JOHNSON COUNTY®
COMMUNITY COLLEGE

DS 280 – SPRING 2019

Big Data Architecture

Final Project

Wes McNall

Crime in KCMO

What insights can we gain from a dataset on crime in KCMO?

- How has crime changed overtime in KCMO?
- When are crimes reported?
- What's the age of those involved in crime?
- Where in KCMO has the most crime?

Dataset Details

- File Type: .csv
- File Size: 138 MB
- Data Source: https://data.kcmo.org
- SQL Script Name: crime.hql
- Anything interesting about dataset?

- A combination of five datasets all from the same website, from 2014-2018 Crime in KCMO.
- Did some minor editing in Python to combine them cleanly
- Interesting columns:
  - Description of the crime
  - Zip code of the crime
  - When the crime was reported
  - An estimation of when the crime could have happened
  - If a firearm was used in the crime

JOHNSON COUNTY
COMMUNITY COLLEGE

DB & Table Details

Database Name: crime_db
Table Name: crime_tbl_partition
Table Type: External

DB & Table Details

Field Names:

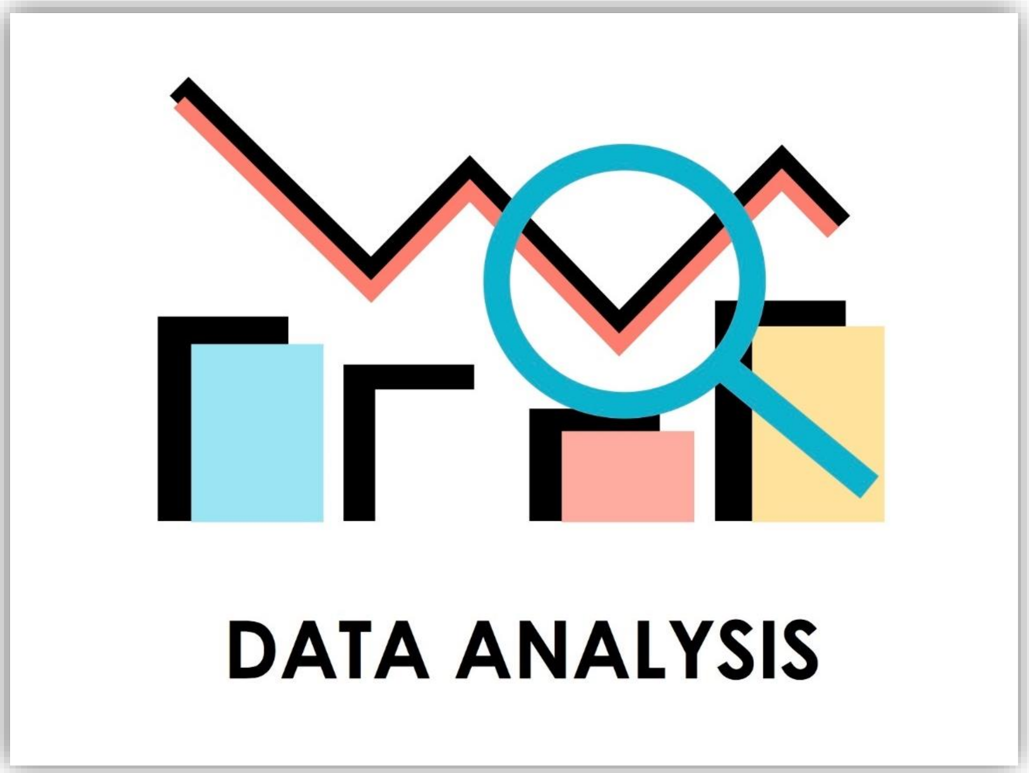
- Address
- Age
- Area
- Beat
- City
- DV_Flag
- Description
- Firearm_Used_Flag
- From_Date
- IBRS
- Invl_No
- Involvement
- Latitude
- Longitude
- Offense
- Race
- Rep_Dist
- Report_No
- Reported_Date
- Sex
- To_Date
- Zip_Code
- Year

DB & Table Details

Anything interesting about the table?

Before partitioning the data, my SQL queries were taking 300+ seconds to complete.

After partitioning by year, those went down to ~20 seconds to complete.

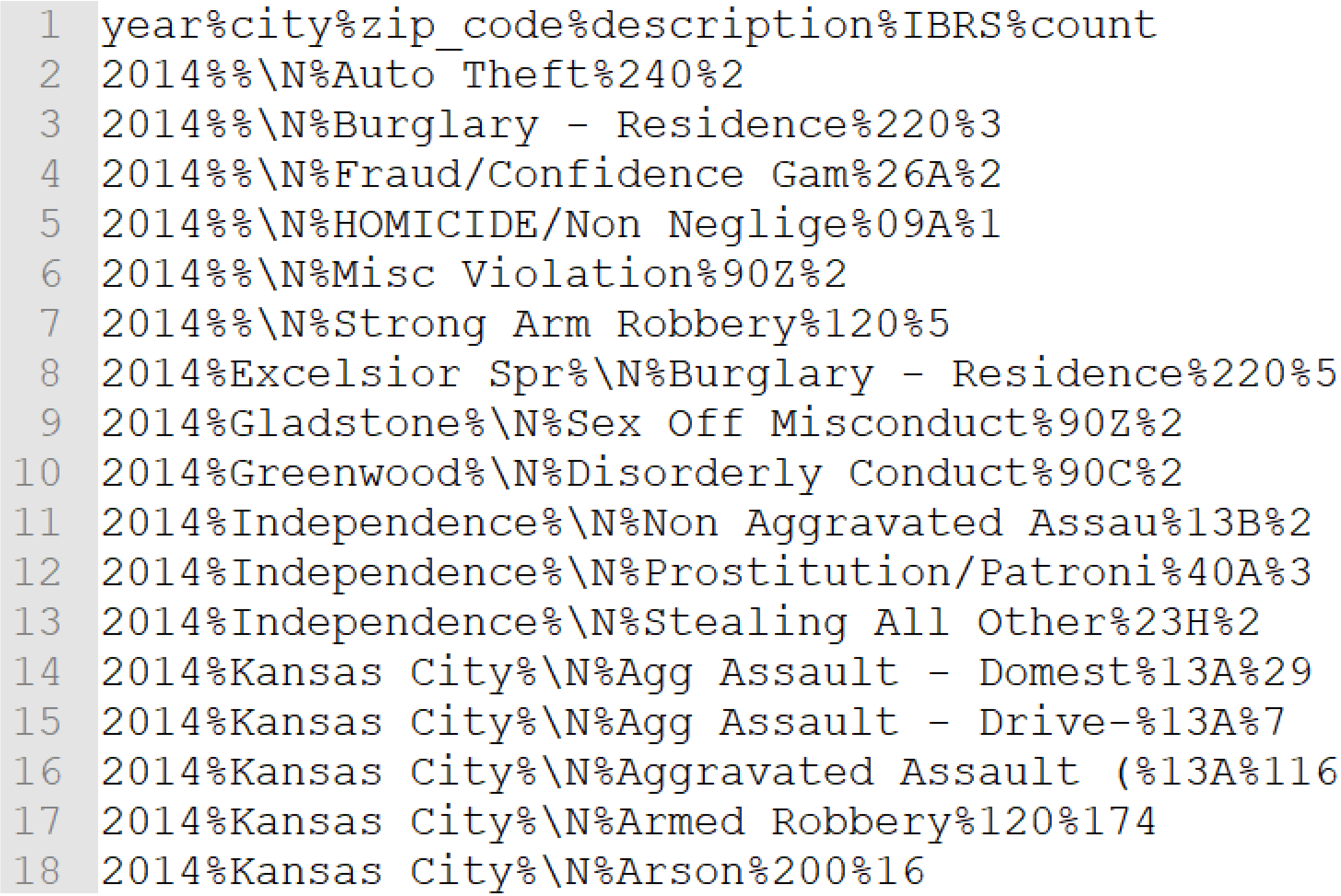

DATA ANALYSIS

Data Discovery - Query 1

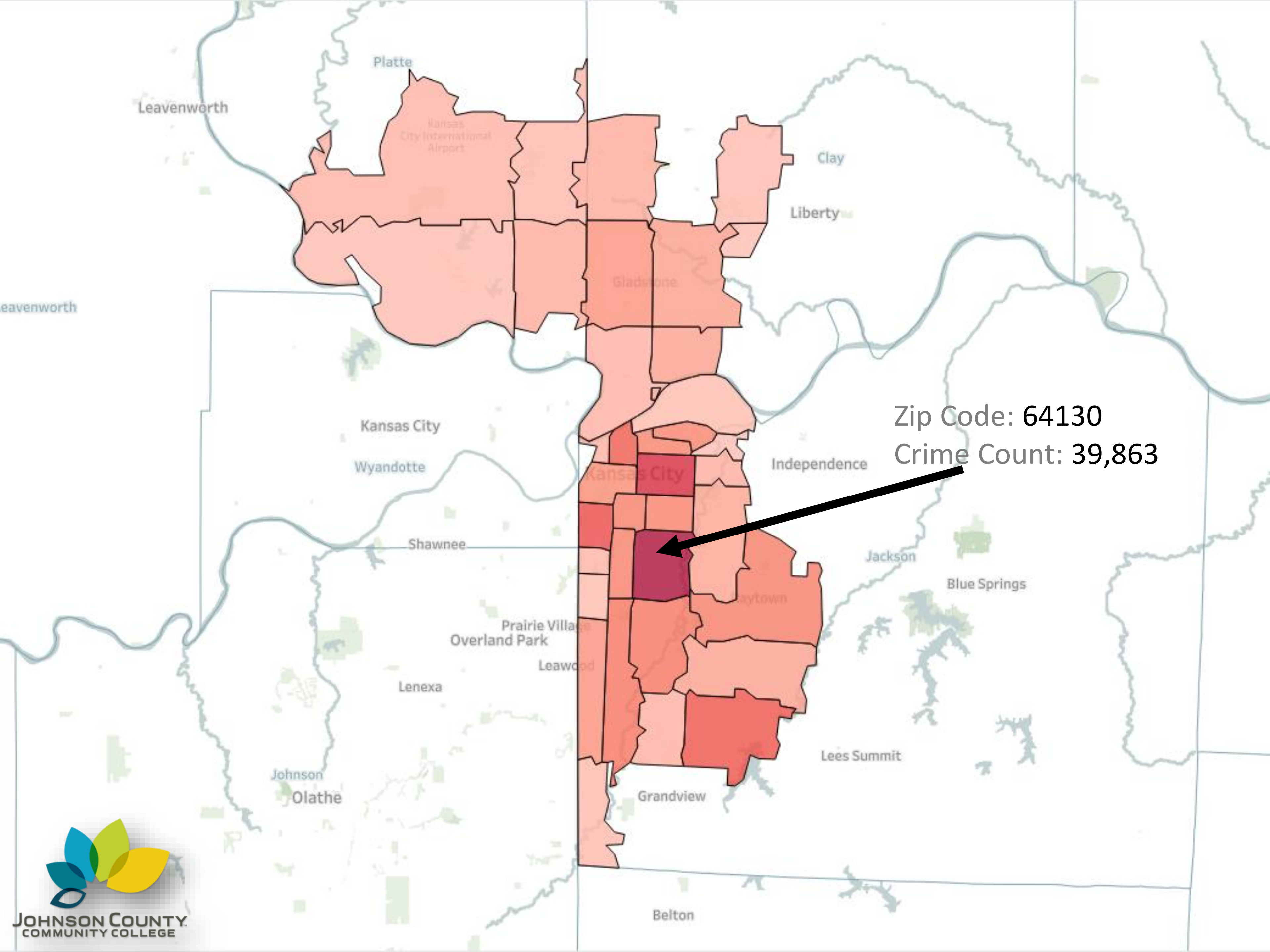
How are the number of crimes broken down by Zip Code? How has the number of crimes within these areas changed year by year?

```
set hive.groupby.orderby.position.alias = true;
SELECT Year, INITCAP(City), Zip_Code, description, ibrs, COUNT(Firearm_Used_Flag)
FROM crime_tbl_partition
GROUP BY 3, 1, 2, 4, 5;
```

Data Discovery - Query 1



```
year%city%zip_code%description%IBRS%count
2014%%\N%Auto Theft%240%2
2014%%\N%Burglary - Residence%220%3
2014%%\N%Fraud/Confidence Gam%26A%2
2014%%\N%HOMICIDE/Non Neglige%09A%1
2014%%\N%Misc Violation%90Z%2
2014%%\N%Strong Arm Robbery%120%5
2014%Excelsior Spr%\N%Burglary - Residence%220%5
2014%Gladstone%\N%Sex Off Misconduct%90Z%2
2014%Greenwood%\N%Disorderly Conduct%90C%2
2014%Independence%\N%Non Aggravated Assau%13B%2
2014%Independence%\N%Prostitution/Patroni%40A%3
2014%Independence%\N%Stealing All Other%23H%2
2014%Kansas City%\N%Agg Assault - Domest%13A%29
2014%Kansas City%\N%Agg Assault - Drive-%13A%7
2014%Kansas City%\N%Aggravated Assault (%13A%116
2014%Kansas City%\N%Armed Robbery%120%174
2014%Kansas City%\N%Arson%200%16
```



Zip Code: 64130
Crime Count: 39,863

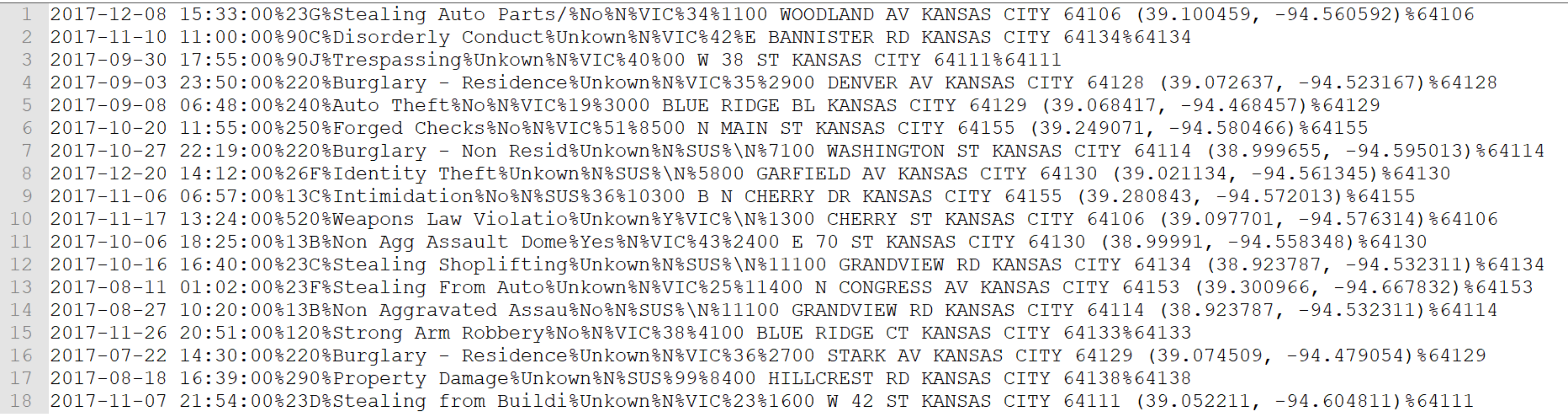
Johnson County Community College

Data Discovery - Query 2

What crimes have happened since I moved to Kansas City?

```
SELECT reported_Date, ibrs, description, INITCAP(CASE WHEN DV_Flag = 'Y' THEN
'Yes' WHEN DV_FLAG = 'N' THEN 'NO' ELSE 'UNKOWN' END), Firearm_Used_Flag,
Involvement, Age, Location, Zip_Code
FROM crime_tbl_partition
WHERE reported_date > '2017-07-01 00:00:00';
```

Data Discovery - Query 2



```
2017-12-08 15:33:00%23G%Stealing Auto Parts/%No%N%VIC%34%1100 WOODLAND AV KANSAS CITY 64106 (39.100459, -94.560592)%64106
2017-11-10 11:00:00%90C%Disorderly Conduct%Unkown%N%VIC%42%E BANNISTER RD KANSAS CITY 64134%64134
2017-09-30 17:55:00%90J%Trespassing%Unkown%N%VIC%40%00 W 38 ST KANSAS CITY 64111%64111
2017-09-03 23:50:00%220%Burglary - Residence%Unkown%N%VIC%35%2900 DENVER AV KANSAS CITY 64128 (39.072637, -94.523167)%64128
2017-09-08 06:48:00%240%Auto Theft%No%N%VIC%19%3000 BLUE RIDGE BL KANSAS CITY 64129 (39.068417, -94.468457)%64129
2017-10-20 11:55:00%250%Forged Checks%No%N%VIC%51%8500 N MAIN ST KANSAS CITY 64155 (39.249071, -94.580466)%64155
2017-10-27 22:19:00%220%Burglary - Non Resid%Unkown%N%SUS%\N%7100 WASHINGTON ST KANSAS CITY 64114 (38.999655, -94.595013)%64114
2017-12-20 14:12:00%26F%Identity Theft%Unkown%N%SUS%\N%5800 GARFIELD AV KANSAS CITY 64130 (39.021134, -94.561345)%64130
2017-11-06 06:57:00%13C%Intimidation%No%N%SUS%36%10300 B N CHERRY DR KANSAS CITY 64155 (39.280843, -94.572013)%64155
2017-11-17 13:24:00%520%Weapons Law Violatio%Unkown%Y%VIC%\N%1300 CHERRY ST KANSAS CITY 64106 (39.097701, -94.576314)%64106
2017-10-06 18:25:00%13B%Non Agg Assault Dome%Yes%N%VIC%43%2400 E 70 ST KANSAS CITY 64130 (38.99991, -94.558348)%64130
2017-10-16 16:40:00%23C%Stealing Shoplifting%Unkown%N%SUS%\N%11100 GRANDVIEW RD KANSAS CITY 64134 (38.923787, -94.532311)%64134
2017-08-11 01:02:00%23F%Stealing From Auto%Unkown%N%VIC%25%11400 N CONGRESS AV KANSAS CITY 64153 (39.300966, -94.667832)%64153
2017-08-27 10:20:00%13B%Non Aggravated Assau%No%N%SUS%\N%11100 GRANDVIEW RD KANSAS CITY 64114 (38.923787, -94.532311)%64114
2017-11-26 20:51:00%120%Strong Arm Robbery%No%N%VIC%38%4100 BLUE RIDGE CT KANSAS CITY 64133%64133
2017-07-22 14:30:00%220%Burglary - Residence%Unkown%N%VIC%36%2700 STARK AV KANSAS CITY 64129 (39.074509, -94.479054)%64129
2017-08-18 16:39:00%290%Property Damage%Unkown%N%SUS%99%8400 HILLCREST RD KANSAS CITY 64138%64138
2017-11-07 21:54:00%23D%Stealing from Buildi%Unkown%N%VIC%23%1600 W 42 ST KANSAS CITY 64111 (39.052211, -94.604811)%64111
```

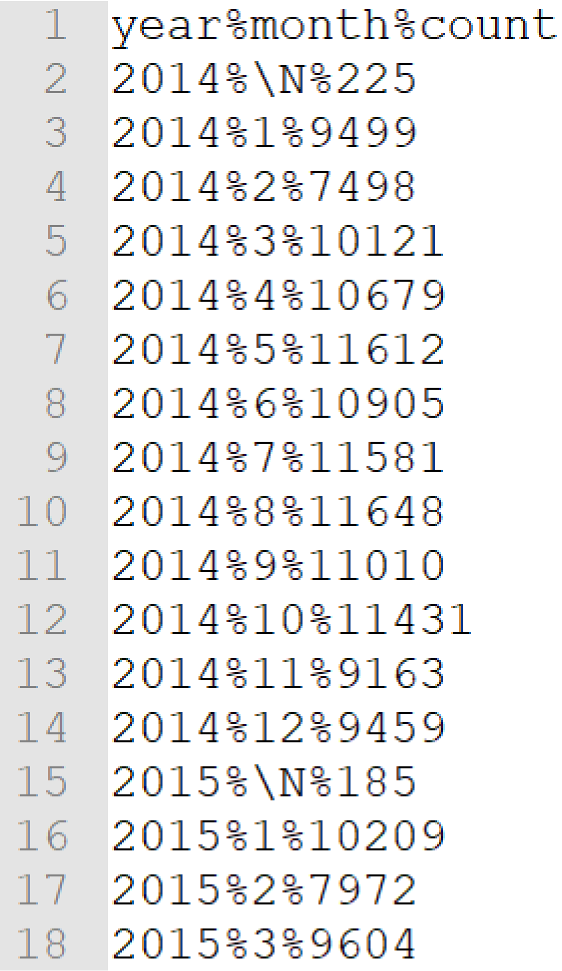
JOHNSON COUNTY
COMMUNITY COLLEGE

Data Discovery - Query 3

How has crime changed on a month level?
Are there any seasonal patterns to it?

```
SELECT Year, MONTH(reported_date), COUNT(Firearm_Used_Flag)
FROM crime_tbl_partition
GROUP BY 1, 2;
```

Data Discovery - Query 3



```
year%month%count
2014%\N%225
2014%1%9499
2014%2%7498
2014%3%10121
2014%4%10679
2014%5%11612
2014%6%10905
2014%7%11581
2014%8%11648
2014%9%11010
2014%10%11431
2014%11%9163
2014%12%9459
2015%\N%185
2015%1%10209
2015%2%7972
2015%3%9604
```

It turns out there isn't much seasonality related to crime. The time of year doesn't really affect the number of crimes that do happen

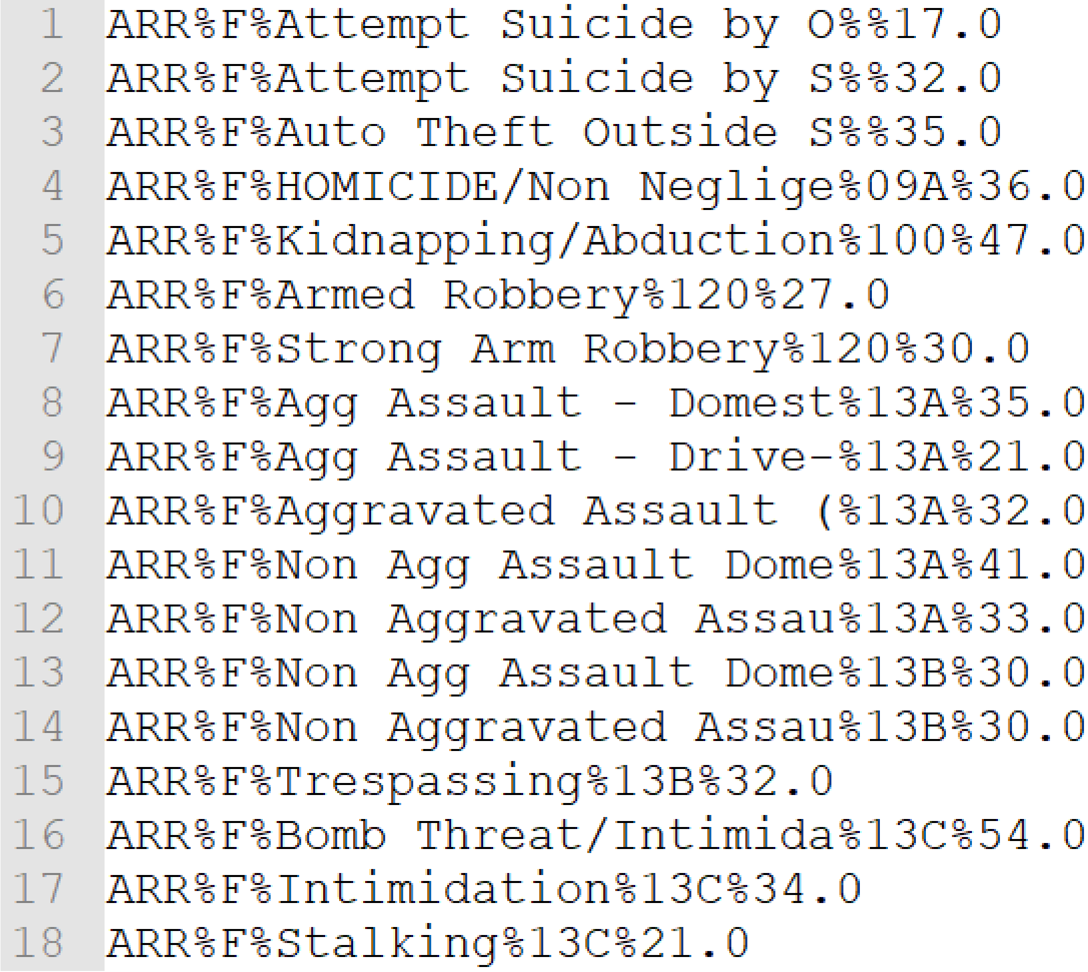
Data Discovery - Query 4

What crimes are committed by what age person?
What age are the victims of crimes?

```
ROUND(PERCENTILE(CAST(Age AS BIGINT), 0.5), 0)
FROM crime_tbl_partition
WHERE age > 0 AND age < 100
GROUP BY involvement, sex, ibrs, description;
```

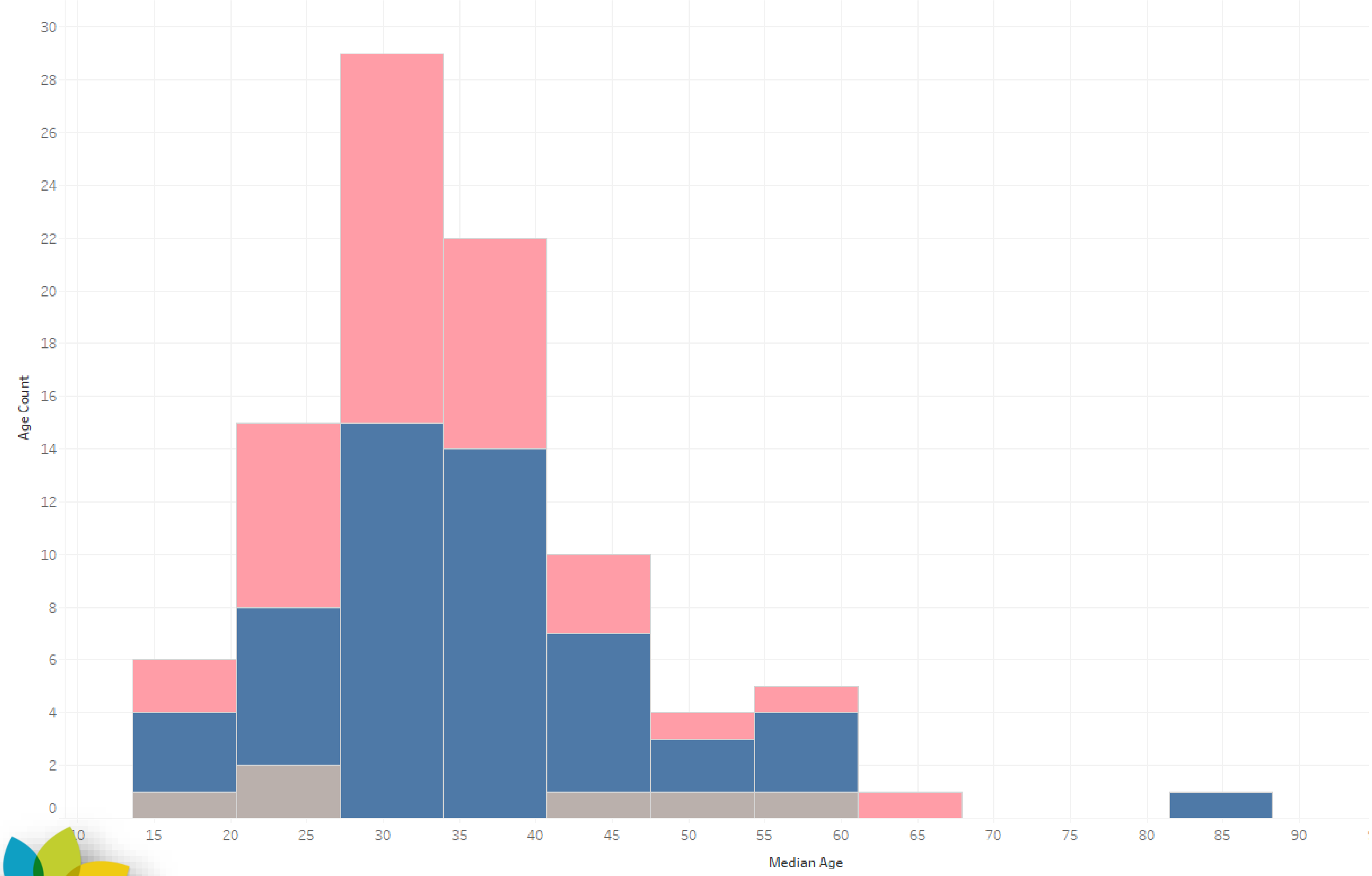
Johnson County
Community College

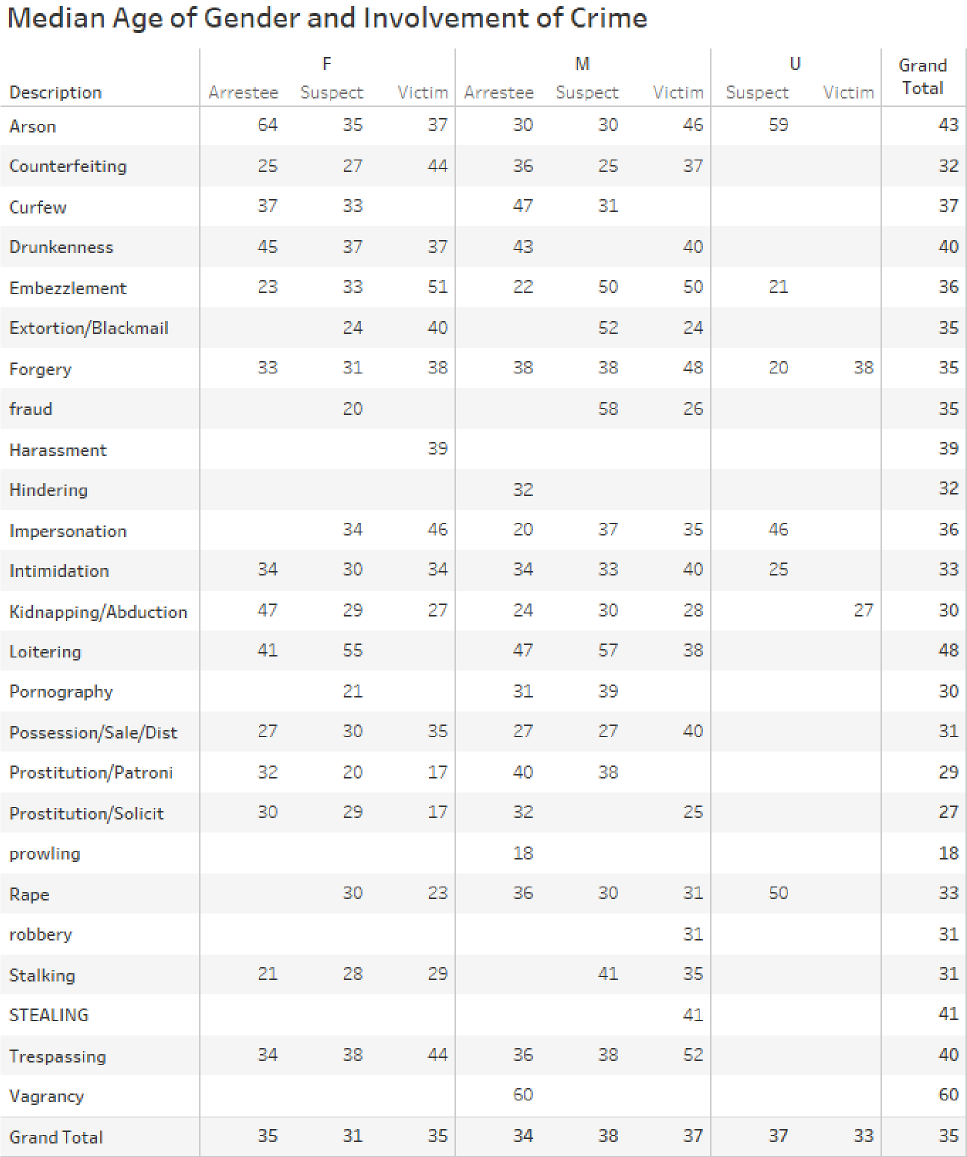
Data Discovery - Query 4



```
ARR%F%Attempt Suicide by O%%17.0
ARR%F%Attempt Suicide by S%%32.0
ARR%F%Auto Theft Outside S%%35.0
ARR%F%HOMICIDE/Non Neglige%09A%36.0
ARR%F%Kidnapping/Abduction%100%47.0
ARR%F%Armed Robbery%120%27.0
ARR%F%Strong Arm Robbery%120%30.0
ARR%F%Agg Assault - Domest%13A%35.0
ARR%F%Agg Assault - Drive-%13A%21.0
ARR%F%Aggravated Assault (%13A%32.0
ARR%F%Non Agg Assault Dome%13A%41.0
ARR%F%Non Aggravated Assau%13A%33.0
ARR%F%Non Agg Assault Dome%13B%30.0
ARR%F%Non Aggravated Assau%13B%30.0
ARR%F%Trespassing%13B%32.0
ARR%F%Bomb Threat/Intimida%13C%54.0
ARR%F%Intimidation%13C%34.0
ARR%F%Stalking%13C%21.0
```

Median Age of Those Arrested or Suspected of Crimes in KCMO by Gender





Median Age of Gender and Involvement of Crime

	F			M			U		Grand Total
Description	Arrestee	Suspect	Victim	Arrestee	Suspect	Victim	Suspect	Victim	
Arson	64	35	37	30	30	46	59		43
Counterfeiting	25	27	44	36	25	37			32
Curfew	37	33		47	31				37
Drunkenness	45	37	37	43		40			40
Embezzlement	23	33	51	22	50	50	21		36
Extortion/Blackmail		24	40		52	24			35
Forgery	33	31	38	38	38	48	20	38	35
fraud		20			58	26			35
Harassment			39						39
Hindering				32					32
Impersonation		34	46	20	37	35	46		36
Intimidation	34	30	34	34	33	40	25		33
Kidnapping/Abduction	47	29	27	24	30	28		27	30
Loitering	41	55		47	57	38			48
Pornography		21		31	39				30
Possession/Sale/Dist	27	30	35	27	27	40			31
Prostitution/Patroni	32	20	17	40	38				29
Prostitution/Solicit	30	29	17	32		25			27
prowling				18					18
Rape		30	23	36	30	31	50		33
robbery						31			31
Stalking	21	28	29		41	35			31
STEALING						41			41
Trespassing	34	38	44	36	38	52			40
Vagrancy				60					60
Grand Total	35	31	35	34	38	37	37	33	35

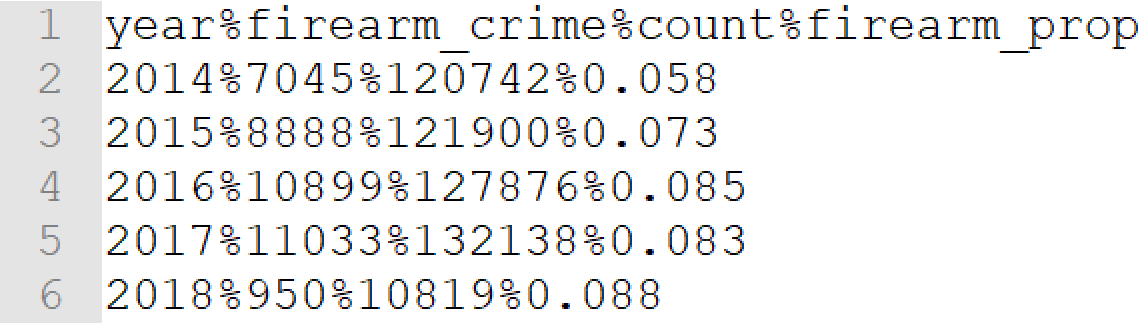
Johnson County
Community College

Data Discovery - Query 5

How have the crimes involving firearms changed over the years?

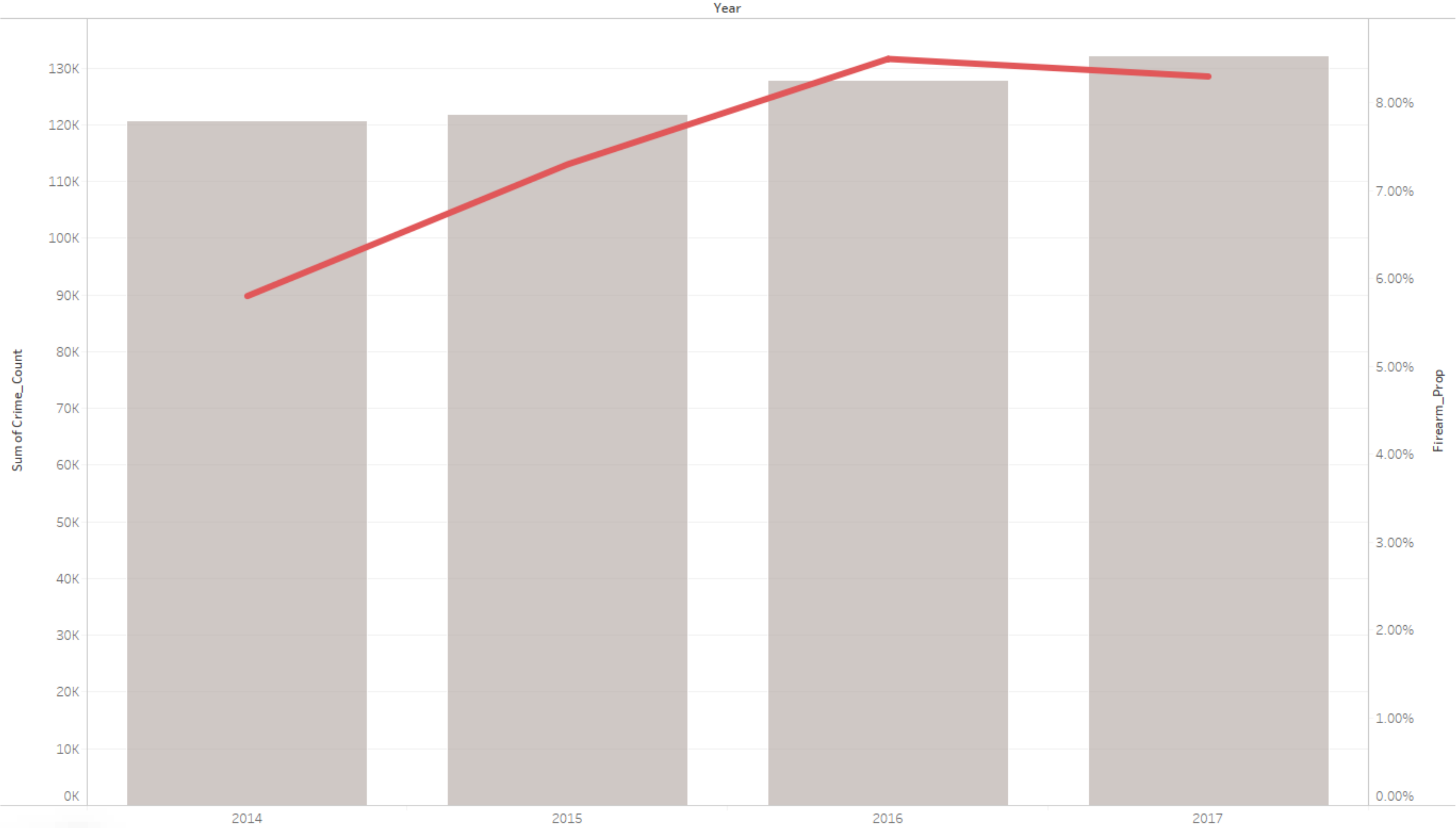
```
SELECT Year, firearm_crime, crime_count, ROUND(firearm_crime / crime_count, 3)
AS firearm_prop
FROM (
SELECT Year, SUM( CASE WHEN Firearm_Used_Flag = 'Y' THEN 1 ELSE 0 END ) AS
firearm_crime, COUNT(Zip_Code) AS crime_count
FROM crime_tbl_partition
GROUP BY Year
)
crime_tbl_partition;
```

Data Discovery - Query 5



```
year%firearm_crime%count%firearm_prop
2014%7045%120742%0.058
2015%8888%121900%0.073
2016%10899%127876%0.085
2017%11033%132138%0.083
2018%950%10819%0.088
```

Total Crime in KCMO and Proportion of Crimes Involving Firearms



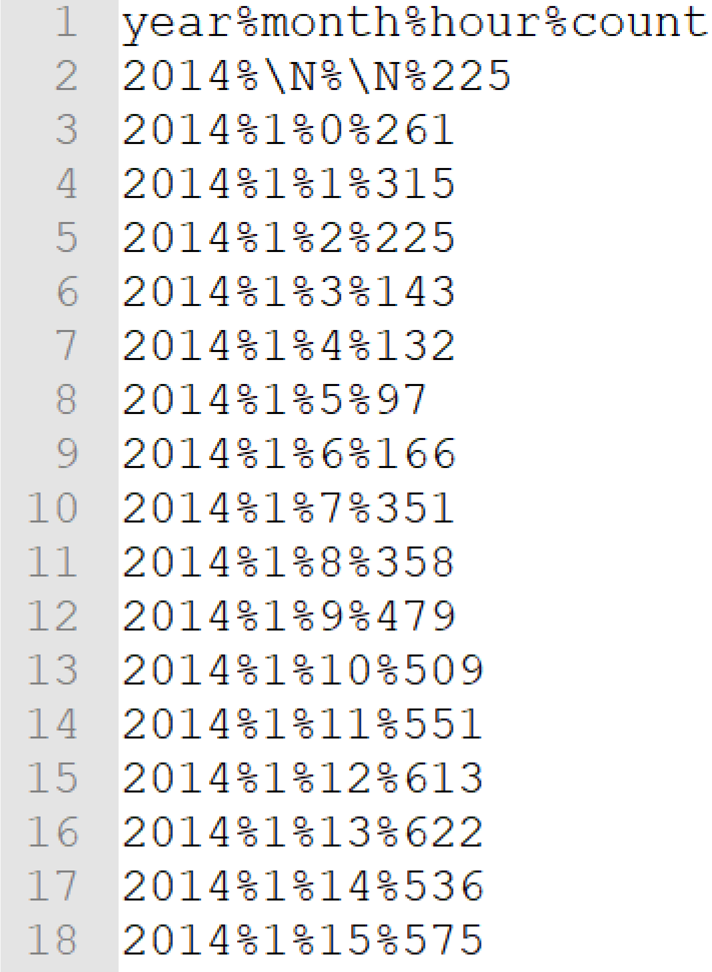
Data Discovery - Query 6

What hour of the day are crimes reported?

```
SELECT Year, MONTH(reported_date) as month, HOUR(reported_date) as hour,
SUM( CASE WHEN Zip_Code = -1 THEN 0 ELSE 1 END )
FROM crime_tbl_partition
GROUP BY 1, 2, 3;
```

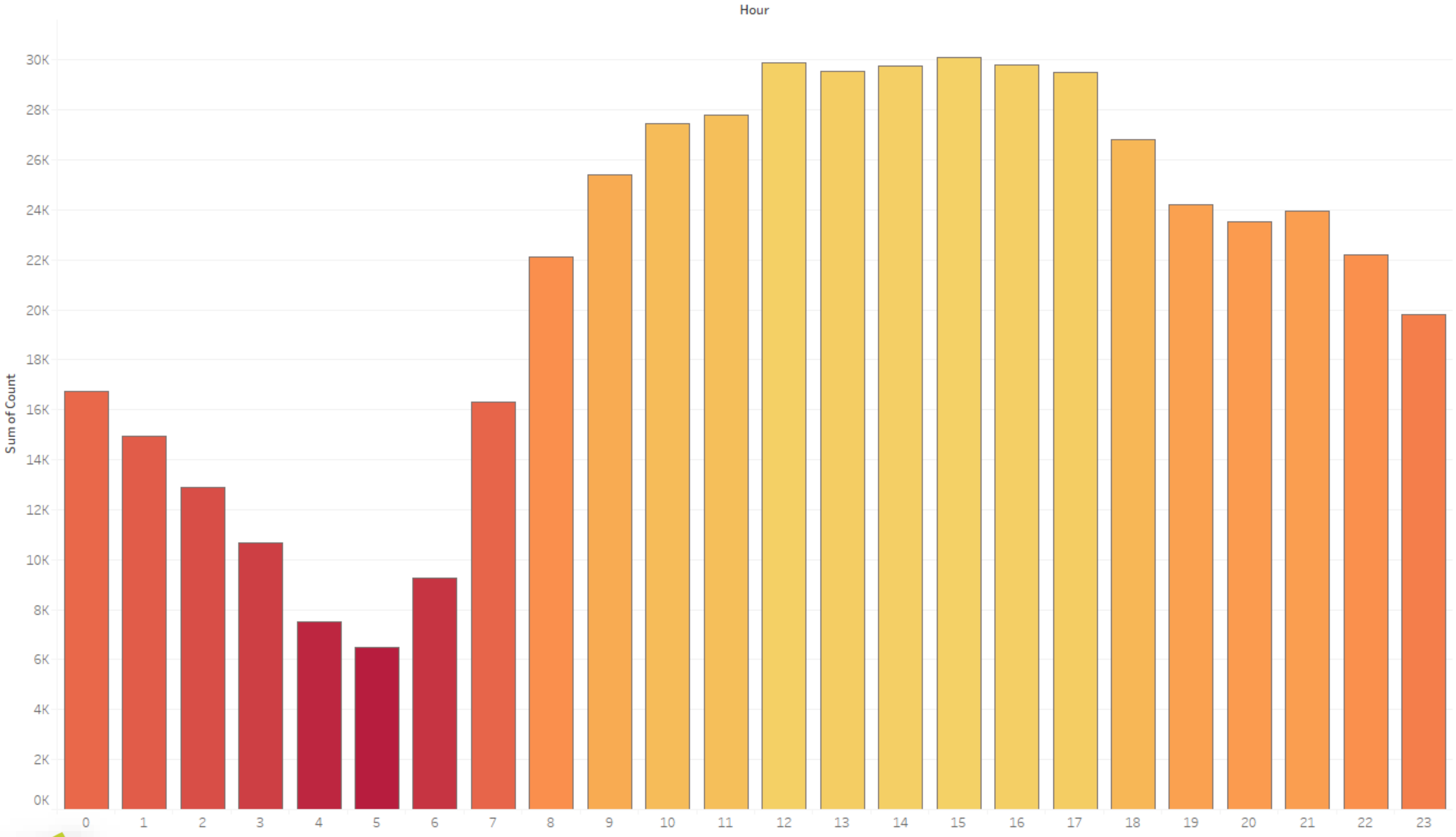
JOHNSON COUNTY
COMMUNITY COLLEGE

Data Discovery - Query 6



```
year%month%hour%count
2014%\N%\N%225
2014%1%0%261
2014%1%1%315
2014%1%2%225
2014%1%3%143
2014%1%4%132
2014%1%5%97
2014%1%6%166
2014%1%7%351
2014%1%8%358
2014%1%9%479
2014%1%10%509
2014%1%11%551
2014%1%12%613
2014%1%13%622
2014%1%14%536
2014%1%15%575
```

Reported Hour of Crime in KCMO



JOHNSON COUNTY COMMUNITY COLLEGE

Business Problem

- By doing this analysis I feel like I have a better understanding of some of the aspects of crime within KCMO
- I feel like if there are any other interesting questions about the dataset I'd be able to figure out how to write SQL queries to get those answers and be able to visualize them
- My clients can conclude that older suspects tend to skew male
- Knowing where the majority of crime happens allows my clients to be able to staff appropriately. Along with knowing what hour of the day most crimes are reported.
- The client can conclude that the 64130 zip code needs the most attention and that the majority of crimes are reported from 12-5pm so to have the most amount of staff on hand during those times

Johnson County
Community College

Business Problem