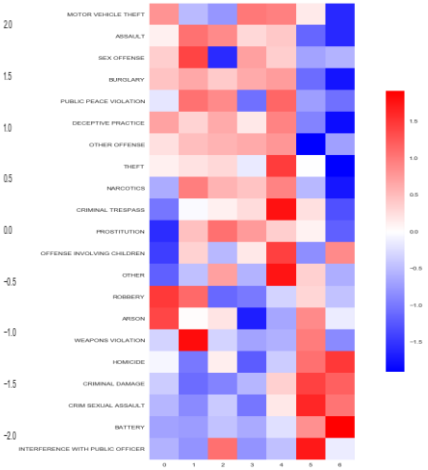


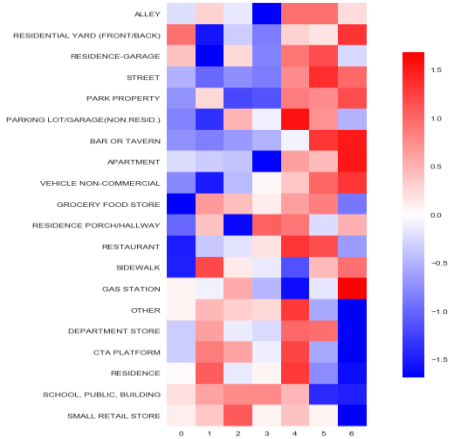


Visualization(EDA)

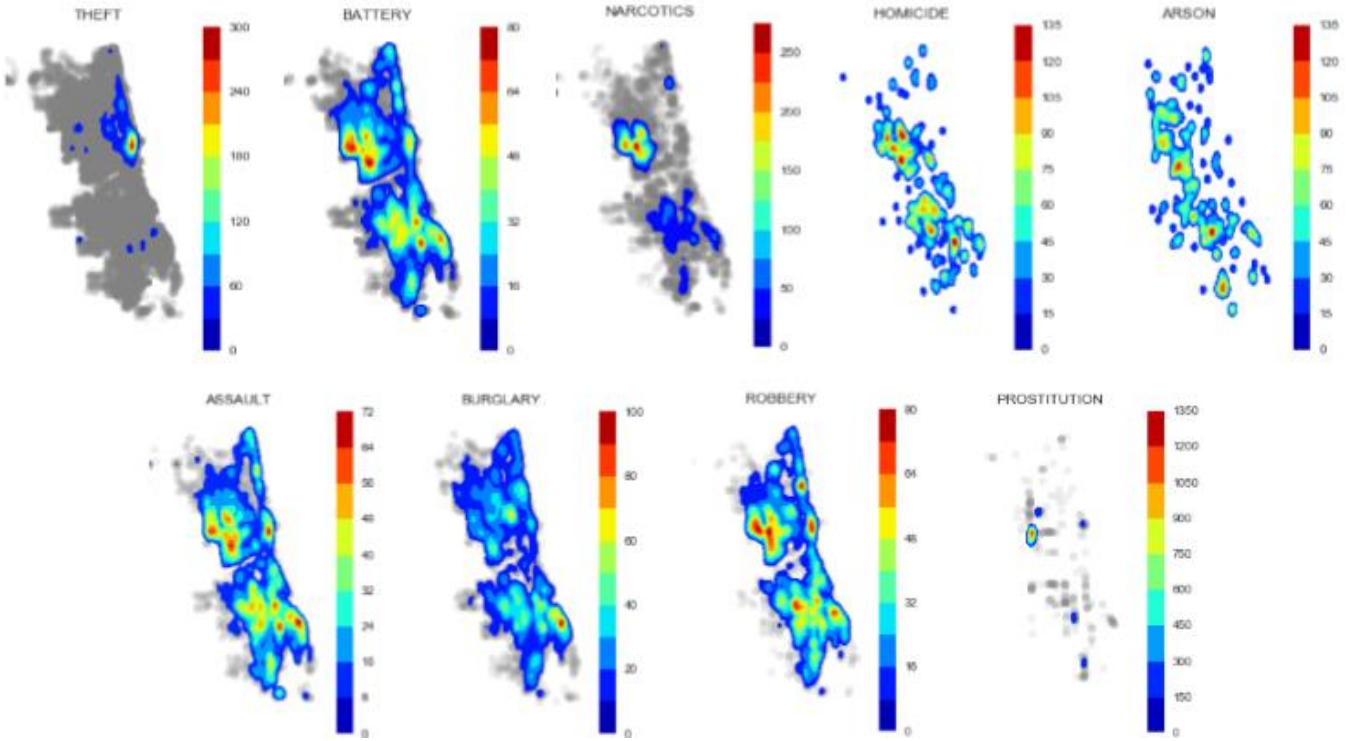








Visualization(EDA)



Modeling

Modeling

```
> table(crime$Primary.Type)

                          ARSON                          ASSAULT
                            209                             8729
                        BATTERY                         BURGLARY
                          25476                             8129
CONCEALED CARRY LICENSE VIOLATION              CRIM SEXUAL ASSAULT
                              8                              660
                 CRIMINAL DAMAGE                 CRIMINAL TRESPASS
                          15242                             3590
             DECEPTIVE PRACTICE                         GAMBLING
                           6770                              206
                       HOMICIDE  INTERFERENCE WITH PUBLIC OFFICER
                            229                              590
                   INTIMIDATION                       KIDNAPPING
                             76                              104
           LIQUOR LAW VIOLATION              MOTOR VEHICLE THEFT
                            193                             5948
                      NARCOTICS                   NON - CRIMINAL
                          12962                               10
                       OBSCENITY        OFFENSE INVOLVING CHILDREN
                             22                             1027
        OTHER NARCOTIC VIOLATION                     OTHER OFFENSE
                              3                             8466
                   PROSTITUTION                  PUBLIC INDECENCY
                            739                                8
          PUBLIC PEACE VIOLATION                         ROBBERY
                           1254                             5503
                     SEX OFFENSE                         STALKING
                            454                               79
                          THEFT                 WEAPONS VIOLATION
                          31719                             1595
```

```
> table(crime$Primary.Type)

   drug  others property  public     sex violent  white
  13158    8476    70297    1874    2959   36216   6906
```

Modeling

```
> crime[c("Primary.Type", "Date")]
# A tibble: 139,886 x 2
   Primary.Type Date
   <chr>        <chr>
 1 drug         10/04/2013 09:05:00 AM
 2 others       02/29/2012 05:33:00 PM
 3 property     12/19/2013 04:30:00 PM
 4 violent      09/15/2013 08:10:00 PM
 5 property     09/22/2012 08:00:00 AM
 6 violent      04/15/2013 10:15:00 PM
 7 property     02/08/2012 09:30:00 AM
 8 property     03/05/2016 12:00:00 PM
 9 property     07/12/2015 01:15:00 PM
10 others       03/22/2012 04:19:00 AM
# ... with 139,876 more rows
```

```
> crime[c("Primary.Type", "Date")]
# A tibble: 139,886 x 2
   Primary.Type Date
   <chr>        <chr>
 1 drug         10/04/2013 09:05:00 AM
 2 others       02/29/2012 05:33:00 PM
 3 property     12/19/2013 04:30:00 PM
 4 violent      09/15/2013 08:10:00 PM
 5 property     09/22/2012 08:00:00 AM
 6 violent      04/15/2013 10:15:00 PM
 7 property     02/08/2012 09:30:00 AM
 8 property     03/05/2016 12:00:00 PM
 9 property     07/12/2015 01:15:00 PM
10 others       03/22/2012 04:19:00 AM
# ... with 139,876 more rows
```

```
> crime[c("Primary.Type", "time", "time.tag")]
# A tibble: 139,886 x 3
   Primary.Type time     time.tag
   <chr>        <times>  <fct>
 1 drug         09:05:00 06-12
 2 others       17:33:00 12-18
 3 property     16:30:00 12-18
 4 violent      20:10:00 18-24
 5 property     08:00:00 06-12
 6 violent      22:15:00 18-24
 7 property     09:30:00 06-12
 8 property     12:00:00 06-12
 9 property     13:15:00 12-18
10 others       04:19:00 00-06
# ... with 139,876 more rows
```

```
> crime[c("Primary.Type", "day", "month")]
# A tibble: 139,886 x 3
   Primary.Type day   month
   <chr>        <chr> <chr>
 1 drug         Fri   Oct
 2 others       Wed   Feb
 3 property     Thu   Dec
 4 violent      Sun   Sep
 5 property     Sat   Sep
 6 violent      Mon   Apr
 7 property     Wed   Feb
 8 property     Sat   Mar
 9 property     Sun   Jul
10 others       Thu   Mar
# ... with 139,876 more rows
```

Modeling – Decision Tree

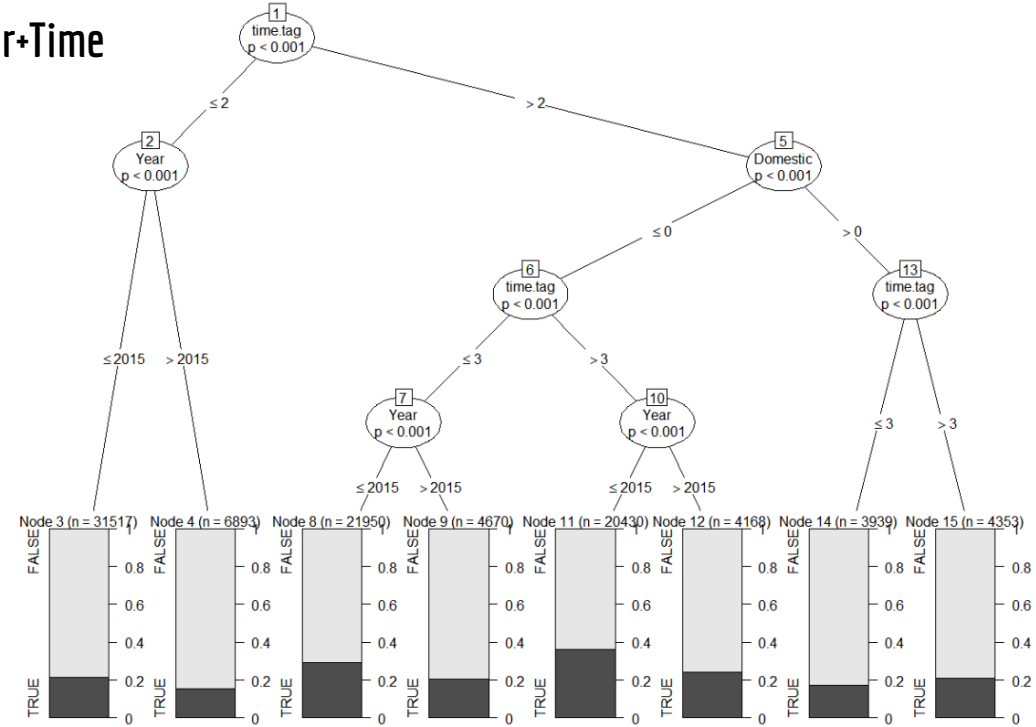
```
for(i in 1:nrow(crime)){
  if(crime$time.tag[i] == "00-06"){
    crime$time.tag[i] <- 1
    }
  if(crime$time.tag[i] == "06-12"){
    crime$time.tag[i] <- 2
    }
  if(crime$time.tag[i] == "12-18"){
    crime$time.tag[i] <- 3
    }
  if(crime$time.tag[i] == "18-24"){
    crime$time.tag[i] <- 4
  }
}
crime$time.tag<-as.numeric(crime$time.tag)
```

```
set.seed(100)
train.index<-sample(nrow(crime),139886*0.7)
train<-crime[train.index,]
test<-crime[-train.index,]
crime_ctree1<-ctree(Arrest~Domestic+Year+time.tag,data=train)
crime_ctree2<-ctree(Arrest~Year+time.tag,data=train)
crime_ctree3<-ctree(Domestic~Arrest+Year+time.tag,data=train)
crime_ctree4<-ctree(Domestic~Year+time.tag,data=train)
```

00-06 -> 1
06-12 -> 2
12-18 -> 3
18-24 -> 4

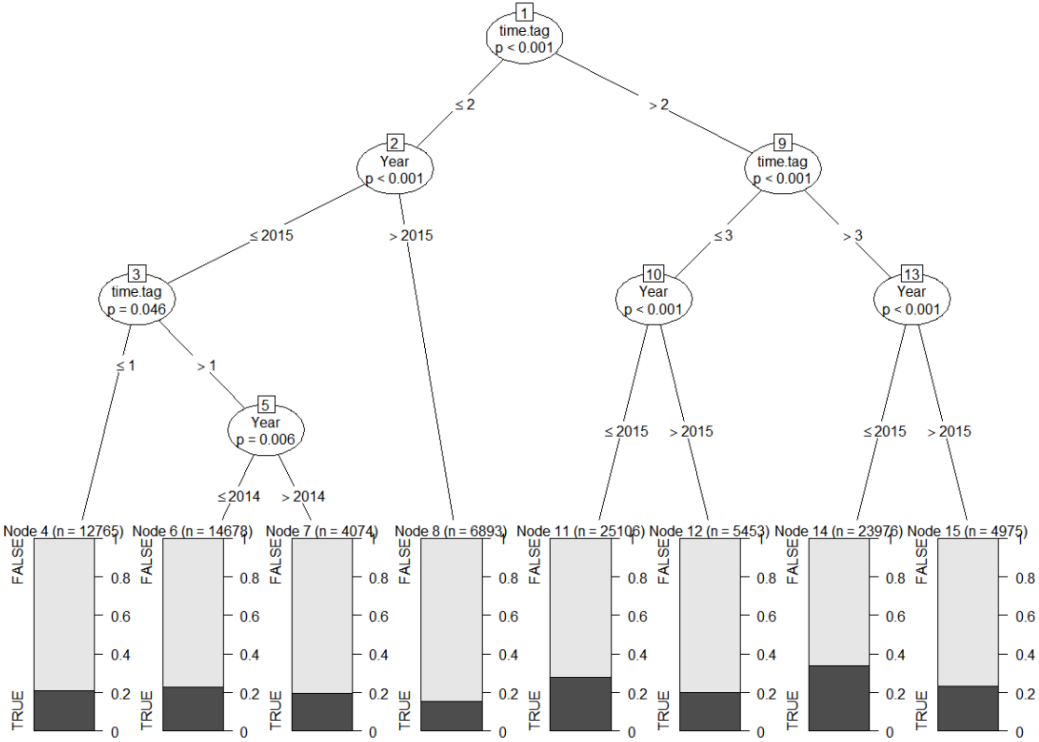
Modeling – Decision Tree

1. Arrest – Domestic+Year+Time



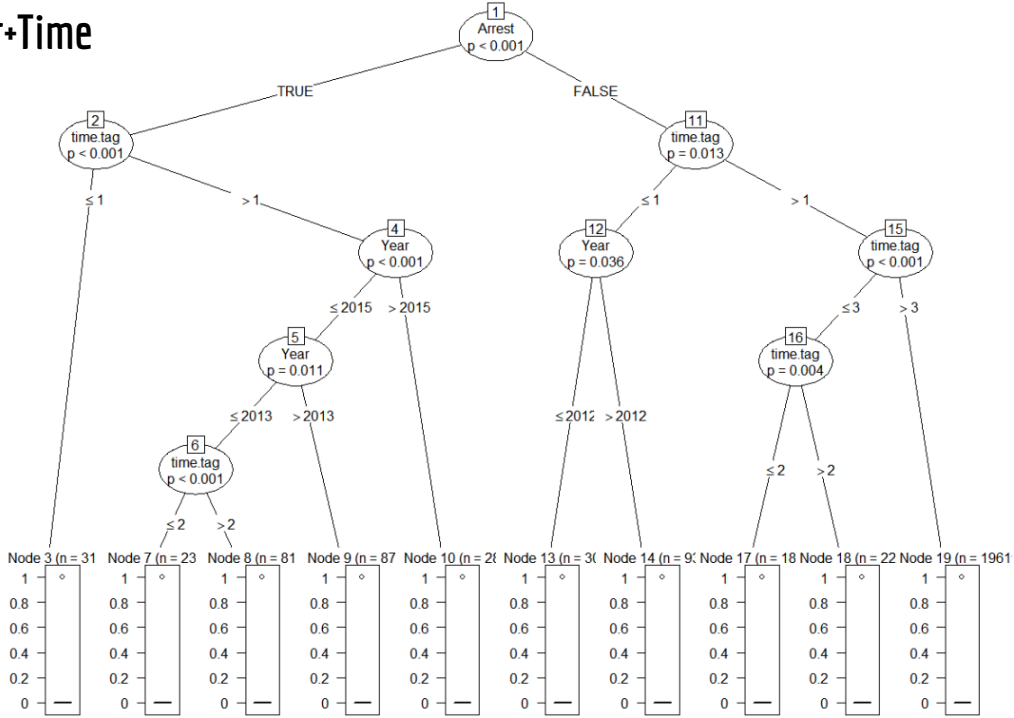
Modeling – Decision Tree

2 . Arrest – Year+Time



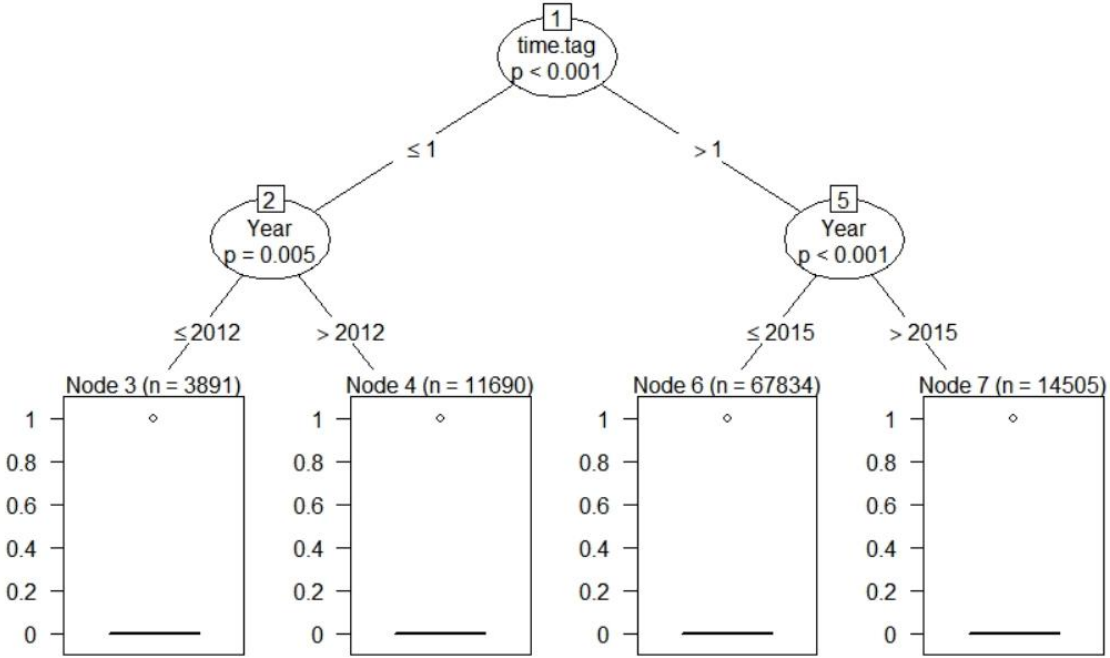
Modeling – Decision Tree

3. Domestic – Arrest+Year+Time



Modeling – Decision Tree

4. Domestic – Year+Time



Modeling – 범죄발생가능성 지수

1. 범죄유형 – 재범률에 따른 점수 부여

Drug, Property <- 5
Others, Public, White, Violent <- 4
Sex <- 2

```
chicago$type_score<-0
↓
for(i in 1:nrow(chicago)){
  if(chicago$Primary.Type[i]=="drug"|chicago$Primary.Type[i]=="property"){
    chicago$type_score[i]<-5
  }
  if(chicago$Primary.Type[i]=="others"|chicago$Primary.Type[i]=="public"|chic
ago$Primary.Type[i]=="white"|chicago$Primary.Type[i]=="violent"){
    chicago$type_score[i]<-4
  }
  if(chicago$Primary.Type[i]=="sex"){
    chicago$type_score[i]<-2
  }
}
```

수형자 특징	출소자의 비율	출소 후 3년 내 재범률			
		재체포	재유죄판결	새로운 장역형으로 인한 재복역률	전체 재복역
총 출소자	100	67.5	46.9	25.4	51.8
폭력 범죄	22.5	61.7	39.9	20.4	48.8
살인	1.7	40.7	20.5	10.8	31.4
납치	0.4	59.4	37.8	25.1	29.5
강간	1.2	46.0	27.4	12.6	43.5
다른 성범죄	2.4	41.4	22.3	10.5	36.0
강도	9.9	70.2	43.5	25.0	54.7
폭행	6.5	65.1	44.2	21.0	51.2
기타 폭력범죄	0.4	51.7	29.8	12.7	40.9
재산범죄	33.5	73.8	53.4	30.5	56.4
주거침입 절도	15.2	74.0	54.2	30.8	56.1
단순절도	9.7	74.6	55.7	32.6	60.1
차량 절도	3.5	78.8	54.3	31.3	59.1
방화	0.5	57.7	41.0	20.1	38.7
사기	2.9	66.3	42.1	22.8	45.4
장물	1.4	77.4	57.2	31.8	62.1
기타 재산범죄	0.3	71.1	47.6	28.5	40.0
약물 범죄	32.6	66.7	47.0	25.2	49.2
약물 소지	7.5	67.5	46.6	23.9	43.6
마약 밀매	20.2	64.2	44.0	24.8	46.1
기타/불특정	4.9	75.5	60.5	28.8	71.8
공공질서침해	9.7	62.2	42.0	21.6	48.0
무기	3.1	70.2	46.6	24.3	55.5
음주 운전	3.3	51.5	51.7	16.6	43.7
기타 공공질서침해	3.3	65.1	48.0	24.4	43.6
기타범죄	1.7	64.7	42.1	20.7	66.9

Modeling - 범죄발생가능성 지수

2. 범죄발생장소 – 빈도에 따른 점수 부여

0.2 이상 <- 5
0.1 이상 0.2 미만 <- 4
0.02 이상 0.1 미만 <- 3
0.01 이상 0.02 미만 <- 2
0.01 미만 <- 1

```
chicago$location_score<-0
y<-summary(chicago$Location.Description)/nrow(chicago)

for(i in 1:nrow(chicago)){
  x<-chicago$Location.Description[i]
  prob<-unname(y[which(x==names(y))])
  if(length(prob)>=1){
    chicago$location_score[i]<-ifelse(prob>=0.2,5,ifelse(prob>=0.1,4,ifelse(p
rob>=0.02,3,ifelse(prob>=0.01,2,1))))
  } else {
    chicago$location_score[i]<-1
  }
}
```

Modeling - 범죄발생가능성 지수

3. 가정범죄 여부에 따른 점수 부여

가정범죄 <- 5
가정범죄 이외의 범죄 유형 <- 0

```
chicago$domestic_score<-0
for(i in 1:nrow(chicago)){
  chicago$domestic_score[i]<-ifelse(chicago$Domestic[i]==TRUE,5,0)
}
```

Modeling - 범죄발생가능성 지수

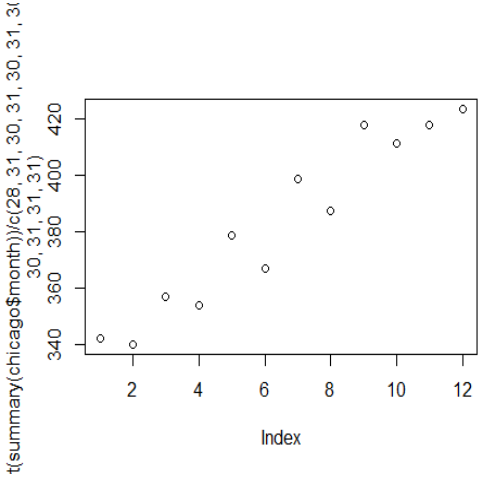
4. 범죄 발생 달 – 빈도에 따른 점수 부여

```
> sort(table(crime$month))
```

Feb	Dec	Nov	Jan	Apr	Mar	Sep	Oct	Jun	May	Aug	Jul
9587	10546	10712	10980	11361	11377	11958	12015	12535	12747	12943	13125

```
chicago$month_score<-0

for(i in 1:nrow(chicago)){
  if(chicago$month[i]=="Jul"|chicago$month[i]=="Jun"|chicago$month[i]=="Aug"|
chicago$month[i]=="May"){
    chicago$month_score[i]<-5
  }
  if(chicago$month[i]=="Oct"|chicago$month[i]=="Sep"){
    chicago$month_score[i]<-4
  }
  if(chicago$month[i]=="Mar"|chicago$month[i]=="Apr"){
    chicago$month_score[i]<-3
  }
  if(chicago$month[i]=="Jan"|chicago$month[i]=="Nov"){
    chicago$month_score[i]<-2
  }
  if(chicago$month[i]=="Feb"|chicago$month[i]=="Dec"){
    chicago$month_score[i]<-1
  }
}
```



5월, 6월, 7월, 8월 <- 5

9월, 10월 <- 4

4월, 5월 <- 3

1월, 11월 <- 2

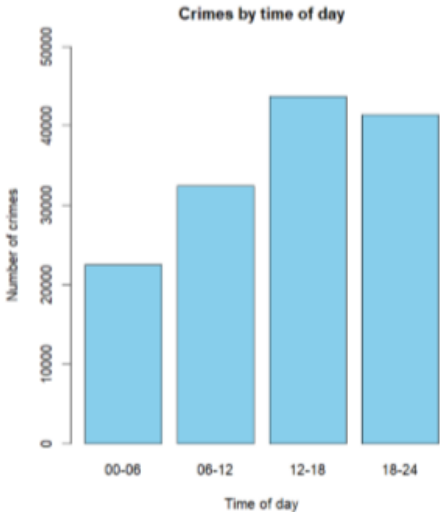
2월, 12월 <- 1

Modeling - 범죄발생가능성 지수

5. 시간대 – 빈도에 따른 점수 부여

12-18 <- 4.5
18-24 <- 4
06-12 <- 3
00-06 <- 2

```
chicago$time_score<-0
↓
for(i in 1:nrow(chicago)){
  x<-chicago$time.tag[i]
  chicago$time_score[i]<-ifelse(x=="12-18",4.5,ifelse(x=="18-24",4,ifelse(x==
"06-12",3,ifelse(x=="00-06",2,0))))
}
```



Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #1

범죄유형: 0.4
범죄 발생 장소: 0.25
가정 범죄 여부: 0.1
범죄 발생 달: 0.1
시간대: 0.15

```
chicago$crime_score<-0.4*chicago$type_score+0.25*chicago$location_score+0.1*c
hicago$domestic_score+0.1*chicago$month_score+0.15*chicago$time_score
```

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #1

- 최종 범죄 발생 가능성 지수가 가장 높은 사례 추출

```
chicago[which.max(chicago$crime_score),]
```

```
##       X.1   X1   X1_1       X      ID Case.Number                   Date
## 1454 1454 1454 200719 2534413 8626214    HV299466 05/23/2012 01:00:00 PM
##                      Block IUCR Primary.Type Description
## 1454 116XX S VINCENNES AVE 1320     property   TO VEHICLE
##      Location.Description Arrest Domestic Beat District Ward
## 1454               STREET  FALSE     TRUE 2212       22   34
##      Community.Area FBI.Code X.Coordinate Y.Coordinate Year
## 1454             75       14      1165794      1827755 2012
##                  Updated.On Latitude Longitude
## 1454 02/04/2016 06:33:39 AM 41.68294 -87.66872
##                          Location          newdate     time time.tag
## 1454 (41.682938571, -87.668723088) 2012-05-23 13:00 13:00:00    12-18
##        newdate1 month day type_score location_score domestic_score
## 1454 2012-05-23   May Wed          5              5              5
##      month_score time_score crime_score
## 1454           5        4.5       4.925
```

- 빈도수를 주로 활용하였기 때문에 빈도가 가장 큰 경우(ex. 재산범죄, 거리에서 범죄 발생)를 모두 포함한 범죄사건이 가장 높은 범죄발생 가능성 지수를 받음.

- 범죄 유형을 제외하고는 재발가능성을 고려하지 않았기 때문에 이러한 결과가 나온 것으로 보임

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #2

범죄유형: 0.6
범죄 발생 장소: 0.1
가정 범죄 여부: 0.1
범죄 발생 달: 0.1
시간대: 0.1

범죄유형에 높은 비중을 부여
Why? 재범률이 높은 범죄유형의
재범가능성을 강조하기 위함.

```
chicago$crime_score<-0.6*chicago$type_score+0.1*chicago$location_score+0.1*ch
icago$domestic_score+0.1*chicago$month_score+0.1*chicago$time_score
```

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #2

- 최종 범죄 발생 가능성 지수가 가장 높은 사례 추출

```r
chicago[which.max(chicago$crime_score),]
```

```
##       X.2  X.1   X1   X1_1       X      ID Case.Number
## 1454 1454 1454 1454 200719 2534413 8626214    HV299466
##                        Date                  Block IUCR Primary.Type
## 1454 05/23/2012 01:00:00 PM 116XX S VINCENNES AVE 1320     property
##      Description Location.Description Arrest Domestic Beat District Ward
## 1454  TO VEHICLE               STREET  FALSE     TRUE 2212       22   34
##      Community.Area FBI.Code X.Coordinate Y.Coordinate Year
## 1454             75       14      1165794      1827755 2012
##                  Updated.On Latitude Longitude
## 1454 02/04/2016 06:33:39 AM 41.68294 -87.66872
##                           Location          newdate     time time.tag
## 1454 (41.682938571, -87.668723088) 2012-05-23 13:00 13:00:00    12-18
##        newdate1 month day type_score location_score domestic_score
## 1454 2012-05-23   May Wed          5              5              5
##      month_score time_score crime_score
## 1454           5        4.5        4.95
```

-거리(street)에서 일어난 재산범죄(property)이자 가정범죄(domestic)이며 5월 수요일 13시경(time)에 일어난 범죄가 가장 높은 점수를 받음.

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #3

범죄유형: 0
범죄 발생 장소: 0.7
가정 범죄 여부: 0.1
범죄 발생 달: 0.1
시간대: 0.1

범죄 유형은 배제하고 범죄 장소에 높은 비중을 부여
why? 범죄 유형에 높은 비중을 부여하는 것은
현실의 범죄 예방에는 큰 도움이 안 될 것이다.
범죄가 어떤 지역에서 발생할 것인지를 예측해서
사전에 예방하는 것이 사회의 궁극적인 목표일 것.

```
chicago$crime_score<-0.7*chicago$location_score+0.1*chicago$domestic_score+0.
1*chicago$month_score+0.1*chicago$time_score
```

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #3

- 최종 범죄 발생 가능성 지수가 가장 높은 사례 추출

```r
chicago[which.max(chicago$crime_score),]
```

```
##     X.2 X.1  X1   X1_1       X      ID Case.Number                    Date
## 889 889 889 889 533020 2867945 9155534    HW301210 06/02/2013 01:53:00 PM
##                  Block IUCR Primary.Type                   Description
## 889 083XX S HALSTED ST  530      violent AGGRAVATED: OTHER DANG WEAPON
##     Location.Description Arrest Domestic Beat District Ward Community.Area
## 889               STREET   TRUE     TRUE  613        6   21             71
##     FBI.Code X.Coordinate Y.Coordinate Year             Updated.On
## 889      04A      1172423      1849658 2013 02/04/2016 06:33:39 AM
##     Latitude Longitude                     Location          newdate
## 889  41.7429 -87.64381 (41.742900733, -87.643814722) 2013-06-02 13:53
##         time time.tag   newdate1 month day type_score location_score
## 889 13:53:00    12-18 2013-06-02   Jun Sun          4              5
##     domestic_score month_score time_score crime_score
## 889              5           5        4.5        4.95
```

-거리(street)에서 일어난 가정범죄(domestic)이며 6월 일요일 14시경(time)에 일어난 범죄가 가장 높은 점수를 받음.

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수 #4

범죄유형: 0
범죄 발생 장소: 0.5
가정 범죄 여부: 0
범죄 발생 달: 0
시간대: 0.5
달을 12월로 고정

범죄유형, 가정범죄여부, 월 모든 것을 배제하고
오직 범죄 장소와 시간대에만 가중치 부여
Why? 장소와 시간대만 고려하면 높은 예측율을
기대할 수 있을 것이기 때문

현재 어떤 범죄가 일어날 가능성이 큰지를 예측하기
위해 month를 고정
Why? 당장 내일의 범죄가 궁금할 뿐이지, 6개월
후의 범죄가 지금 궁금하지는 않기 때문

```
chicago12<-subset(chicago,month=="Dec")
```

```
chicago$crime_score<-0.5*chicago$location_score+0.5*chicago$time_score
```

Modeling - 범죄발생가능성 지수

최종 범죄 발생 가능성 지수

- 최종 범죄 발생 가능성 지수가 가장 높은 사례 추출

```
##       X.2  X.1   X1   X1_1       X      ID Case.Number↓
## 1635 1635 1635 1635 968445 3303830 9888401    HX538614↓
##                        Date              Block IUCR Primary.Type↓
## 1635 12/11/2014 04:55:00 PM 005XX N WALLER AVE  486      violent↓
##                  Description Location.Description Arrest Domestic Beat↓
## 1635 DOMESTIC BATTERY SIMPLE               STREET  FALSE     TRUE 1512↓
##      District Ward Community.Area FBI.Code X.Coordinate Y.Coordinate Year↓
## 1635       15   29             25      08B      1138247      1902907 2014↓
##                  Updated.On Latitude Longitude↓
## 1635 02/04/2016 06:33:39 AM 41.88971 -87.76775↓
##                          Location          newdate     time time.tag↓
## 1635 (41.88970788, -87.767754466) 2014-12-11 16:55 16:55:00    12-18↓
##        newdate1 month day type_score location_score domestic_score↓
## 1635 2014-12-11   Dec Thu          4               5              5↓
##      month_score time_score crime_score↓
## 1635           1        4.5        4.55↵
```

-12월에는 거리(street)에서 목요일 17시경(time)에 일어난 범죄가 가장 높은 점수를 받음.

Modeling - Multilinear Regression

1. 모든 변수(7개) 고려

Response Variable : Arrest(체포 여부)
Explanatory Variable : Primary.Type(범죄 유형) / Domestic(가정 범죄 여부) / Ward(숫자로 주어진 위치, 구역 변수. factor형으로 변환 후 사용) / Year(년도) / time.tag(시간대) / month(달) / day(요일)

```
model<-glm(Arrest~Primary.Type+Domestic+Ward+Year+time.tag+month+day,data=chi
cago,family="binomial")↓
```

Modeling – Multilinear Regression

```
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         67.349373  10.692863   6.299 3.00e-10 ***
## Primary.Typeothers  -6.064322   0.106742 -56.813  < 2e-16 ***
## Primary.Typeproperty -6.834833  0.103839 -65.822  < 2e-16 ***
## Primary.Typepublic  -3.357512   0.119535 -28.088  < 2e-16 ***
## Primary.Typesex     -5.244595   0.110058 -47.653  < 2e-16 ***
## Primary.Typeviolent -5.789168   0.104109 -55.607  < 2e-16 ***
## Primary.Typewhite   -6.544573   0.108651 -60.235  < 2e-16 ***
## DomesticTRUE        -0.354786   0.021437 -16.550  < 2e-16 ***
## Ward2                0.432213   0.070349   6.144 8.05e-10 ***
## Ward3                0.378892   0.075194   5.039 4.68e-07 ***
## Ward4               -0.051981   0.087757  -0.592 0.553633
## Ward5                0.159399   0.077907   2.046 0.040754 *
## Ward6                0.318278   0.072249   4.405 1.06e-05 ***
## Ward7                0.221849   0.075515   2.938 0.003305 **
## Ward8               -0.023257   0.077398  -0.300 0.763802
## Ward9                0.308257   0.075414   4.088 4.36e-05 ***
## Ward10               0.270130   0.081703   3.306 0.000946 ***
## Ward11               0.126694   0.089790   1.411 0.158241
## Ward12               0.208034   0.089283   2.330 0.019803 *
## Ward13               0.322239   0.089052   3.619 0.000296 ***
## Ward14               0.139927   0.088330   1.584 0.113160
## Ward15               0.317983   0.075165   4.230 2.33e-05 ***
## Ward16               0.312002   0.075375   4.139 3.48e-05 ***
## Ward17               0.304101   0.072628   4.187 2.83e-05 ***
## Ward18               0.048926   0.085343   0.573 0.566449
## Ward19              -0.219984   0.110201  -1.996 0.045911 *
## Ward20               0.309558   0.073251   4.226 2.38e-05 ***
## Ward21               0.378183   0.073176   5.168 2.36e-07 ***
## Ward22               0.216203   0.090871   2.379 0.017349 *
## Ward23               0.019542   0.095356   0.205 0.837621
## Ward24               0.279887   0.071206   3.931 8.47e-05 ***
## Ward25               0.086245   0.089683   0.962 0.336217
## Ward26              -0.017855   0.090108  -0.198 0.842925
## Ward27               0.209974   0.074350   2.824 0.004741 **
## Ward28               0.478752   0.069500   6.889 5.64e-12 ***
## Ward29               0.133128   0.079150   1.682 0.092576 .
## Ward30               0.171385   0.088015   1.947 0.051508 .
## Ward31               0.256830   0.086850   2.957 0.003105 **
## Ward32              -0.150329   0.091555  -1.642 0.100599
## Ward33              -0.036364   0.100900  -0.360 0.718551
## Ward34               0.183717   0.076569   2.399 0.016424 *
## Ward35              -0.006289   0.095069  -0.066 0.947255
## Ward36               0.135042   0.098396   1.372 0.169926
## Ward37               0.437722   0.074877   5.846 5.04e-09 ***
## Ward38              -0.177555   0.102870  -1.726 0.084344 .
## Ward39              -0.269110   0.110306  -2.440 0.014701 *
## Ward40               0.091260   0.097378   0.937 0.348673
## Ward41               0.051018   0.095178   0.536 0.591940
## Ward42               0.671631   0.068218   9.845  < 2e-16 ***
## Ward43              -0.342381   0.102892  -3.328 0.000876 ***
## Ward44               0.267002   0.085361   3.128 0.001760 **
## Ward45               0.185158   0.096648   1.916 0.055392 .
## Ward46               0.577328   0.084717   6.815 9.44e-12 ***
## Ward47               0.001187   0.103661   0.011 0.990865
## Ward48               0.423111   0.096706   4.375 1.21e-05 ***
## Ward49               0.399606   0.086917   4.598 4.28e-06 ***
## Ward50              -0.198462   0.103753  -1.913 0.055770 .
## Year                -0.031180   0.005310  -5.872 4.31e-09 ***
## time.tag06-12       -0.088257   0.025603  -3.447 0.000566 ***
## time.tag12-18        0.149925   0.023485   6.384 1.73e-10 ***
## time.tag18-24        0.286647   0.023468  12.214  < 2e-16 ***
## monthAug            -0.130128   0.036181  -3.597 0.000322 ***
## monthDec            -0.149688   0.038439  -3.894 9.86e-05 ***
## monthFeb             0.048752   0.038495   1.266 0.205358
## monthJan            -0.043717   0.037580  -1.163 0.244715
## monthJul            -0.131773   0.036092  -3.651 0.000261 ***
## monthJun            -0.070415   0.036131  -1.949 0.051310 .
## monthMar             0.016112   0.036737   0.439 0.660970
## monthMay            -0.075798   0.035989  -2.106 0.035190 *
## monthNov            -0.154339   0.038373  -4.022 5.77e-05 ***
## monthOct            -0.111227   0.036848  -3.018 0.002540 **
## monthSep            -0.147742   0.036974  -3.996 6.45e-05 ***
## dayMon              -0.023607   0.028536  -0.827 0.408085
## daySat               0.028720   0.028082   1.023 0.306433
## daySun              -0.013855   0.028478  -0.486 0.626614
## dayThu               0.048693   0.028145   1.730 0.083617 .
## dayTue               0.063824   0.028026   2.277 0.022769 *
## dayWed               0.054116   0.028008   1.932 0.053338 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## (Dispersion parameter for binomial family taken to be 1)

##     Null deviance: 160125  on 139885  degrees of freedom
## Residual deviance: 112584  on 139808  degrees of freedom
## AIC: 112740

## Number of Fisher Scoring iterations: 7
```

Modeling - Multilinear Regression

2. LASSO로 무의미한 변수 삭제

```
### LASSO

library(glmnet) #ridge:alpha=0, lasso:alpha=1

crimes<-chicago[,c("Primary.Type","Arrest","Domestic","Ward","Year","time.tag","month","day")]

crimes$Primary.Type<-as.numeric(crimes$Primary.Type)
crimes$Arrest<-as.numeric(crimes$Arrest)
crimes$Domestic<-as.numeric(crimes$Domestic)
crimes$Ward<-as.numeric(crimes$Ward)
crimes$Year<-as.numeric(crimes$Year)
crimes$time.tag<-as.numeric(crimes$time.tag)
crimes$month<-as.numeric(crimes$month)
crimes$day<-as.numeric(crimes$day)

x_vars<-as.matrix(crimes[,-2])
y_var<-crimes$Arrest
lambda_seq <- 10^seq(2, -2, by = -.1)
set.seed(1)
train<-sample(1:nrow(x_vars),nrow(x_vars)/2)
test<-(-train)
y_test<-y_var[test]
cv_output<-cv.glmnet(x_vars[train,],y_var[train],alpha=1,lambda=lambda_seq)
best_lam<-cv_output$lambda.min
lasso_best<-glmnet(x_vars[train,],y_var[train],alpha=1,lambda=best_lam)
pre<-predict(lasso_best,s=best_lam,newx=x_vars[test,])
final<-cbind(y_var[test],pred)

coef(lasso_best) #범죄유형, 연도, 시간대가 유의미하다고 나왔는데 결과가 너무 별로라서 신뢰성이 매우 낮음
```

```
> coef(lasso_best)
8 x 1 sparse Matrix of class "dgCMatrix"
                    s0
(Intercept)  13.162247866
Primary.Type -0.039768798
Domestic      .
Ward          .
Year         -0.006371142
time.tag      0.027959689
month         .
day           .
>
```

Modeling - Multilinear Regression

2. LASSO로 무의미한 변수 삭제

Response Variable : Arrest(체포 여부)
Explanatory Variable : Primary.Type(범죄 유형) / Year(년도) / time.tag(시간대)

```
model2<-glm(Arrest~Primary.Type+Year+time.tag,data=chicago,family="binomial")
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           64.950612  10.626534   6.112 9.83e-10 ***
## Primary.Typeothers    -6.200298   0.106412 -58.267  < 2e-16 ***
## Primary.Typeproperty -6.881979   0.103647 -66.398  < 2e-16 ***
## Primary.Typepublic    -3.385439   0.119356 -28.364  < 2e-16 ***
## Primary.Typesex       -5.330894   0.109749 -48.574  < 2e-16 ***
## Primary.Typeviolent   -5.945720   0.103708 -57.331  < 2e-16 ***
## Primary.Typewhite     -6.531746   0.108307 -60.308  < 2e-16 ***
## Year                  -0.029897   0.005277  -5.665 1.47e-08 ***
## time.tag06-12         -0.064173   0.025251  -2.541    0.011 *
## time.tag12-18          0.192274   0.023134   8.311  < 2e-16 ***
## time.tag18-24          0.307032   0.023199  13.235  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 160125  on 139885  degrees of freedom
## Residual deviance: 113705  on 139875  degrees of freedom
## AIC: 113727
##
## Number of Fisher Scoring iterations: 7
```

Conclusion

Conclusion

How to use?

Identifying the characteristics of crimes that are more likely to occur or crimes that are less likely to be arrested and deploying stronger police forces in the area or time of day

THANK YOU