



Mined
Your
Business



Kelly Appleton

Jake Lombardo

Chris Pegoli



Agenda

- ❖ Problem Description
 - Company and Client Information
 - Our Task
- ❖ Data Exploration
 - Data Set
- ❖ Data Preprocessing
- ❖ Data Mining
 - Techniques / Algorithms / Models
- ❖ Algorithm / Model
 - Performance & Results
- ❖ Conclusions
 - Lessons Learned





Problem Description



BREAKING NEWS TEAM COVERAGE

TWO OFFICERS SHOT; ONE KILLED
» ROCHESTER

BREAKING NEWS

16-YEAR-OLD VICTIM OF 14TH HOMICIDE THIS YEAR
CLIFFORD AVE AND ARBUTUS ST, ROCHESTER

BREAKING NEWS

MAN SERIOUSLY HURT IN SHOOTING
HUDSON AVE, ROCHESTER

CONTINUING COVERAGE

3 KILLED, 4 INJURED IN DRIVE-BY
ROCHESTER

#13WHAM
5:35 55°

RIGHT NOW

MAN SHOT OUTSIDE DELI; NO SUSPECTS IN CUSTODY
NORTH GOODMAN STREET & SHORT STREET, ROCHESTER

BREAKING NEWS

MAN SHOT OVERNIGHT
JEFFERSON AVE, ROCHESTER

MANHUNT UNDERWAY AFTER SHOOTING
AT A BACKYARD PARTY IN ROCHESTER, NY

FOX NEWS ALERT

NY GOV ANDREW CUOMO ANNOUNCES STATE WILL HONOR LIFE AND

BREAKING NEWS

RPD NAME VICTIMS IN DOUBLE HOMICIDE
DRIVING PARK AVENUE, ROCHESTER

NEW AT 11

MURDER ARREST

WABASH STREET, ROCHESTER - AUGUST 4, 2020

BREAKING NEWS

MAN SHOT, TAKEN TO HOSPITAL
LYELL AVE AND WHITNEY ST, ROCHESTER

BREAKING NEWS

MASS SHOOTING LEAVES ONE DEAD, THREE INJURED
HUDSON AVE & RIDGE RD, ROCHESTER

BREAKING OVERNIGHT

TRIPLE SHOOTING INVESTIGATION UNDERWAY
DEWEY AVE, ROCHESTER



Problem Description

- ❖ About our company
- ❖ Our Client
 - Rochester Police Department (RPD)
 - Scruff McGruff, police chief
- ❖ Rochester, NY is riddled with crime
 - Focus on shootings and homicides
- ❖ Our Task:




































- Analyze data & find trends to help the RPD & McGruff reduce crime in Rochester
- Findings will be used for:
 - Decision-making on spending
 - Resource allocation
 - Management of personnel
 - Patrolling
 - And more!





Data Exploration

❖ RPD Open Data Portal

Data	Start Here	Featured Maps	Explore and Download	Codebook
 Crime  Violent and Property Crime		 Public Crime Map  Neighborhood History Map		
 Homicides  Murder and Non-Negligent Manslaughter		 Incident Locations		
 Shootings  Fatal and Non-Fatal Shootings		 No Map		
Sex Offenders  Level 2 & 3 Offenders from NYS Registry		 Sex Offenders by Location		
 Boundaries  Geographic Outlines of Police Boundaries		 Sections and Patrol Beats		
Personnel  Sworn and Civilian Roster		 No Map		



Crime

- 2011 - Present
- 37 attributes

Statute Attempted	Text
Weapon Description	Text
Larceny Type	Text
Location Type	Text
Geo Beat	Text
Geo Section	Text
Geo Section Num	Number

Data Exploration

Geocode Address	Text	Address StreetFull	Text
Geocode Street	Text	Address City	Text
Case Number	Text	Address State	Text
Occurred From Date Year	Number	Patrol Beat	Text
Occurred From Date Month	Number	Patrol Section	Text
Occurred From Time	Text	Case Status	Text
Occurred From Timestamp	Date or Time	Statute Title	Text
Occurred Through Date Year	Number	Statute Section	Text
Occurred Through Date Month	Number	Statute Subsection	Text
Occurred Through Time	Text	Statute Degree	Text
Occurred Through Timestamp	Date or Time	Statute Class	Text
Reported Date Year	Number	Statute Category	Text
Reported Date Month	Number	Statute Text	Text
Reported Time	Text	Statute Description	Text
Reported Timestamp	Date or Time	Statute Crime Category	Number

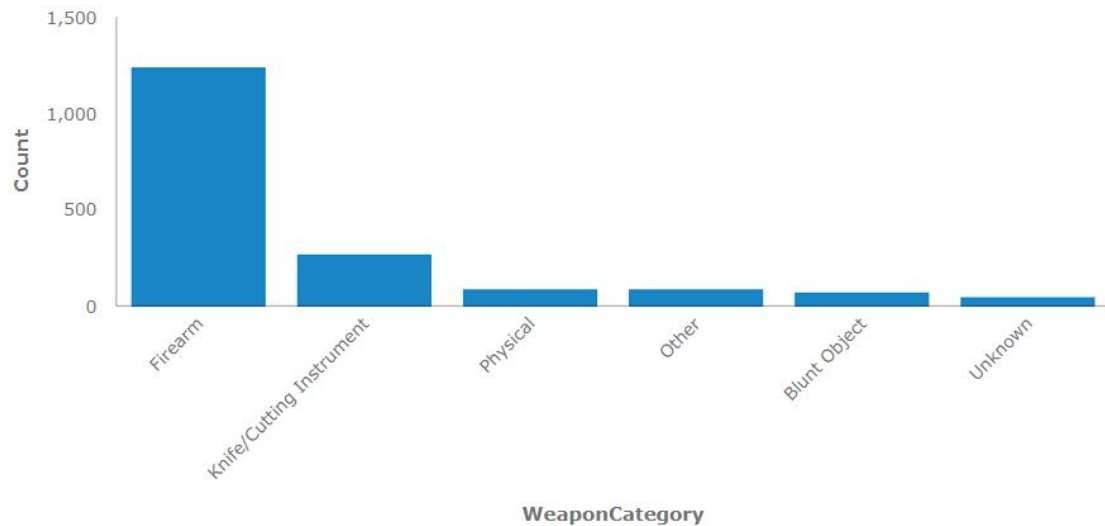


Data Exploration

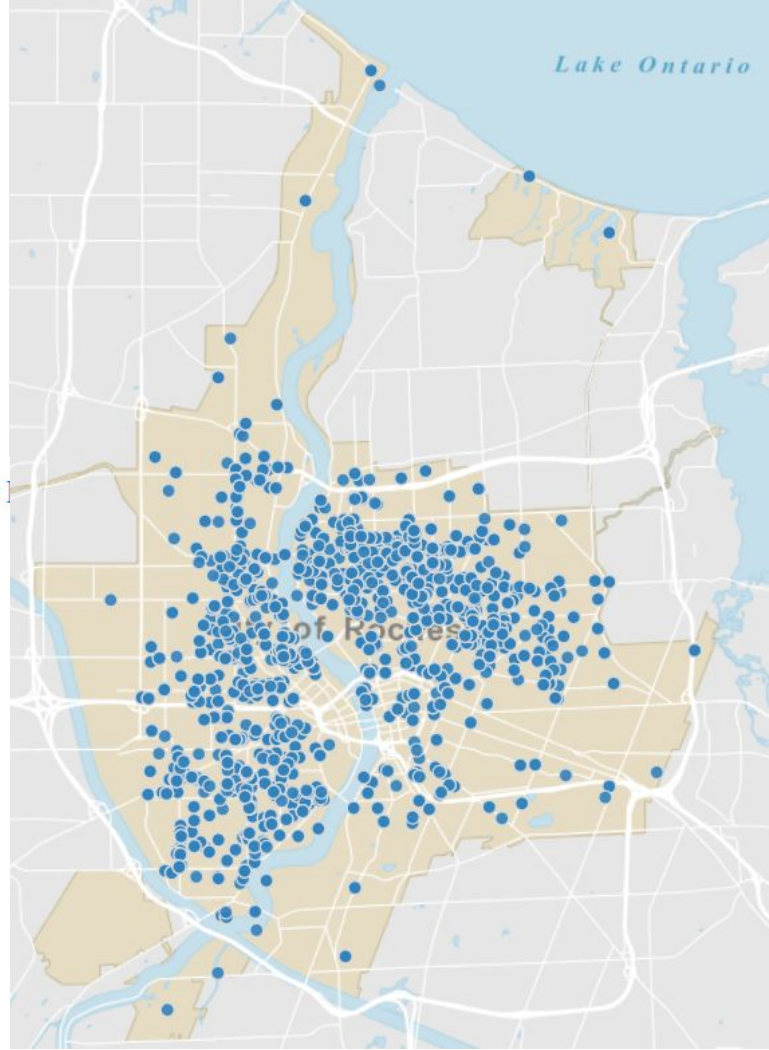
❖ Homicides

- 2000 - Present
- 29 attributes

Count of WeaponCategory



[Show Table] [Close]



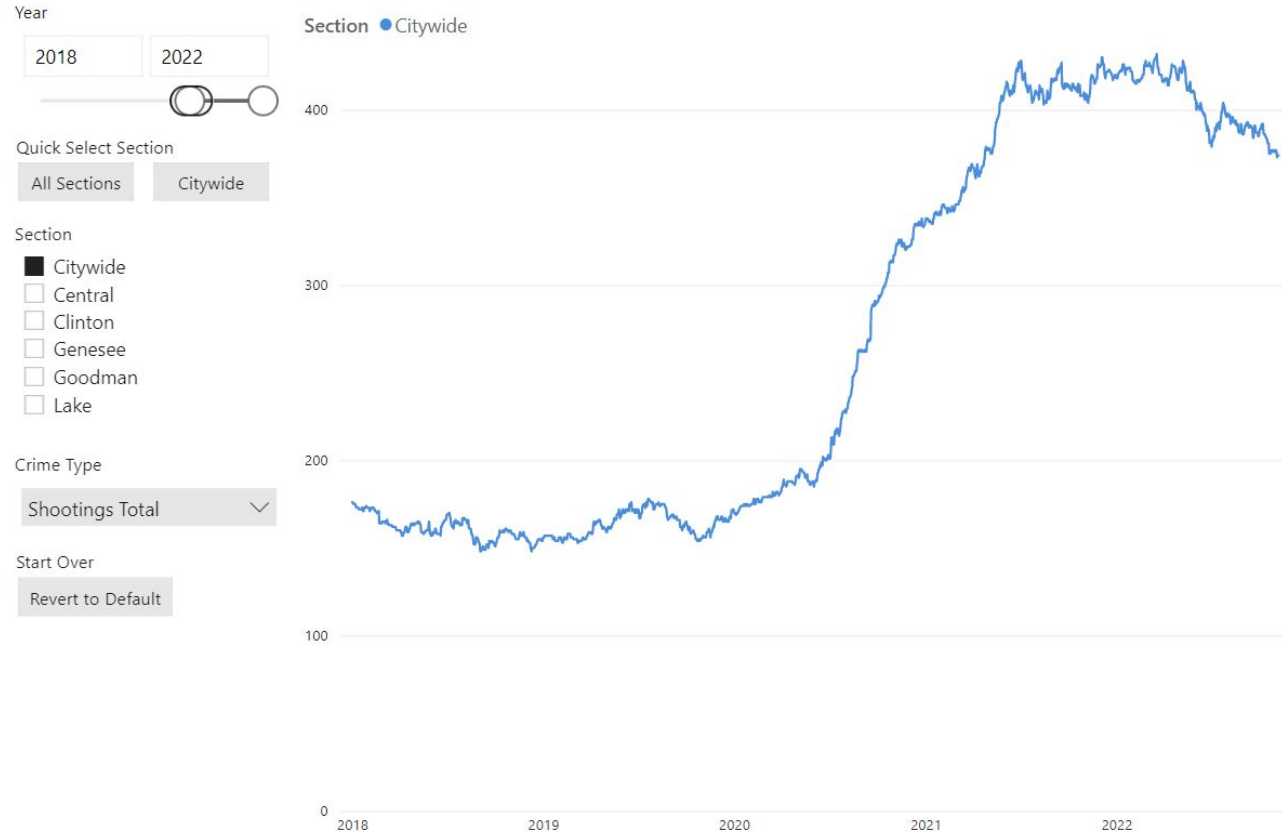


Data Exploration



Shootings

- 2012 - Present
- 15 attributes





Yearly
Statistics

- 2022
- 2021
- 2020
- 2019
- 2018
- 2017
- 2016
- 2015
- 2014
- 2013
- 2012

Rolling 365 Day
Chart

Incidents

275

with multiple victims

30

Multiple Victim Rate

10.9%

Victims

315

fatal shootings

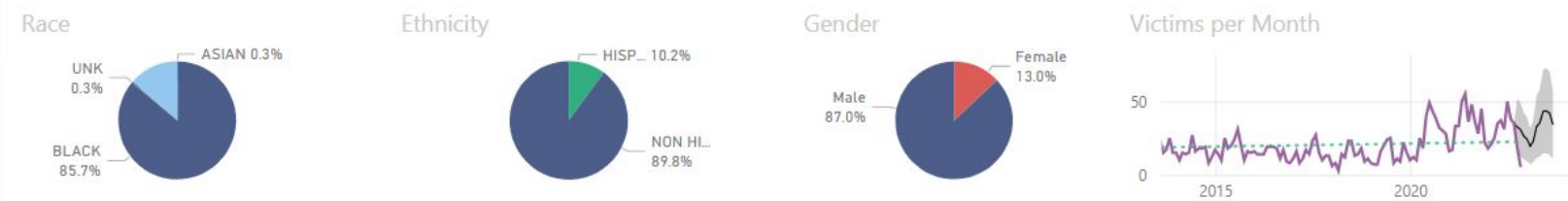
57

Fatality Rate

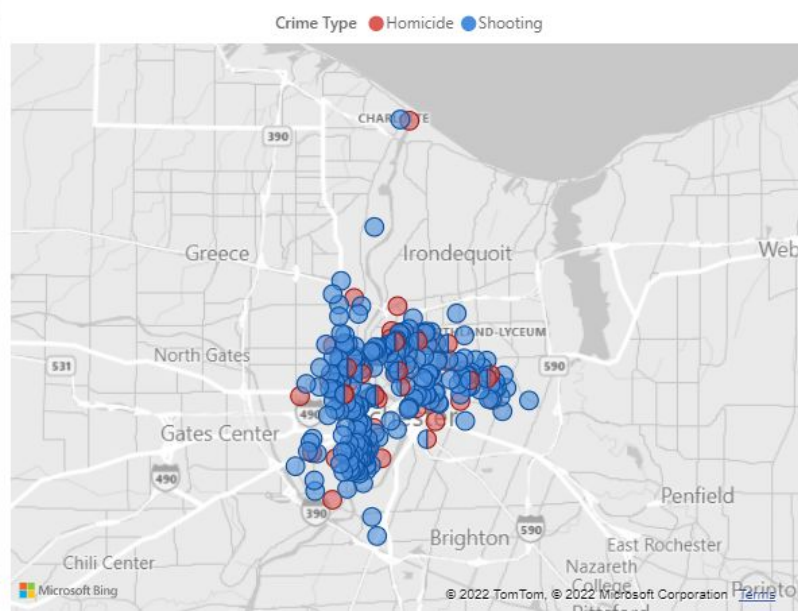
18.1%

Median Age

29



Occurred Date	Case Number	Address	Age	Multiple Victims	Crime Type
2022-11-11	22-236708	479 Bay St	22	No	Shooting
2022-11-10	22-236531	69 Myrtle St	45	No	Shooting
2022-11-05	22-232641	445 Flint St	29	No	Shooting
2022-11-05	22-233345	175 Selye Ter	4	No	Shooting
2022-11-02	22-230405	95 Glenwood Ave	24	No	Shooting
2022-11-01	22-229312	99 Weld St	17	No	Homicide
2022-10-31	22-228495	64 Sullivan St	37	No	Shooting
2022-10-31	22-229013	223 Genesee St	15	No	Shooting
2022-10-28	22-226158	444 Central Ave	20	No	Shooting
2022-10-25	22-224108	70 Reynolds St	20	No	Shooting
2022-10-24	22-223769	271 Flint St	23	No	Shooting
2022-10-23	22-222821	73 Weld St	21	No	Shooting
2022-10-23	22-222891	368 Verona St	26	No	Shooting
2022-10-21	22-221361	470 Lyell Ave	36	No	Shooting
2022-10-10	22-212577	27 Oregon St	37	No	Shooting
2022-10-07	22-211191	1060 E Main St	30	No	Homicide
2022-10-06	22-210462	25 Austin St	38	No	Shooting
2022-10-05	22-209350	127 Emerson St	15	No	Shooting

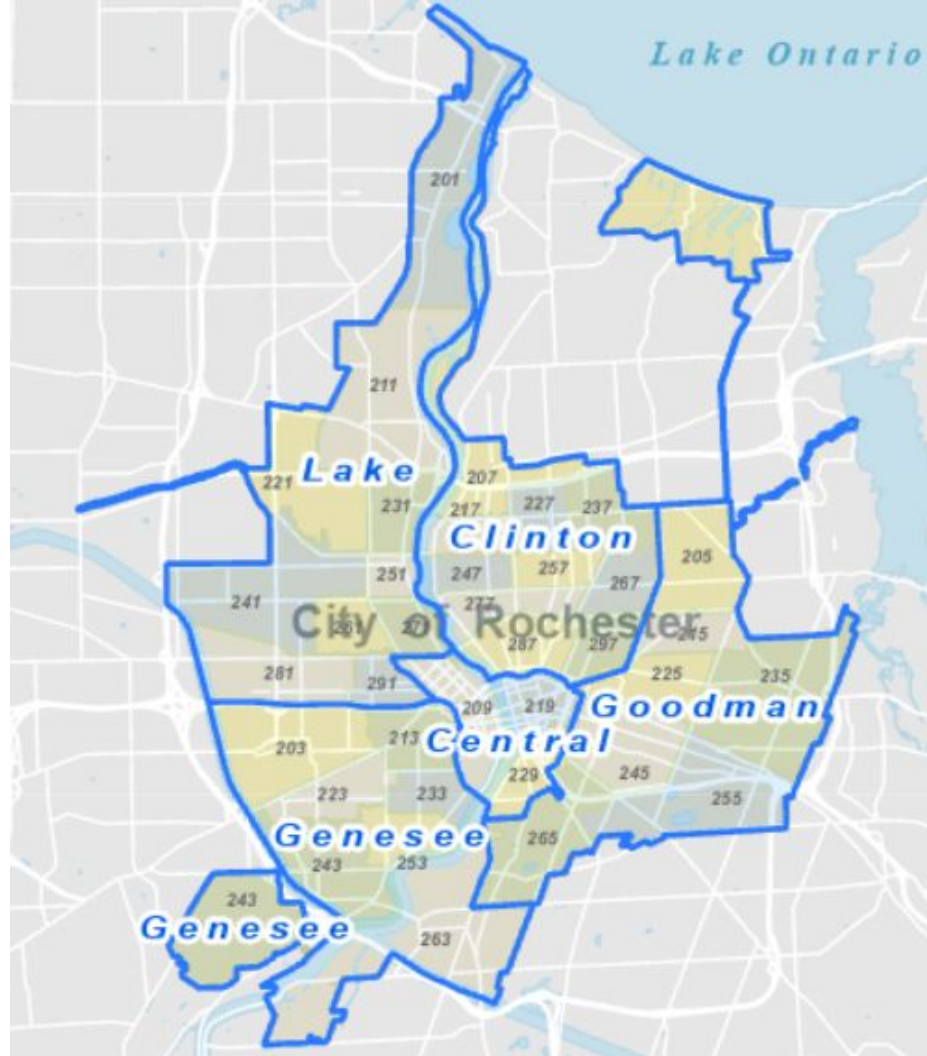




Data Exploration

❖ Boundaries

- Central
- Clinton
- Genesee
- Goodman
- Lake





Data Preprocessing

❖ What data do we want to analyze?

- Data set selection
- Attribute selection
- Year range
 - 2018 - Oct 12, 2022

❖ Eliminate attributes

- Duplicated
- Redundant
- Inaccurate

❖ Ignore values

- "Unknown"

❖ Fixed Formatting Errors

- PersonAgeBand 5-14, May 14 in Excel

❖ Discretization

- Day of Week
- Season
- Gun?

❖ No missing values or outliers



Data Preprocessing

- ❖ Crime
 - OccurredFrom_Date_Year
 - OccurredFrom_Date_Month
 - DayOfWeek
 - Season
 - Statute_Text
 - Section
 - Gun?

- ❖ Homicides
 - OccurredMonth
 - OccurredYear
 - DayOfWeek
 - Section
 - WeaponCategory
 - Season

- ❖ Shootings
 - DayOfWeek
 - Occurred_Month
 - Season
 - Occurred_Year
 - Multiple_Shooting
 - Victim_Age
 - Victim_Age_Band
 - Homicide?

* Even after trimming the attributes, ignored attributes like OccurredYear when Clustering in Weka



Data Mining

Techniques / Algorithms / Models

- ❖ Count / Frequency
- ❖ Visualization
 - Attribute X vs Attribute Y
 - Bar graphs
- ❖ K-means Clustering
 - Attribute selection
 - What to ignore
 - K value



Data Mining Crime

❖ Instances: 40547

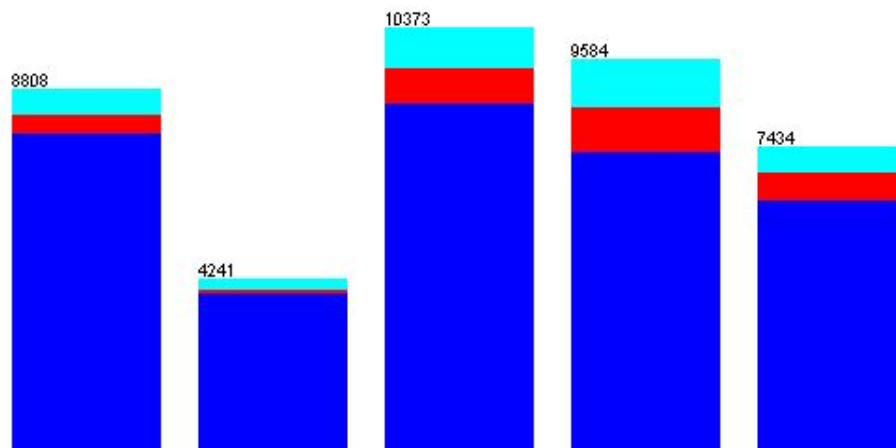
❖ Gun?	Count	Freq.
➤ Yes	3215	7.9%
➤ No	3735	9.2%
➤ Unknown	33597	82.9%



- ❖ Aqua = No Gun
- ❖ Red = Gun
- ❖ Blue = Unknown
- ❖ Most crimes in Lake
 - Runner-up: Clinton
- ❖ More crimes (reported) involving a gun in Clinton

Selected attribute			
Name: Section		Type: Nominal	
Missing: 107 (0%)		Distinct: 5	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Goodman	8808	8808
2	Central	4241	4241
3	Lake	10373	10373
4	Clinton	9584	9584
5	Genesee	7434	7434

Class: Gun? (Nom) Visualize All

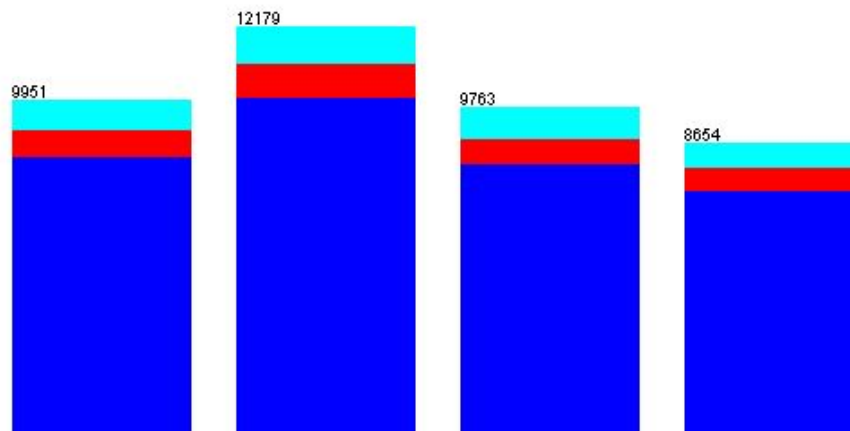




- ❖ Aqua = No Gun
- ❖ Red = Gun
- ❖ Blue = Unknown
- ❖ More crime in Summer
- ❖ More crime in August
 - Not represented here

Selected attribute			
Name: Season		Type: Nominal	
Missing: 0 (0%)		Distinct: 4	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Autumn	9951	9951
2	Summer	12179	12179
3	Spring	9763	9763
4	Winter	8654	8654

Class: Gun? (Nom) ▼ Visualize All



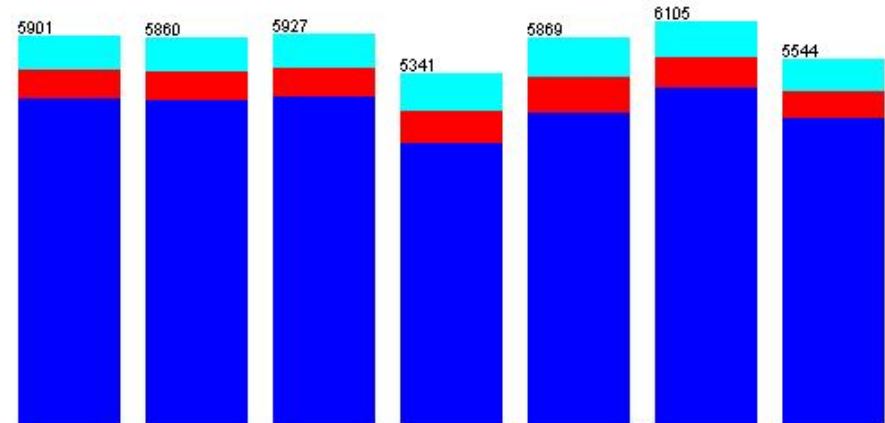


- ❖ Aqua = No Gun
- ❖ Red = Gun
- ❖ Blue = Unknown
- ❖ Most crimes occur on Fridays
- ❖ All days are very close in numbers
- ❖ The least amount of crime occurs on Sundays
 - Interesting because the second most shootings occur on Sunday and the most homicides occur on Sunday

Selected attribute			
Name: DayOfWeek		Type: Nominal	
Missing: 0 (0%)		Distinct: 7	Unique: 0 (0%)

No.	Label	Count	Weight
1	Wed	5901	5901
2	Tue	5860	5860
3	Mon	5927	5927
4	Sun	5341	5341
5	Sat	5869	5869
6	Fri	6105	6105
7	Thu	5544	5544

Class: Gun? (Nom)	Visualize All
-------------------	---------------





Data Mining

Shootings

❖ Instances: 1170

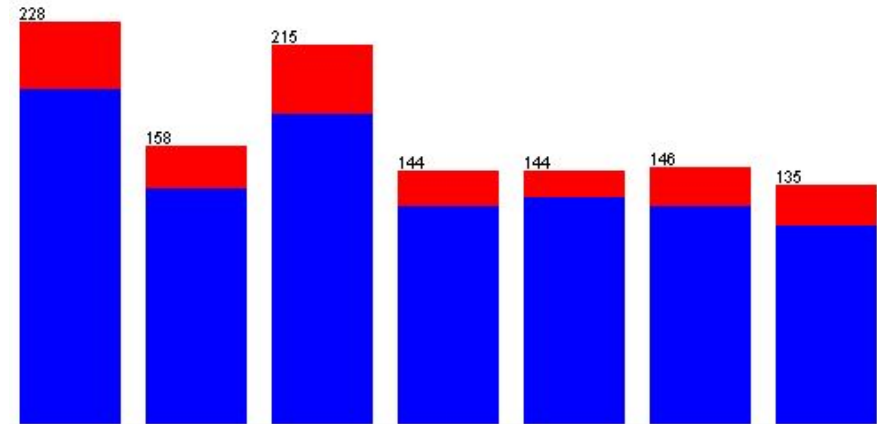
❖ Homicide?	Count	Freq.
➤ Yes	181	15.5%
➤ No	989	84.5%



- ❖ Red = Homicide (death)
- ❖ Blue = Not Homicide
- ❖ Most shootings, regardless of death, occur on Saturday
 - Runner-up: Sunday
- ❖ More shootings in June, followed by July
 - Not represented here

Selected attribute			
Name: DayOfWeek		Type: Nominal	
Missing: 0 (0%)		Distinct: 7	Unique: 0 (0%)
No.	Label	Count	Weight
1	Sat	228	228
2	Wed	158	158
3	Sun	215	215
4	Mon	144	144
5	Tue	144	144
6	Fri	146	146
7	Thu	135	135

Class: Homicide? (Nom) Visualize All

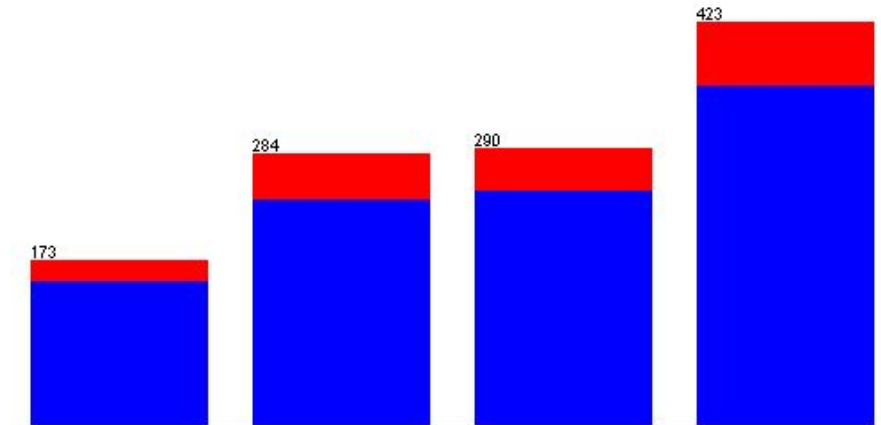




- ❖ Red = Homicide (death)
- ❖ Blue = Not Homicide
- ❖ Most shootings occur in the Summer
- ❖ Section was not available in Shootings dataset
 - Addresses provided
 - Latitude & Longitude

Selected attribute			
Name: Season		Type: Nominal	
Missing: 0 (0%)		Distinct: 4	Unique: 0 (0%)
No.	Label	Count	Weight
1	Winter	173	173
2	Autumn	284	284
3	Spring	290	290
4	Summer	423	423

Class: Homicide? (Nom) Visualize All





Data Mining

Homicides

❖ Instances: 244

❖ Weapon Category

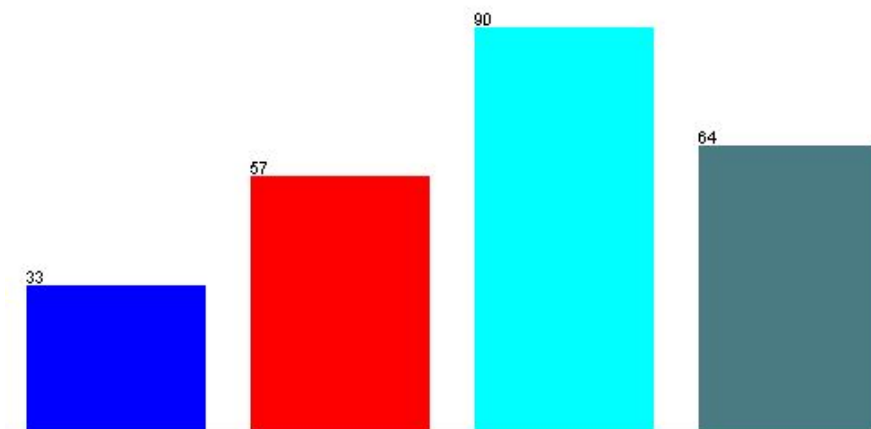
	Count	Freq.
➤ Firearm	188	77.0%
➤ Knife / Cutting Instrument	32	13.1%
➤ Other	24	9.8%



- ❖ Most homicides occur in the Summer
- ❖ Most homicides occur in June
 - Not represented here

Selected attribute			
Name: Season		Type: Nominal	
Missing: 0 (0%)		Distinct: 4	Unique: 0 (0%)
No.	Label	Count	Weight
1	Winter	33	33
2	Spring	57	57
3	Summer	90	90
4	Autumn	64	64

Class: Season (Nom) Visualize All

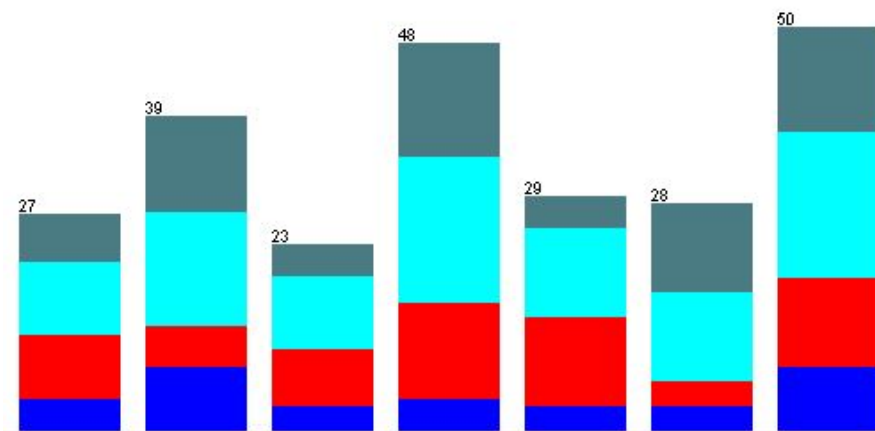




- ❖ Green(ish) = Autumn
- ❖ Aqua = Summer
- ❖ Red = Spring
- ❖ Blue = Winter
- ❖ Most homicides occur on Sundays
 - Runner-up (very close): Saturdays

Selected attribute			
Name: DayOfWeek		Type: Nominal	
Missing: 0 (0%)		Distinct: 7	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Mon	27	27
2	Fri	39	39
3	Tue	23	23
4	Sat	48	48
5	Wed	29	29
6	Thu	28	28
7	Sun	50	50

Class: Season (Nom) Visualize All



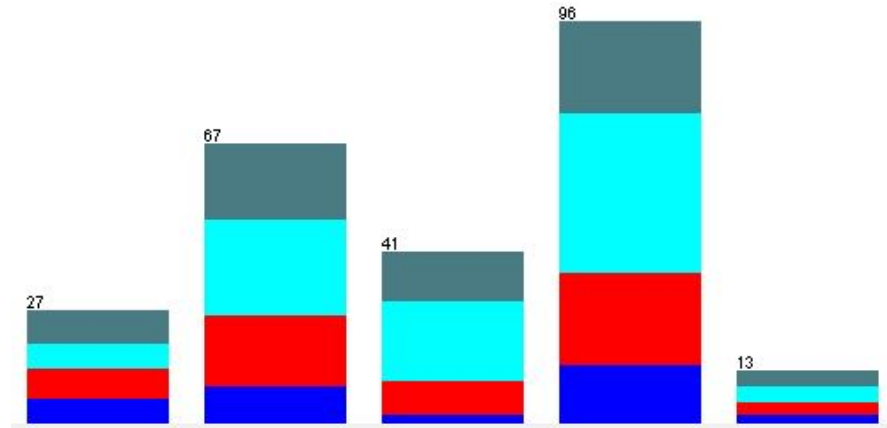


- ❖ Green(ish) = Autumn
- ❖ Aqua = Summer
- ❖ Red = Spring
- ❖ Blue = Winter
- ❖ Most homicides occur in Clinton



Selected attribute			
Name: Section		Type: Nominal	
Missing: 0 (0%)		Distinct: 5	Unique: 0 (0%)
No.	Label	Count	Weight
1	Goodman	27	27
2	Lake	67	67
3	Genesee	41	41
4	Clinton	96	96
5	Central	13	13

Class: Season (Nom) Visualize All





Algorithm / Model Performance & Results

❖ Clustering

- K Simple Means
 - $K = 2, 3, 5, 7$

❖ Crime

- Ignored Year
- Removed “unknown” from Gun?

❖ Attributes

- 6
- Month, Day of Week, Season, Statute, Section, Gun?



Algorithm / Model

Performance & Results

❖ Crime, K = 2

Number of iterations: 4
Within cluster sum of squared errors: 17953.218462718156

Initial starting points (random):

Cluster 0: 2,Thu,Winter,Larceny,Goodman,No

Cluster 1: 5,Sat,Spring,Robbery,Lake,Yes

Final cluster centroids:

Attribute	Cluster#		
	Full Data	0	1
	(6950.0)	(3457.0)	(3493.0)
=====			
OccurredFrom_Date_Month	6.5931	5.4839	7.6908
DayOfWeek	Sat	Thu	Sat
Season	Summer	Spring	Summer
Statute_Text	Aggravated Assault	Aggravated Assault	Aggravated Assault
Section	Clinton	Clinton	Lake
Gun?	No	No	Yes

Clustered Instances

0	3457 (50%)
1	3493 (50%)



Algorithm / Model Performance & Results

❖ Crime, K = 3

Number of iterations: 4
Within cluster sum of squared errors: 16565.834243917765

Initial starting points (random):

Cluster 0: 2,Thu,Winter,Larceny,Goodman,No

Cluster 1: 5,Sat,Spring,Robbery,Lake,Yes

Cluster 2: 6,Sun,Summer,'Aggravated Assault',Genesee,No

Final cluster centroids:

Attribute	Full Data (6950.0)	Cluster#		
		0 (1741.0)	1 (2278.0)	2 (2931.0)
OccurredFrom_Date_Month	6.5931	4.6295	6.4899	7.8396
DayOfWeek	Sat	Thu	Sat	Sun
Season	Summer	Winter	Spring	Summer
Statute_Text	Aggravated Assault	Aggravated Assault	Robbery	Aggravated Assault
Section	Clinton	Goodman	Lake	Clinton
Gun?	No	No	Yes	No

Clustered Instances

0	1741 (25%)
1	2278 (33%)
2	2931 (42%)



Algorithm / Model

Performance & Results

❖ Crime, K = 5

Number of iterations: 6

Within cluster sum of squared errors: 14762.427893485758

Initial starting points (random):

Cluster 0: 2,Thu,Winter,Larceny,Goodman,No

Cluster 1: 5,Sat,Spring,Robbery,Lake,Yes

Cluster 2: 6,Sun,Summer,'Aggravated Assault',Genesee,No

Cluster 3: 12,Sat,Winter,Robbery,Lake,Yes

Cluster 4: 2,Sat,Winter,'Aggravated Assault',Clinton,Yes

Clustered Instances

0	1367 (20%)
1	1204 (17%)
2	1475 (21%)
3	1252 (18%)
4	1652 (24%)

Final cluster centroids:

Attribute	Full Data (6950.0)	Cluster# 0 (1367.0)	1 (1204.0)	2 (1475.0)	3 (1252.0)	4 (1652.0)
OccurredFrom_Date_Month	6.5931	4.8149	3.3978	7.6983	10.0256	6.8051
DayOfWeek	Sat	Thu	Sat	Sun	Sat	Sat
Season	Summer	Winter	Spring	Summer	Autumn	Summer
Statute_Text	Aggravated Assault	Aggravated Assault	Robbery	Aggravated Assault	Robbery	Aggravated Assault
Section	Clinton	Goodman	Lake	Genesee	Lake	Clinton
Gun?	No	No	Yes	No	Yes	Yes



Algorithm / Model

Performance & Results

❖ Crime, K = 7

Number of iterations: 4
Within cluster sum of squared errors: 13472.33224127969

Initial starting points (random):

Cluster 0: 2,Thu,Winter,Larceny,Goodman,No
Cluster 1: 5,Sat,Spring,Robbery,Lake,Yes
Cluster 2: 6,Sun,Summer,'Aggravated Assault',Genesee,No
Cluster 3: 12,Sat,Winter,Robbery,Lake,Yes
Cluster 4: 2,Sat,Winter,'Aggravated Assault',Clinton,Yes
Cluster 5: 1,Sat,Winter,'Aggravated Assault',Lake,No
Cluster 6: 6,Fri,Summer,'Aggravated Assault',Lake,Yes

Clustered Instances

0	584 (8%)
1	763 (11%)
2	1790 (26%)
3	893 (13%)
4	804 (12%)
5	1143 (16%)
6	973 (14%)

Final cluster centroids:

Attribute	Full Data (6950.0)	Cluster# 0 (584.0)	1 (763.0)	2 (1790.0)	3 (893.0)	4 (804.0)	5 (1143.0)	6 (973.0)
OccurredFrom_Date_Month	6.5931	5.9606	3.557	7.3849	10.0269	6.1704	4.4917	7.5632
DayOfWeek	Sat	Thu	Sat	Sun	Sat	Sat	Sat	Fri
Season	Summer	Winter	Spring	Summer	Autumn	Winter	Winter	Summer
Statute_Text	Aggravated Assault	Robbery	Robbery	Aggravated Assault	Robbery	Aggravated Assault	Aggravated Assault	Aggravated Assault
Section	Clinton	Goodman	Lake	Clinton	Lake	Clinton	Lake	Lake
Gun?	No	No	Yes	No	Yes	Yes	No	Yes



Algorithm / Model Performance & Results

❖ Clustering

- K Simple Means
 - $K = 2, 3, 5, 7$

❖ Shooting

- Ignored Year, Victim Age, Victim Age Band

❖ Attributes

- 5
- Day of Week, Month, Season, Multiple Shooting, Homicide?



Algorithm / Model Performance & Results

❖ Shooting, K = 2

```
Number of iterations: 3
Within cluster sum of squared errors: 1704.3043507294788

Initial starting points (random):

Cluster 0: Sat,11,Autumn,No,No
Cluster 1: Tue,6,Summer,No,No
```

Final cluster centroids:

Attribute	Cluster#		
	Full Data	0	1
	(1170.0)	(442.0)	(728.0)
=====			
DayOfWeek	Sat	Sat	Sun
Occurred_Month	6.788	8.9321	5.4863
Season	Summer	Autumn	Summer
Multiple_Shooting	No	No	No
Homicide?	No	No	No

Clustered Instances

0	442 (38%)
1	728 (62%)



Algorithm / Model Performance & Results

❖ Shooting, K = 3

Number of iterations: 4

Within cluster sum of squared errors: 1587.5501189616261

Initial starting points (random):

Cluster 0: Sat,11,Autumn,No,No

Cluster 1: Tue,6,Summer,No,No

Cluster 2: Sun,9,Autumn,Yes,No

Final cluster centroids:

Attribute	Full Data (1170.0)	Cluster#		
		0 (420.0)	1 (627.0)	2 (123.0)
DayOfWeek	Sat	Sat	Sun	Sun
Occurred_Month	6.788	8.9571	5.4817	6.0407
Season	Summer	Autumn	Summer	Summer
Multiple_Shooting	No	No	No	Yes
Homicide?	No	No	No	No

Clustered Instances

0	420 (36%)
1	627 (54%)
2	123 (11%)



Algorithm / Model Performance & Results

❖ Shooting, K = 5

```
Number of iterations: 7  
Within cluster sum of squared errors: 1313.258582506489
```

```
Initial starting points (random):
```

```
Cluster 0: Sat,11,Autumn,No,No  
Cluster 1: Tue,6,Summer,No,No  
Cluster 2: Sun,9,Autumn,Yes,No  
Cluster 3: Tue,10,Autumn,No,Yes  
Cluster 4: Thu,8,Summer,No,No
```

```
Final cluster centroids:
```

Attribute	Full Data	Cluster#				
		0	1	2	3	4
	(1170.0)	(314.0)	(281.0)	(83.0)	(89.0)	(403.0)
=====						
DayOfWeek	Sat	Sat	Tue	Sun	Sun	Sat
Occurred_Month	6.788	10.2261	3.7153	5.0964	5.9888	6.7767
Season	Summer	Autumn	Spring	Spring	Spring	Summer
Multiple_Shooting	No	No	No	Yes	No	No
Homicide?	No	No	No	No	Yes	No

```
Clustered Instances
```

```
0      314 ( 27%)  
1      281 ( 24%)  
2       83 (  7%)  
3       89 (  8%)  
4      403 ( 34%)
```




Algorithm / Model Performance & Results

❖ Clustering

- K Simple Means
 - $K = 2, 3, 5, 7$

❖ Homicides

- Ignored Year

❖ Attributes

- 5
- Day of Week, Month, Season, Section, Weapon Category



Algorithm / Model Performance & Results

❖ Homicide, K = 2

Number of iterations: 3
Within cluster sum of squared errors: 488.9957126184171

Initial starting points (random):

Cluster 0: 11, Sat, Lake, Firearm, Autumn
Cluster 1: 3, Sat, Clinton, Firearm, Spring

Final cluster centroids:

Attribute	Cluster#		
	Full Data	0	1
	(244.0)	(115.0)	(129.0)
=====			
OccurredMonth	6.6598	8.4087	5.1008
DayOfWeek	Sun	Sun	Sat
Section	Clinton	Lake	Clinton
WeaponCategory	Firearm	Firearm	Firearm
Season	Summer	Autumn	Spring

Clustered Instances

0	115 (47%)
1	129 (53%)



Algorithm / Model Performance & Results

❖ Homicide, $K = 3$

Number of iterations: 5

Within cluster sum of squared errors: 417.09234254959324

Initial starting points (random):

Cluster 0: 11, Sat, Lake, Firearm, Autumn

Cluster 1: 3, Sat, Clinton, Firearm, Spring

Cluster 2: 10, Mon, Clinton, Knife/CuttingInstrument, Autumn

Final cluster centroids:

Attribute	Full Data (244.0)	Cluster#		
		0 (84.0)	1 (73.0)	2 (87.0)
OccurredMonth	6.6598	8.8095	4.0274	6.7931
DayOfWeek	Sun	Sun	Sat	Fri
Section	Clinton	Lake	Clinton	Clinton
WeaponCategory	Firearm	Firearm	Firearm	Firearm
Season	Summer	Autumn	Spring	Summer

Clustered Instances

0	84 (34%)
1	73 (30%)
2	87 (36%)



Algorithm / Model Performance & Results

❖ Homicide, K = 5

Number of iterations: 5

Within cluster sum of squared errors: 389.7705360636493

Initial starting points (random):

Cluster 0: 11, Sat, Lake, Firearm, Autumn

Cluster 1: 3, Sat, Clinton, Firearm, Spring

Cluster 2: 10, Mon, Clinton, Knife/CuttingInstrument, Autumn

Cluster 3: 1, Fri, Central, Other, Winter

Cluster 4: 7, Fri, Lake, Knife/CuttingInstrument, Summer

Clustered Instances

0 68 (28%)

1 60 (25%)

2 17 (7%)

3 17 (7%)

4 82 (34%)

Final cluster centroids:

Attribute	Full Data (244.0)	Cluster# 0 (68.0)	1 (60.0)	2 (17.0)	3 (17.0)	4 (82.0)
OccurredMonth	6.6598	9.2794	3.95	7.8824	2.9412	6.9878
DayOfWeek	Sun	Sat	Sat	Mon	Fri	Fri
Section	Clinton	Lake	Clinton	Clinton	Goodman	Clinton
WeaponCategory	Firearm	Firearm	Firearm	Knife/CuttingInstrument	Other	Firearm
Season	Summer	Autumn	Spring	Autumn	Winter	Summer



Algorithm / Model

Performance & Results

❖ Homicide, K = 7

Number of iterations: 6
Within cluster sum of squared errors: 344.56333543345727

Initial starting points (random):

Cluster 0: 11, Sat, Lake, Firearm, Autumn
Cluster 1: 3, Sat, Clinton, Firearm, Spring
Cluster 2: 10, Mon, Clinton, Knife/CuttingInstrument, Autumn
Cluster 3: 1, Fri, Central, Other, Winter
Cluster 4: 7, Fri, Lake, Knife/CuttingInstrument, Summer
Cluster 5: 4, Mon, Clinton, Other, Spring
Cluster 6: 3, Wed, Lake, Other, Spring

Clustered Instances

0	45 (18%)
1	33 (14%)
2	21 (9%)
3	24 (10%)
4	35 (14%)
5	59 (24%)
6	27 (11%)

Final cluster centroids:

Attribute	Full Data (244.0)	Cluster# 0 (45.0)	1 (33.0)	2 (21.0)	3 (24.0)	4 (35.0)	5 (59.0)	6 (27.0)
OccurredMonth	6.6598	9.9778	4.2727	8.6667	3.5	7.3714	6.7797	4.1111
DayOfWeek	Sun	Sat	Sat	Thu	Fri	Fri	Mon	Wed
Section	Clinton	Lake	Clinton	Clinton	Goodman	Lake	Clinton	Lake
WeaponCategory	Firearm	Firearm	Firearm	Knife/CuttingInstrument	Firearm	Firearm	Firearm	Firearm
Season	Summer	Autumn	Spring	Autumn	Winter	Summer	Summer	Spring



Conclusions

Lessons Learned

- ❖ Maybe shouldn't live here...
- ❖ Problem solution
 - More personnel / patrols in the Summer, on weekends, in sections of Clinton & Lake
- ❖ Questions raised
 - Section size vs infrastructure
 - Legal vs illegal guns
 - Rochester vs other cities (especially those in different climates)
 - Time of day/night of occurrence
- ❖ Importance of each data mining step
 - Preprocessing
- ❖ Importance of data visualization
- ❖ Difference between working with discrete vs continuous values