CMP7205 – Applied Statistics

REPORT TITLE

Applied Statistics Report

Zan Zver, 18133498

BIRMINGHAM CITY University

MSc (Hons) in Big Data Analytics
Faculty of Computing, Engineering and the Built Environment
Birmingham City University

Contents

List of Tables

List of Figures

Abstract

Summary of the entire report

1 Executive Summary

This is the Executive Summary section

1.1 Subsection

briefly introduce the problem, major questions and interesting findings of your report. Although present at the start of the report, this must be written after the rest of the report is complete and must be no longer than half a page;

2 Introduction

This is the Introduction section

2.1 Subsection

describe the purpose of the report - introduce the problem domain with suitable citations and list a set of related statistical questions;

3 Methodology

This is the Methodology section

3.1 Exploratory Data Analysis

list and justify the number of statistical techniques selected (such as those taught) to answer the previously outlined questions (or intermediate questions within) by showing how they are apt and what information can be extracted from them. This concerns marking criterion 2 listed later.

Viewing data can be done in different ways, in this case Exploratory Data Analysis (or EDA) was used. Purpose of EDA is to give visual representation of the data which can be used to reveal data patterns.

3.2 Correlation Analysis

list and justify the number of statistical techniques selected (such as those taught) to answer the previously outlined questions (or intermediate questions within) by showing how they are apt and what information can be extracted from them. This concerns marking criterion 2 listed later.

Two things causing an event is correlation, but is it cassation? By isolating the variables, we can see if they depend on each other in order for event to happen.

3.3 Regression Analysis

list and justify the number of statistical techniques selected (such as those taught) to answer the previously outlined questions (or intermediate questions within) by showing how they are apt and what information can be extracted from them. This concerns marking criterion 2 listed later.

4 Datasets

This is the Datasets section
Source: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2

4.1 Dataset description

describe and justify the datasets used to show what information is contained in the datasets and how it aids to answer the relevant questions. If the description is very long, then summarise the key information and detail the columns in the datasets in appropriate appendices. Also, properly cite the dataset in the References section and include their exact hyperlinks;

Column Name	Description	Type
ID	Unique identifier for the record.	Number
Case Number	The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.	Plain Text
Date	Date when the incident occurred. this is sometimes a best estimate.	Date and Time
Block	The partially redacted address where the incident occurred, placing it on the same block as the actual address.	Plain Text
IUCR	The Illinois Unifrom Crime Reporting code. This is directly linked to the Primary Type and Description. See the list of IUCR codes at https://data.cityofchicago.org/d/c7ck-438e.	Plain Text
Primary Type	The primary description of the IUCR code.	Plain Text
Description	The secondary description of the IUCR code, a subcategory of the primary description.	Plain Text
Location Description	Description of the location where the incident occurred.	Plain Text
Arrest	Indicates whether an arrest was made.	Checkbox

Domestic	Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.	Checkbox
Beat	Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts. See the beats at https://data.cityofchicago.org/d/aerh-rz74.	Plain Text
District	Indicates the police district where the incident occurred. See the districts at https://data.cityofchicago.org/d/fthy-xz3r.	Plain Text
Ward	The ward (City Council district) where the incident occurred. See the wards at https://data.cityofchicago.org/d/sp34-6z76.	Number
Community Area	Indicates the community area where the incident occurred. Chicago has 77 community areas. See the community areas at https://data.cityofchicago.org/d/cauq-8yn6.	Plain Text
FBI Code	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS). See the Chicago Police Department listing of these classifications at https://ucr.fbi.gov/nibrs/2011/resources/nibrs-offense-codes/view	Plain Text
X Coordinate	The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.	Number

Y Coordinate	The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.	Number
Year	Year the incident occurred.	Number
Updated On	Date and time the record was last updated.	Date and Time
Latitude	The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.	Number
Longitude	The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.	Number
Location	The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.	Location

Table 1: Main dataset attributes

(Data Pre-processing): discuss any data pre-processing that is required to prepare the data for subsequent analysis.

Table X is representing removed attributes and reasons why they are removed.

Column Name	Description
X.Coordinate	Not useful to us. Location can be obtain from Latitude and Longitude coordinate.
Y.Coordinate	Not useful to us. Location can be obtain from Latitude and Longitude coordinate.
Location	Not useful to us. Location can be obtain from Latitude and Longitude coordinate.

Year	Attribute Year is the same as in Date section. Since we don't want duplicate values, it is removed.

Table 2: Removed attributes

Column Name	New data type	Description
ID	integer	ID is a positive whole number. To eliminate decimal places, integer was chosen.
Case Number	character	Case Number is combination of two letters and 6 numbers. Ideally we would split two character into one column and numbers into another. At the moment we cannot verify if there are any exceptions where there would be more (or less) than 2 characters. Due to aforementioned reasons, the whole column is saved as a character.
Date	POSIXct	Date is originally stored as a string. To have more control over that, we are transforming it as POSIXct with the year-month-day hour:minute:second format. Originally time is indicated with AM/PM, but this is converted to 24h notation as well.
Block	character	Block is saved as a character due to mix of numbers and letters. This contains house number and street. House number has first 3 or 4 characters are numbers, while last 1 or 2 are X symbols. This is done to anonymize the data. We could separate house number and street into two different columns, but this will not be done since house number is not much of the use at the moment. A lot of information can be gathered from Location column anyway.

IUCR	character	IUCR column is saved as character due to mix of numbers and letters. There is a predetermined list of IUCR codes that we do check against, to remove all of the codes which are not set as predetermined.
Primary Type	character	This is a short crime description.
Description	character	Longer crime description.
Location Description	character	Location description where crime was committed.
Arrest	logical	Logical information if person was arrested (as TRUE) or not (as FALSE).
Domestic	logical	Logical information if crime was domestic (as TRUE) or not (as FALSE).
Beat	integer	Beat is a positive whole number that indicates location of the crime (as an ID). Numbers are cross referenced with data in PoliceBeat report. Mismatches are removed.
District	integer	This is a positive whole number that indicates what police district handled the crime. District number is verified with the data found in PoliceDistrict file. Mismatches are removed.
Ward	integer	Ward is a positive whole number that locates where crime happened. Number is an ID of City Council district.
Community Area	integer	Community Area is an ID presented as whole positive number. It is checked against CommAreas file. Only matches are kept.
FBI Code	character	This is a crime code indicated by FBI. It contains letters and numbers, therefore it is saved as a character.

Updated On	POSIXct	Date is originally stored as a string. To have more control over that, we are transforming it as POSIXct with the year-month-day hour:minute:second format. Originally time is indicated with AM/PM, but this is converted to 24h notation as well.
Latitude	numeric	LAT is saved as numeric since it represents positive numbers with decimal places. It would be useful to have a range to limit LAT that is out of range.
Longitude	numeric	LON is saved as numeric since it represents negative numbers with decimal places. It would be useful to have a range to limit LON that is out of range.

Table 3: Dataset attributes with new types

Illinois Uniform Crime Reporting (IUCR)
https://data.cityofchicago.org/Public-Safety/Chicago-Police-Department-Illinois-Uniform-Crime-R/c7ck-438e

Column Name	Description	Type
IUCR	Crime ID	Plain Text
Primary description	First (main) description of the crime	Plain Text
Secondary description	Additional description of the crime	Plain Text
Index code	There are 2 codes, "I" (Index) and "N" (Non-Index). "I" code indicates crimes that are collected nation-wide (by FBI) while "N" is for other (usually smaller) crimes.	Plain Text
Active	Whether the code is active. Retired codes (No) are present in this dataset for historical reference. There is a filtered view for this dataset showing only active codes.	Boolean

Table 4: IUCR supporting dataset attributes

Beats
https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Beats-current-/aerh-rz74

Column Name	Description	Type
The_geom	List of locations (as multi polygon) containing LAT and LON.	Multipolygon - location
District	ID of the police district	Number
Sector	Geographically divided area with associated ID	Number
Beat	No description.	Number
Beat_num	Indicates Beat ID from main dataset	Number

Table 5: Beats supporting dataset attributes

Police Districts
https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Districts-current-/fthy-xz3r

Column Name	Description	Type
The_geom	List of locations (as multi polygon) containing LAT and LON where districts operate.	Multipolygon - location
Dist_label	Name of the district	Plain text
Dist_num	Number of the district	Plain text

Table 6: Police districts supporting dataset attributes

Community Areas
https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-

Areas-current-/cauq-8yn6

Column Name	Description	Type
The_geom	List of locations (as multi polygon) containing LAT and LON.	Multipolygon - location
Perimeter	No description.	Number
Area	No description.	Number
Comarea_	No description.	Number
Comarea_ID	No description.	Number
Area_numbe	ID of the area. This is the the same as in main dataset. Note, column name does not have r in number.	Number
Community	Name of the community / region.	Plain Text
Area_num_1	ID of the area. This is the the same as in main dataset. Note, column name does not have r in number.	Number
Shape_area	No description.	Number
Shape_len	No description.	Number

Table 7: Community area supporting dataset attributes

FBI codes
https://ucr.fbi.gov/nibrs/2011/resources/nibrs-offense-codes/view
Group A Offenses

Offense Code	Offense Description	Crime Against
Arson		
200	Arson	Property
Assault Offenses		
13A	Aggravated Assault	Person
13B	Simple Assault	Person
13C	Intimidation	Person
Bribery		
510	Bribery	Property
Burglary/Breaking & Entering		
220	Burglary/Breaking & Entering	Property

Counterfeiting/Forgery		
250	Counterfeiting/Forgery	Property
Destruction/Damage/Vandalism of Property		
290	Destruction/Damage/Vandalism of Property	Property
Drug/Narcotic Offenses		
35A	Drug/Narcotic Violations	Society
35B	Drug Equipment Violations	Society
Embezzlement		
270	Embezzlement	Property
Extortion/Blackmail		
210	Extortion/Blackmail	Property
Fraud Offenses		
26A	False Pretenses/Swindle/Confidence Game	Property
26B	Credit Card/Automated Teller Machine Fraud	Property
26C	Extortion/Blackmail	Property
26D	Welfare Fraud	Property
26E	Wire Fraud	Property
Gambling Offenses		
39A	Betting/Wagering	Property
39B	Operating/Promoting/Assisting Gambling	Property
39C	Gambling Equipment Violations	Property
39D	Sports Tampering	Property
Homicide Offenses		
09A	Murder & Non-negligent Manslaughter	Property
09B	Negligent Manslaughter	Property
09C	Justifiable Homicide	Person/Not a Crime
Kidnapping/Abduction		
100	Kidnapping/Abduction	Person
Larceny/Theft Offenses		

23A	Pocket-picking	Person
23B	Purse-snatching	Person
23C	Shoplifting	Person
23D	Theft From Building	Person
23E	Theft From Coin-Operated Machine or Device	Person
23F	Theft From Motor Vehicle	Person
23G	Theft of Motor Vehicle Parts or Accessories	Person
23H	All Other Larceny	Person
Motor Vehicle	Theft	
240	Motor Vehicle Theft	Property
Pornography /	Obscene Material	
370	Pornography/Obscene Material	Society
Prostitution Offenses		
40A	Prostitution	Society
40B	Assisting or Promoting Prostitution	Society
Robbery		
120	Robbery	Property
Sex Offenses, Forcible		
11A	Forcible Rape	Person
11B	Forcible Sodomy	Person
11C	Sexual Assault With An Object	Person
11D	Forcible Fondling	Person
Sex Offenses, Nonforcible		
36A	Incest	Person
36B	Statutory Rape	Person
Stolen Property Offenses		
280	Stolen Property Offenses	Property
Weapon Law Violations		
520	Weapon Law Violations	Society

Table 8: FBI supporting dataset attributes group A

Group B Offenses

Offense Code	Offense Description	Crime Against
90A	Bad Checks	Property
90B	Curfew/Loitering/Vagrancy Violations	Society
90C	Disorderly Conduct	Society
90D	Driving Under the Influence	Society
90E	Drunkenness	Society
90F	Family Offenses, Nonviolent	Society
90G	Liquor Law Violations	Society
90H	Peeping Tom	Society
90I	Runaway	Not a Crime
90J	Trespass of Real Property	Society
90Z	All Other Offenses	Person, Property, or Society

Table 9: FBI supporting dataset attributes group A

TO DO:
1- data import
import data to R, break data to date & time, remove X and y coordinate, remove location
DONE!!!!!!!
2- remove empty data and remove outliers
3- see how data is shaped, what correlations are formed, etc...

4.2 Exploratory Data Analysis

4.2.1 Arrest

Arrest description Most of the people are not arrested at the end of the police interaction. Why is that? Is it due to minor arrests?

```
  FALSE    TRUE
5138669 1818347
```

Figure 1: Text representation of all values in Arrest column

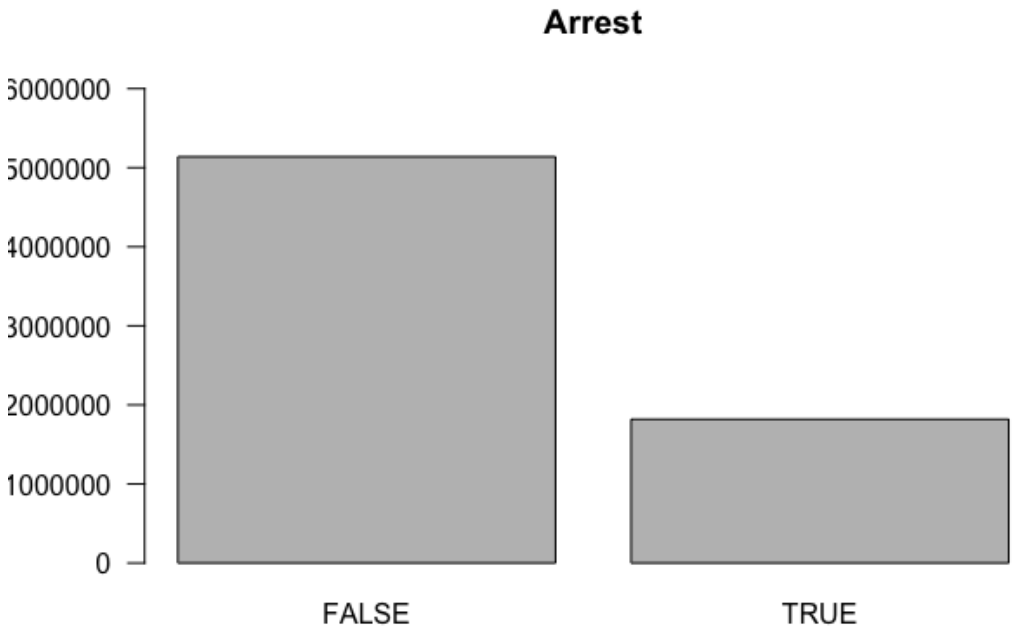

Figure 2: Graph representation of all values in Arrest column

4.2.2 Beat

Graph presents top 50

```
  421   423   624  1834   511  1533  1112   823   414  1522  2533   621   612   321   631   825  1011   522
54592 54343 50006 49241 48406 48096 46846 46442 44285 44181 42887 42842 42411 40803 40243 40041 39733 39300
  512   713  1532  1122   831  1121   632  1831  1531   122   832   623   324  1132   712  1133   513   312
39245 39132 38353 38064 37891 37516 37221 37206 36781 36721 36586 36268 36203 36093 36033 35881 35771 35698
  523  2534   813   611   833   725  1523   835   834   323   331   424  2512   412
35007 34867 34797 34354 34300 33572 33324 33089 33008 32995 32625 32543 32529 32354
```

Figure 3: Text representation of top 50 occurred values in Beat column

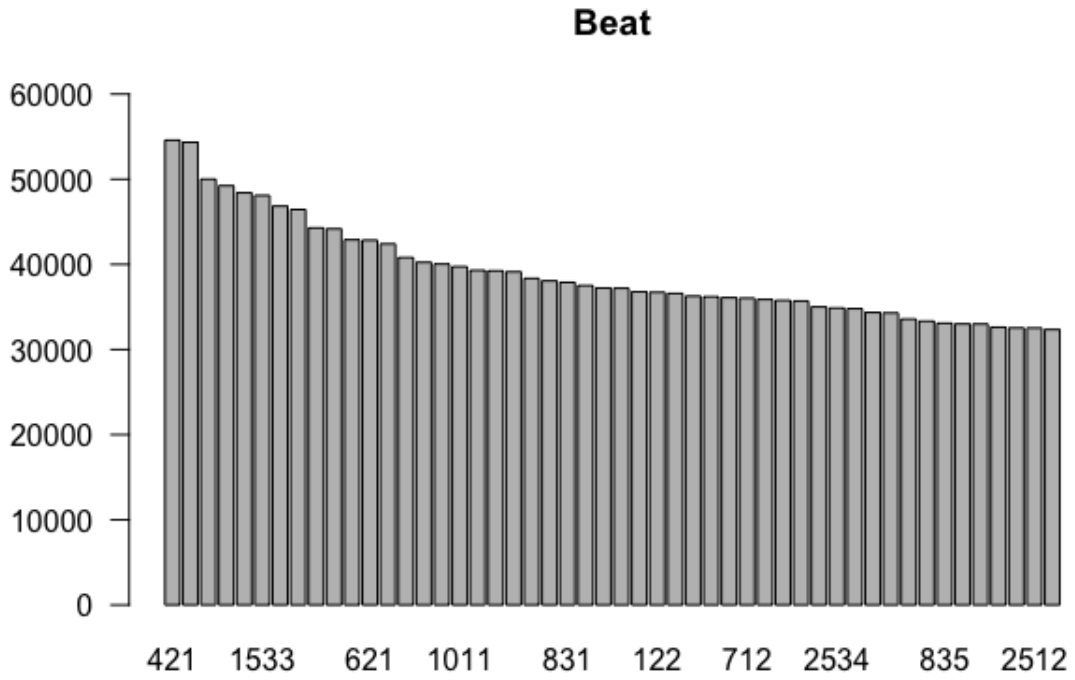

Figure 4: Graph of top 50 occurred values in Beat column

4.2.3 Block

Graph presents top 30 Top 2 blocks are the highest and have similar number of encounters. Are they close? Why are top 8 blocks "higher"?

001XX N STATE ST	100XX W OHARE ST	076XX S CICERO AVE
13670	13625	9903
008XX N MICHIGAN AVE	0000X N STATE ST	0000X W TERMINAL ST
9285	8249	6097
064XX S DR MARTIN LUTHER KING JR DR	063XX S DR MARTIN LUTHER KING JR DR	023XX S STATE ST
5819	5469	4659
001XX W 87TH ST	012XX S WABASH AVE	006XX N MICHIGAN AVE
4537	4291	4232
008XX N STATE ST	009XX W BELMONT AVE	057XX S CICERO AVE
4103	4021	3959
0000X S STATE ST	075XX S STONY ISLAND AVE	071XX S JEFFERY BLVD
3749	3717	3695
002XX W 87TH ST	007XX N MICHIGAN AVE	0000X W 95TH ST
3689	3652	3634
038XX W ROOSEVELT RD	011XX W WILSON AVE	046XX W NORTH AVE
3597	3565	3556
065XX S DR MARTIN LUTHER KING JR DR	005XX N MICHIGAN AVE	022XX S STATE ST
3495	3473	3388
005XX E BROWNING AVE	085XX S COTTAGE GROVE AVE	009XX N MICHIGAN AVE
3298	3242	3220

Figure 5: Text of top 30 occurred values in Block column

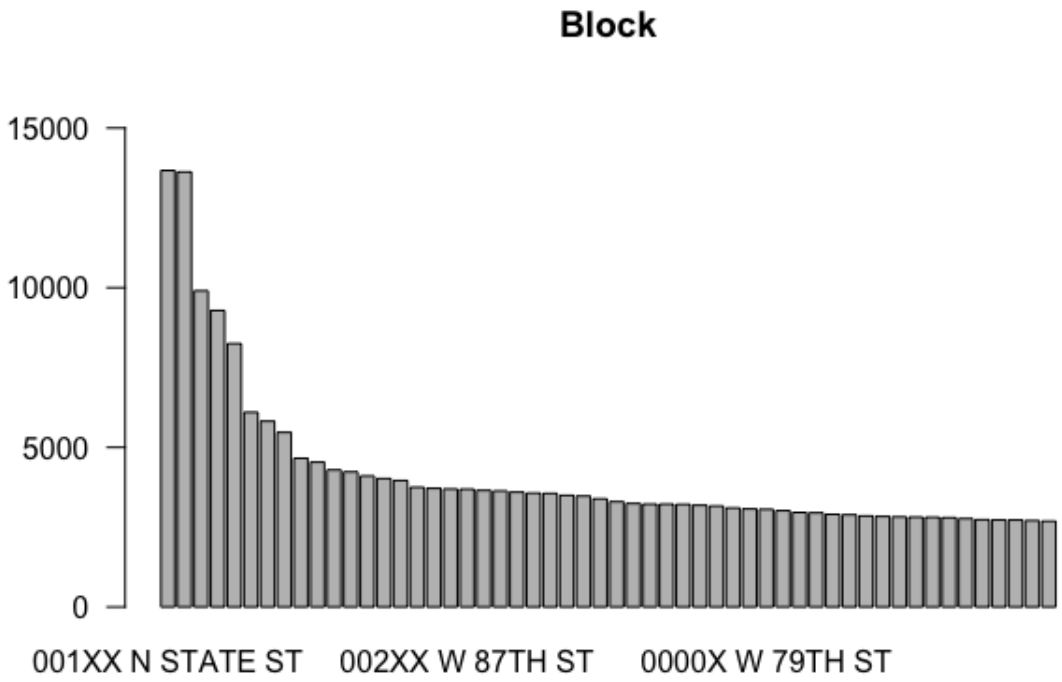

Figure 6: Graph of top 30 occurred values in Block column

4.2.4 Community area

Community area description What is going one with one community area? Why is it so high? What crimes are in here?

25	8	43	23	28	24	29	67	71	49	68	69	32	66	44	22
437335	243872	229705	217907	208759	203456	202954	200733	197742	185511	182687	173385	171182	170520	153177	143990
61	6	26	27	46	19	30	53	42	7	1	3	38	2	15	73
140893	139873	131656	130598	128712	127420	117127	114255	111980	108204	107154	101655	96318	88658	88086	82811
16	35	40	77	31	58	21	63	70	14	56	75	33	65	4	51
78426	77371	73856	69305	68121	67248	64485	63516	63112	62183	57328	55527	53346	51248	49569	45815
60	41	17	20	5	39	76	48	45	52	54	10	59	50	11	64
44454	44340	42872	41813	41013	39902	38909	38033	35756	34415	31387	30144	28342	28168	27739	27553
62	34	72	57	13	37	18	36	74	55	12	47	9	0		
26790	26312	25232	24917	23413	23268	16513	15783	15644	15360	12808	10442	6856	67		

Figure 7: Text representation of all values in CommunityArea column

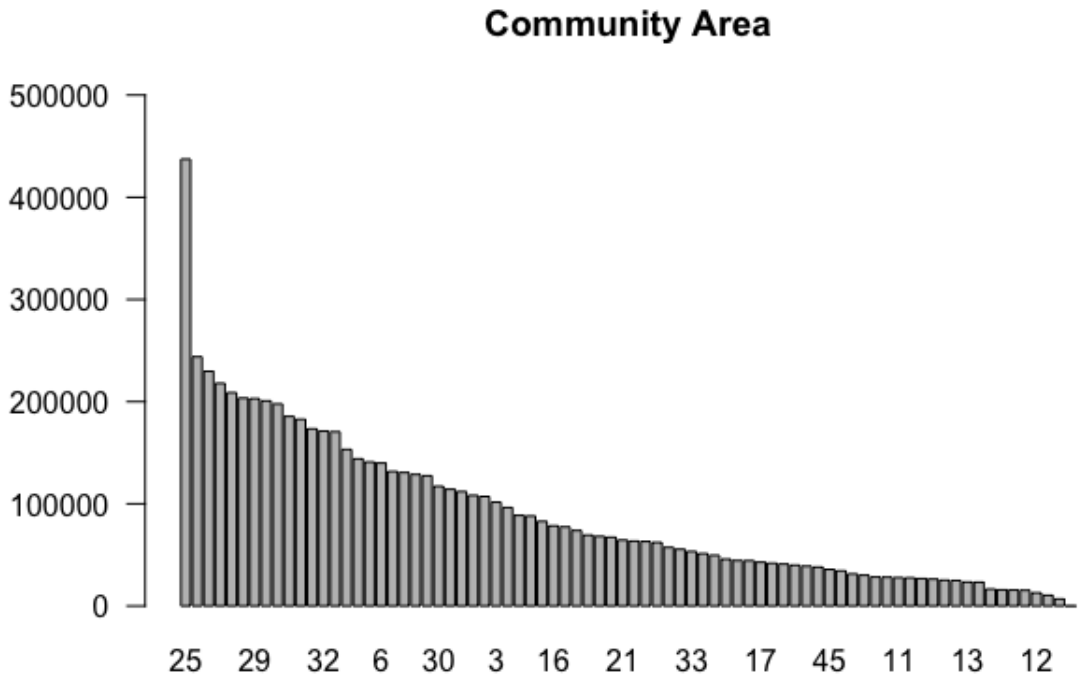

Figure 8: Graph representation of all values in CommunityArea column

4.2.5 Date

Graph presents top 54 Most of the arrests seem to be done on 1/1. This needs to be checked. Why? Is it insertion error? What is the most frequent date?

2007-01-01 00:01:00	2008-01-01 00:01:00	2012-01-01 00:01:00	2007-01-01 00:00:00	2003-01-01 00:00:00
172	166	156	145	140
2003-01-01 00:01:00	2011-01-01 00:01:00	2004-01-01 00:00:00	2016-01-01 00:01:00	2010-01-01 00:01:00
139	136	132	132	130
2010-01-01 00:00:00	2008-01-01 00:00:00	2009-01-01 00:00:00	2009-01-01 00:01:00	2006-01-01 00:01:00
125	119	117	113	112
2013-01-01 00:01:00	2011-01-01 00:00:00	2005-01-01 00:01:00	2012-01-01 00:00:00	2017-01-01 00:01:00
110	107	99	99	99
2006-01-01 00:00:00	2018-01-01 00:01:00	2017-01-01 00:00:00	2005-01-01 00:00:00	2013-01-01 09:00:00
97	97	96	94	92
2003-01-01 12:00:00	2002-10-01 00:00:00	2020-05-31 16:00:00	2021-01-01 00:00:00	2001-01-01 00:00:00
91	90	90	89	88
2013-01-01 00:00:00	2003-08-01 12:00:00	2003-11-01 12:00:00	2022-01-01 00:00:00	2003-12-01 00:00:00
87	86	86	86	85
2011-01-01 09:00:00	2016-01-01 00:00:00	2003-10-01 00:00:00	2004-06-01 12:00:00	2007-08-01 00:00:00
85	85	84	84	84
2015-01-01 00:01:00	2014-01-01 00:01:00	2012-01-01 09:00:00	2001-01-01 00:01:00	2002-06-01 00:00:00
84	83	82	80	80
2003-08-01 00:00:00	2007-09-01 00:00:00	2020-05-31 15:00:00	2002-01-01 00:00:00	2002-08-01 00:00:00
80	79	79	78	78
2006-09-01 00:00:00	2003-11-01 00:00:00	2014-01-01 00:00:00	2003-09-01 00:00:00	
78	77	77	76	

Figure 9: Text of top 54 occurred values in Date column

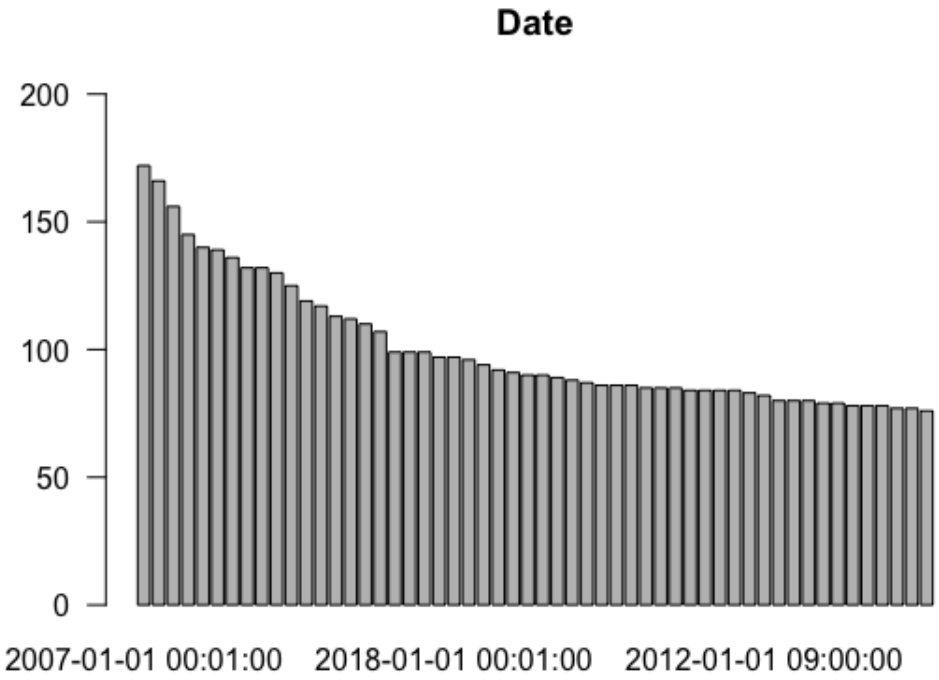

Figure 10: Graph of top 54 occurred values in Date column

4.2.6 Description

Graph presents top 30 What are the top committed crimes? Are any of them related?

SIMPLE	DOMESTIC BATTERY SIMPLE	$500 AND UNDER
769286	597002	539969
TO VEHICLE	TO PROPERTY	OVER $500
385592	366436	358247
FORCIBLE ENTRY	AUTOMOBILE	POSS: CANNABIS 30GMS OR LESS
257828	257686	256848
FROM BUILDING	RETAIL THEFT	TELEPHONE THREAT
248681	197721	134561
TO LAND	UNLAWFUL ENTRY	POSS: CRACK
117891	105791	101712
HARASSMENT BY TELEPHONE	ARMED: HANDGUN	POSS: HEROIN(WHITE)
97468	85275	84842
STRONGARM - NO WEAPON	AGGRAVATED: OTHER DANG WEAPON	AGGRAVATED: HANDGUN
83899	83472	78557
CREDIT CARD FRAUD	AGGRAVATED:KNIFE/CUTTING INSTR	UNLAWFUL POSS OF HANDGUN
59850	51958	48785
SOLICIT ON PUBLIC WAY	POCKET-PICKING	FINANCIAL ID THEFT: OVER $300
42210	38466	37856
FRAUD OR CONFIDENCE GAME	VIOLATE ORDER OF PROTECTION	CRIMINAL DEFACEMENT
37068	36212	35990

Figure 11: Text of top 30 occurred values in Description column

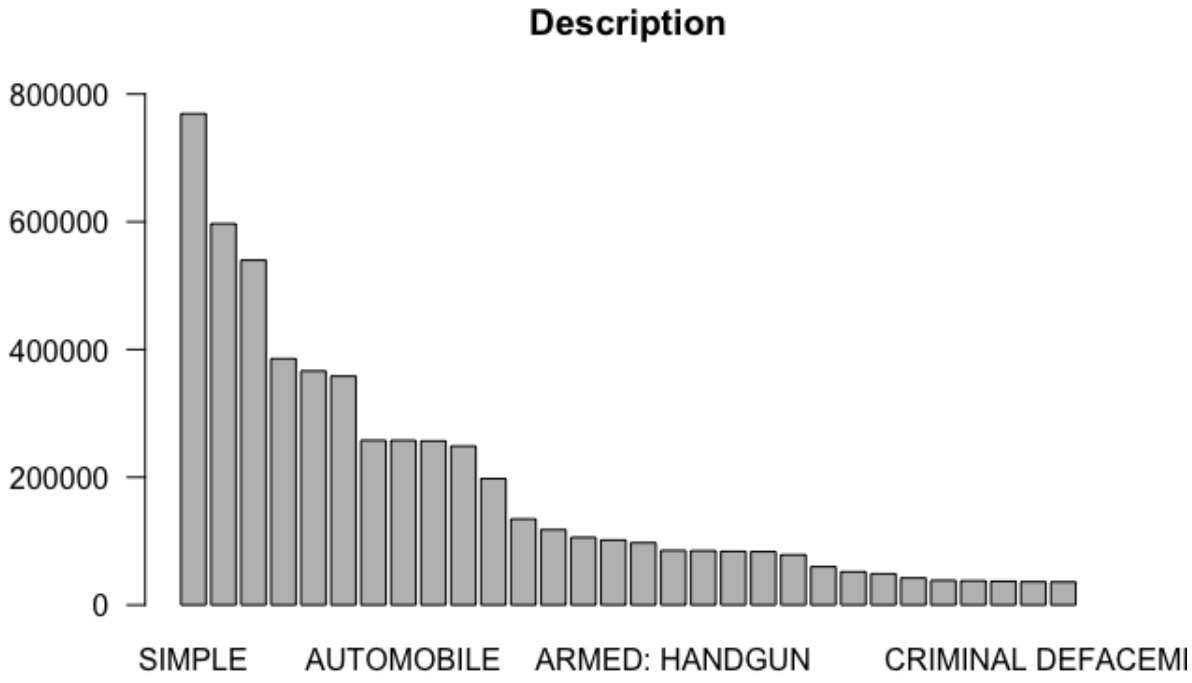

Figure 12: Graph of top 30 occurred values in Description column

4.2.7 District

District description All of the districts seem to be working fine except 31 and 21. Why? What is the reason?

```
     8     11      6      7      4     25      3     12      9      2      5     18     19     10     15      1
470598 449882 410914 407123 398649 396104 354231 341050 340387 314996 310733 310059 309837 302017 300555 283891
    14     16     22     24     17     20     31     21
265086 231668 229213 208990 200001 120850    178      4
```

Figure 13: Text representation of all values in District column

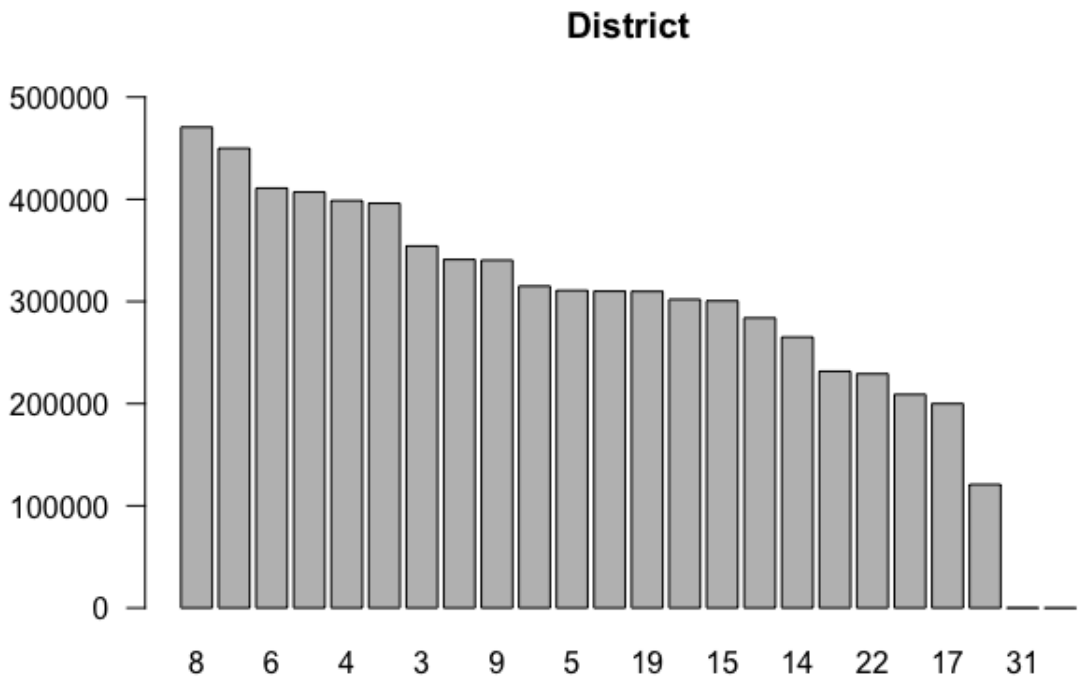

Figure 14: Graph representation of all values in District column

4.2.8 Domestic

Domestic description

```
  FALSE    TRUE
5985713  971303
```

Figure 15: Text representation of all values in Domestic column

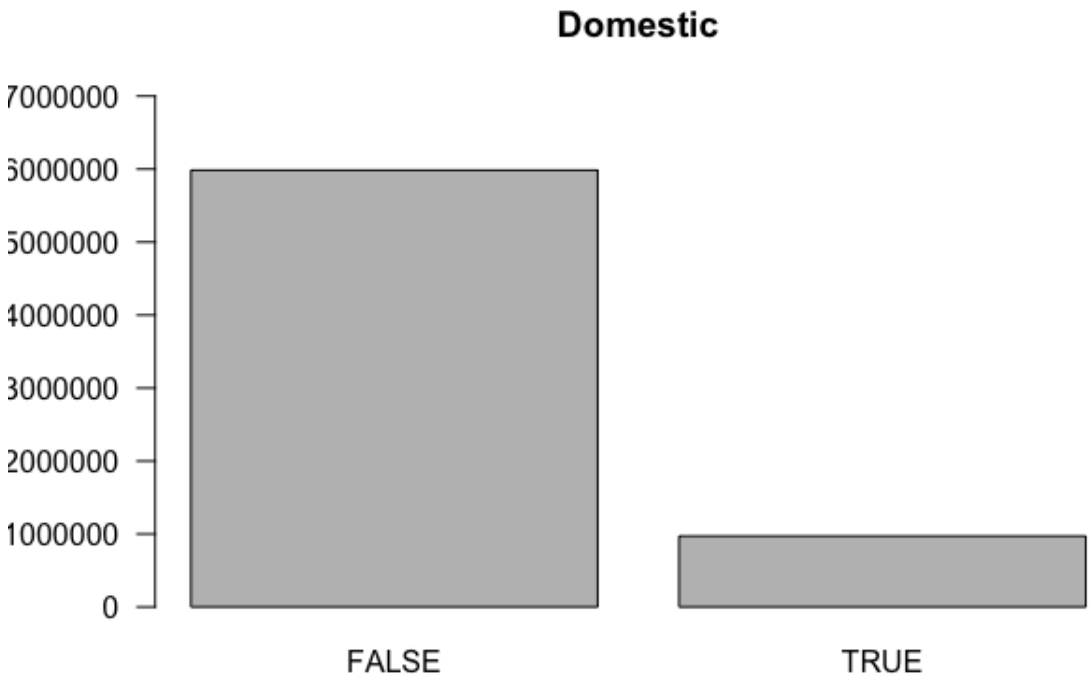

Figure 16: Graph representation of all values in Domestic column

4.2.9 FBI code

FBI code description Most of the FBI codes are 06 related, have a look why.

06	08B	14	26	18	05	08A	07	03	11	04B	04A	15	16
1471191	1091592	796297	681646	628465	386299	348946	322196	262739	257653	189146	122637	97044	61274
24	10	02	17	20	19	22	01A	09	13	12	01B		
54072	38637	33281	32894	27639	13693	12682	11750	11528	2450	1192	73		

Figure 17: Text representation of all values in FBICode column

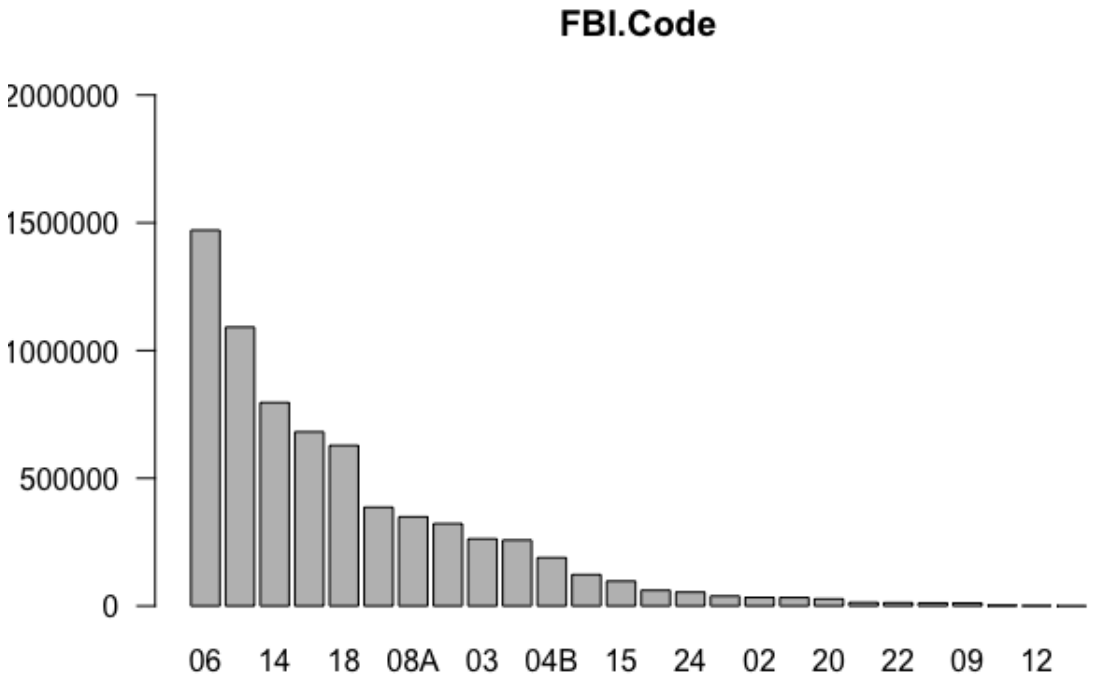

Figure 18: Graph representation of all values in FBICode column

4.2.10 IUCR

Graph presents top 30 IUCR codes seem to be in the groups of 3, investigate!

```
  0486   0820   0460   1320   1310   0810   0560   0610   0910   1811   0890   0860   2820   1330   0620   2027
597002 539979 456089 372149 366436 358255 309858 257828 257686 257507 248681 197721 134561 117891 105791 102957
  2825   031A   0320   2024   0430   143A   1150   051A   1506   041A   0870   0840   1130   4387
 97468  92846  89170  87402  64038  63316  59850  56545  42210  40354  38466  37858  37068  36212
```

Figure 19: Text of top 30 occurred values in IUCR column

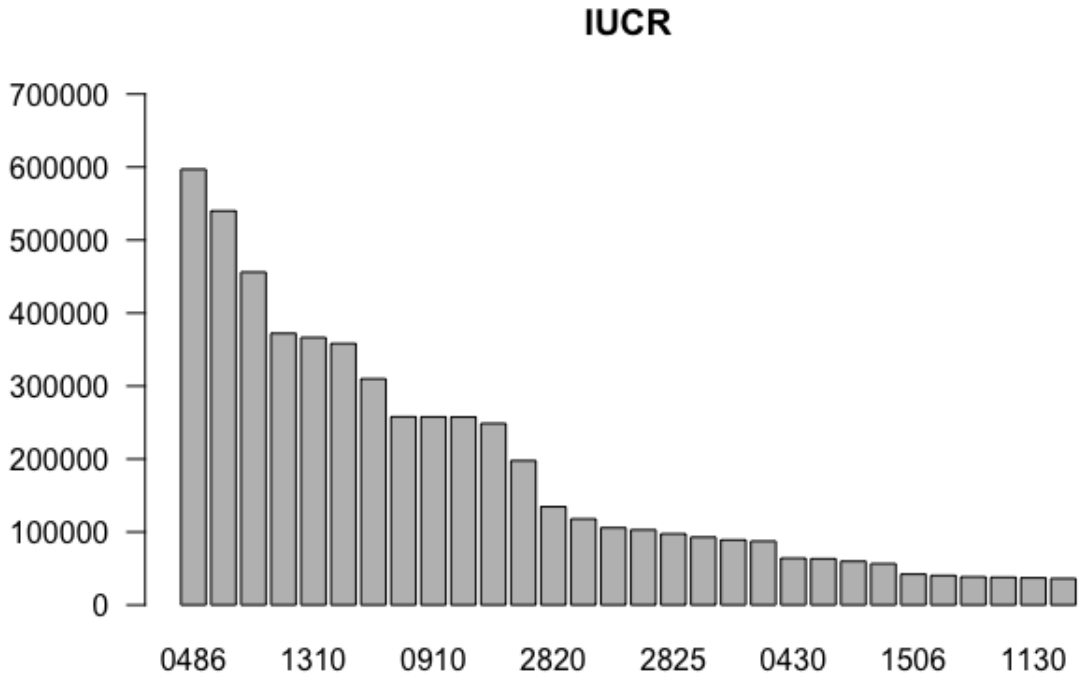

Figure 20: Graph of top 30 occurred values in IUCR column

4.2.11 Latitude

Graph presents top 30 First 3 LATs have higher offset, why? How are they correlated to LON?

```
41.976290414 41.754592961 41.883500187 41.897895128 41.909664252 41.896888586 41.788987036 41.885487535
       13612         9853         8048         4750         3425         3288         2951         2886
41.868180939  41.88233367 41.736259984 41.979006297 41.721627204 41.868541914 41.904192368 41.737094305
        2748         2627         2583         2578         2530         2464         2464         2373
41.891990384 42.019399237  41.68995741 41.739265865 41.882394062 41.899410159 41.736148121 41.750940757
        2366         2293         2242         2203         2188         2181         2134         2130
41.907153315 41.929743818 41.706070186 41.742710224 41.864493678 41.891694878
        2088         2041         2033         2014         2002         2001
```

Figure 21: Text of top 30 occurred values in LAT column

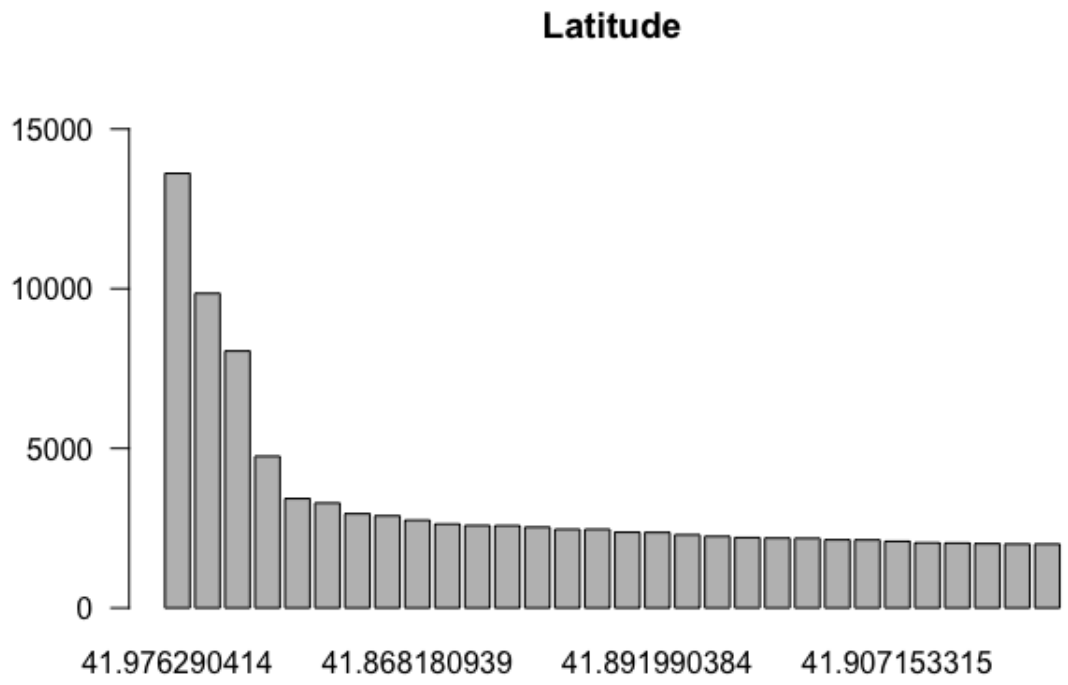

Figure 22: Graph of top 30 occurred values in LAT column

4.2.12 Longitude

Graph presents top 30 First 3 LONs have higher offset, why? How are they correlated to LAT?

-87.905227221	-87.741528537	-87.627876698	-87.624096605	-87.742728815	-87.628203192	-87.74147999	-87.726422045
13612	9853	8048	4754	3425	3288	2951	2886
-87.709271389	-87.627841791	-87.628068782	-87.906463155	-87.624485177	-87.647000785	-87.639235361	-87.572998178
2748	2627	2583	2578	2530	2464	2464	2373
-87.611461502	-87.675049485	-87.637460623	-87.604893749	-87.627844798	-87.624131266	-87.629070243	-87.625185222
2366	2293	2242	2203	2188	2181	2134	2130
-87.639680572	-87.684273777	-87.653645803	-87.634088181	-87.639158	-87.626155832		
2088	2041	2033	2014	2002	2001		

Figure 23: Text of top 30 occurred values in LON column

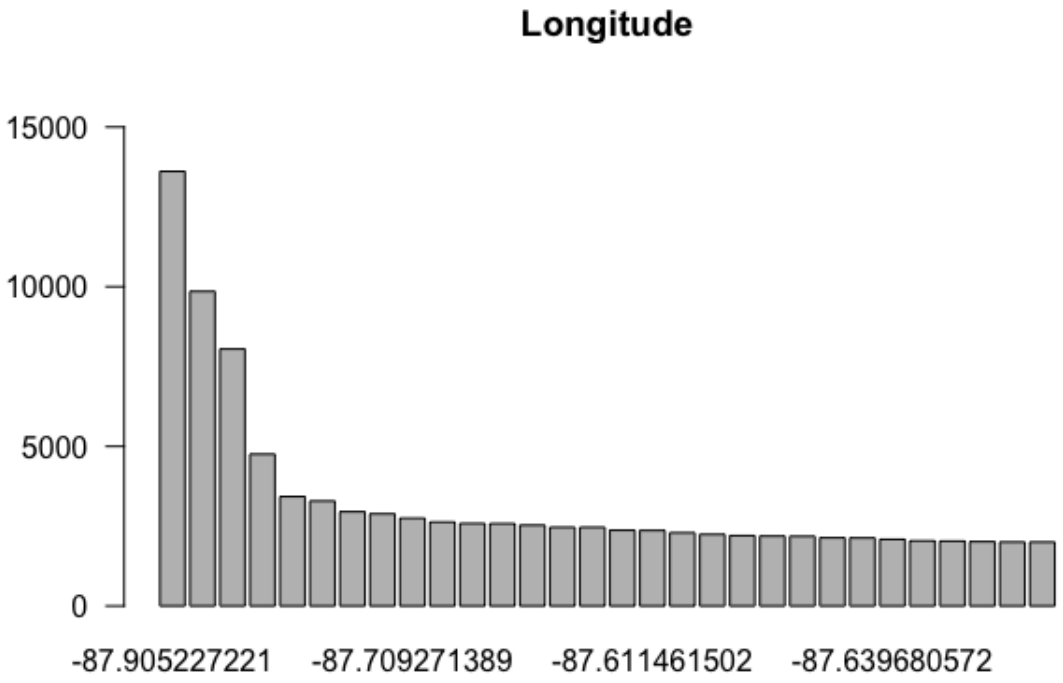

Figure 24: Graph of top 30 occurred values in LON column

4.2.13 Location description

Graph presents top 57 What are the top 3 locations and what crime does happen on them?

STREET	RESIDENCE	APARTMENT
1783788	1149049	807013
SIDEWALK	OTHER	PARKING LOT/GARAGE(NON.RESID.)
681400	239182	181757
ALLEY	SMALL RETAIL STORE	SCHOOL, PUBLIC, BUILDING
155563	132841	131386
RESIDENCE-GARAGE	VEHICLE NON-COMMERCIAL	RESTAURANT
122142	115048	114766
RESIDENCE PORCH/HALLWAY	DEPARTMENT STORE	GROCERY FOOD STORE
112275	87915	86124
GAS STATION	RESIDENTIAL YARD (FRONT/BACK)	PARK PROPERTY
78820	74800	53139
COMMERCIAL / BUSINESS OFFICE	CHA PARKING LOT/GROUNDS	BAR OR TAVERN
50597	44921	38823
CTA PLATFORM	DRUG STORE	CHA APARTMENT
34700	31340	30852
CTA TRAIN	SCHOOL, PUBLIC, GROUNDS	BANK
28309	27716	27488
HOTEL/MOTEL	CTA BUS	VACANT LOT/LAND
26667	22972	22856
CONVENIENCE STORE	DRIVEWAY - RESIDENTIAL	CHA HALLWAY/STAIRWELL/ELEVATOR
22488	20665	20374
HOSPITAL BUILDING/GROUNDS	TAVERN/LIQUOR STORE	PARKING LOT / GARAGE (NON RESIDENTIAL)
20124	18929	17661
POLICE FACILITY/VEH PARKING LOT	CHURCH/SYNAGOGUE/PLACE OF WORSHIP	AIRPORT/AIRCRAFT
16759	14132	13157
NURSING HOME/RETIREMENT HOME	GOVERNMENT BUILDING/PROPERTY	CONSTRUCTION SITE
13149	12970	12664
SCHOOL, PRIVATE, BUILDING	ABANDONED BUILDING	CURRENCY EXCHANGE
12614	11055	10563
OTHER (SPECIFY)	CTA GARAGE / OTHER PROPERTY	ATHLETIC CLUB
10497	10223	8916
WAREHOUSE	RESIDENCE - PORCH / HALLWAY	ATM (AUTOMATIC TELLER MACHINE)
8884	8333	8257
BARBERSHOP	CTA BUS STOP	RESIDENCE - YARD (FRONT / BACK)
7824	7668	7638
TAXICAB	RESIDENCE - GARAGE	MEDICAL/DENTAL OFFICE
7039	6840	6620

Figure 25: Text of top 57 occurred values in LocationDescription column

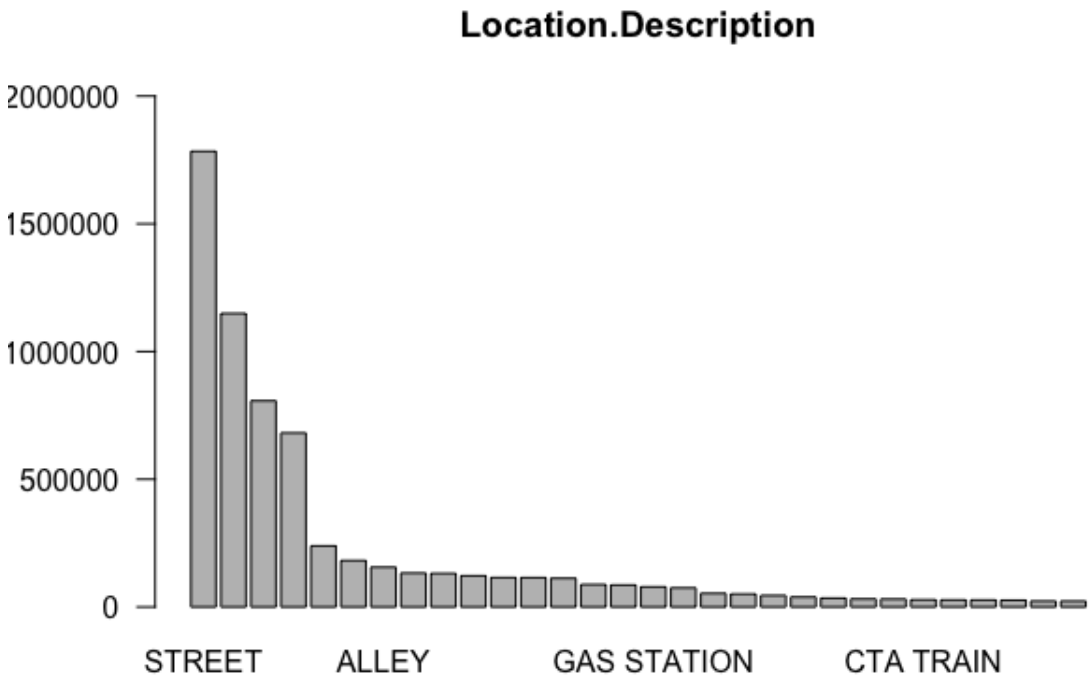

Figure 26: Graph of top 57 occurred values in LocationDescription column

4.2.14 PrimaryType

Graph presents top 30 Why are top 2 crime types this?

THEFT	BATTERY	CRIMINAL DAMAGE
1471191	1278551	796297
NARCOTICS	ASSAULT	OTHER OFFENSE
666586	454700	432483
BURGLARY	MOTOR VEHICLE THEFT	DECEPTIVE PRACTICE
386299	322196	299993
ROBBERY	CRIMINAL TRESPASS	WEAPONS VIOLATION
262739	193526	96060
PROSTITUTION	OFFENSE INVOLVING CHILDREN	PUBLIC PEACE VIOLATION
61213	48459	48303
SEX OFFENSE	CRIM SEXUAL ASSAULT	INTERFERENCE WITH PUBLIC OFFICER
25701	24196	17552
GAMBLING	LIQUOR LAW VIOLATION	HOMICIDE
13403	12682	11823
ARSON	KIDNAPPING	CRIMINAL SEXUAL ASSAULT
11635	5950	5054
STALKING	INTIMIDATION	CONCEALED CARRY LICENSE VIOLATION
4179	4117	971
OBSCENITY	PUBLIC INDECENCY	OTHER NARCOTIC VIOLATION
727	180	133

Figure 27: Text of top 30 occurred values in PrimaryType column

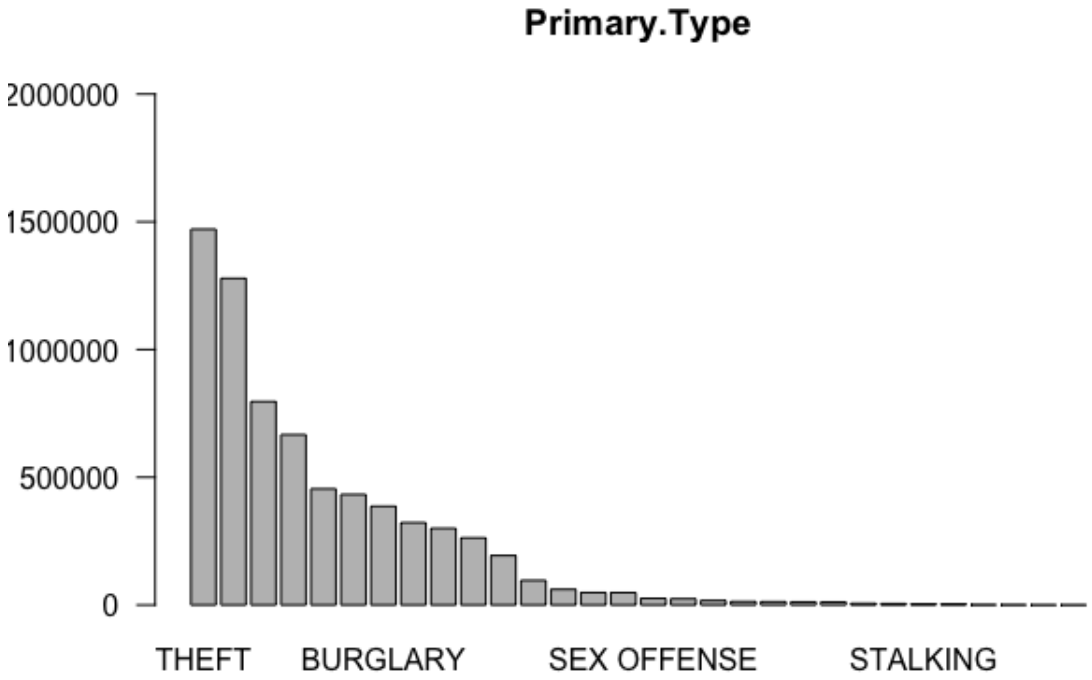

Figure 28: Graph of top 30 occurred values in PrimaryType column

4.2.15 UpdatedOn

Graph presents top 30 Why are there so many updates on that time? Is it an error? Did something happen?

2018-02-10 15:50:01	2018-02-28 15:56:25	2016-02-04 06:33:39	2022-10-08 16:45:51	2018-05-04 15:51:04
2760472	2667209	345154	112910	84250
2022-10-09 16:44:48	2020-12-17 15:44:58	2019-06-30 15:56:27	2017-02-14 15:49:42	2015-08-17 15:03:40
57899	31132	25051	18121	10614
2019-07-19 16:09:50	2022-09-18 16:45:51	2019-01-18 09:37:14	2019-01-10 15:16:50	2021-01-16 15:49:23
8779	8507	5118	4635	3782
2018-10-26 16:01:05	2018-02-09 15:44:29	2022-08-31 16:51:30	2019-01-23 11:26:56	2020-03-18 15:52:17
3492	3195	3181	2683	2161
2021-06-25 17:06:33	2018-10-25 16:10:05	2021-02-19 15:45:56	2020-06-07 15:47:20	2006-03-31 22:03:38
1828	1597	1389	1151	809
2018-08-25 15:59:23	2018-11-07 16:21:48	2018-07-01 15:55:31	2018-06-03 15:52:24	2018-06-27 16:15:34
776	760	744	742	742

Figure 29: Text of top 30 occurred values in UpdatedOn column

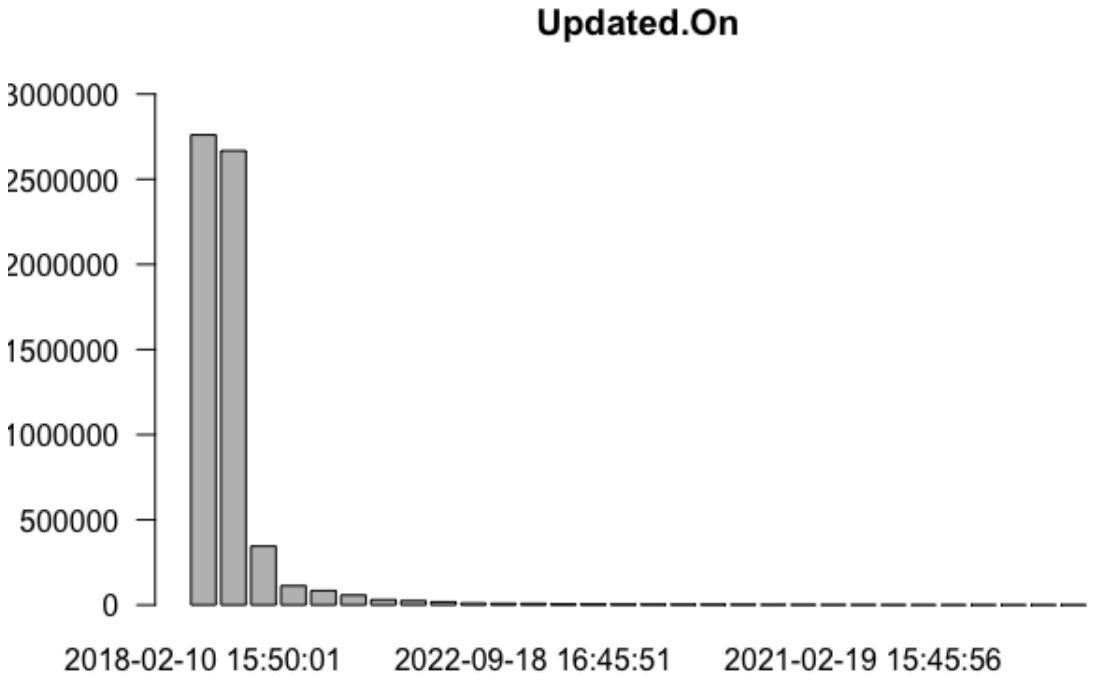

Figure 30: Graph of top 30 occurred values in UpdatedOn column

4.2.16 Ward

Ward description

```
    28     42     24     27      2     17      6     20      3     21     16     34      8      7     37     15
319325 287923 273163 260234 249838 231276 230132 228910 212515 210968 200723 195734 192119 188679 188541 182028
     9      5     29      4      1     10     26     18     32     25     49     44     31     11     30     14
179916 166646 161877 136661 133126 118438 117867 115121 114971 100268  99449  96836  95464  94922  94895  93573
    46     35     23     43     13     12     22     40     36     38     41     47     50     45     48     33
 93337  90945  88739  87219  85853  84913  81263  73372  73217  72667  72657  72021  71733  71051  69996  68934
    39     19
 65217  61744
```

Figure 31: Text representation of all values in Ward column

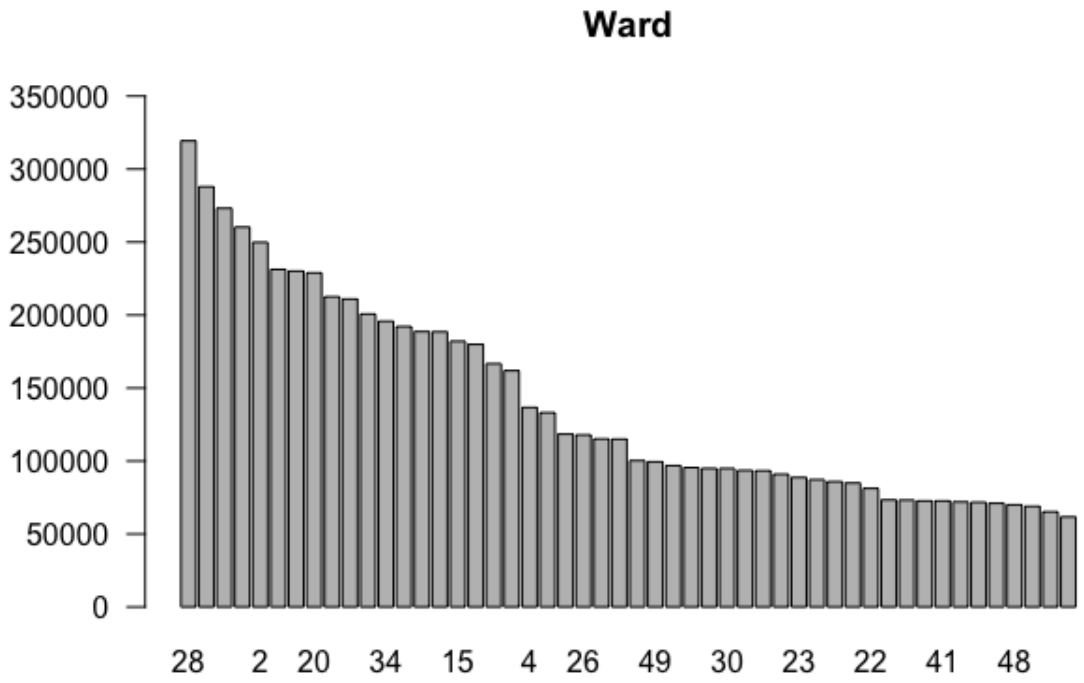

Figure 32: Graph representation of all values in Ward column

5 Correlation Analysis

This is the Correlation Analysis section

6 Regression Analysis

This is the Regression Analysis section

7 Results and Discussion

This is the Results and Discussion section

7.1 Subsection

Conduct the statistical analysis and report the results. Show how the statistical assumptions required for applying a statistical test are fulfilled. Discuss and evaluate the results to effectively answer the questions outlined earlier;
o Detail the results in tables and figures as appropriate. Each table and figure must be numbered, and must have appropriate legends (if required) and captions. The captions must be a set of grammatically complete sentences (up to 3 lines long) that concisely describe the information in the tables and the figures. No such table and figure should be included that is not discussed in the text. The numerical comparisons should be supported by appropriate tests of statistical significance.

8 Conclusions

This is the conclusions section.

8.1 Subsection

Summarize the key findings of your report so that if someone does not have the time to go through the whole report, he/she can still understand the important results.

9 References

This is the references section

9.1 Subsection