Crime in Baltimore

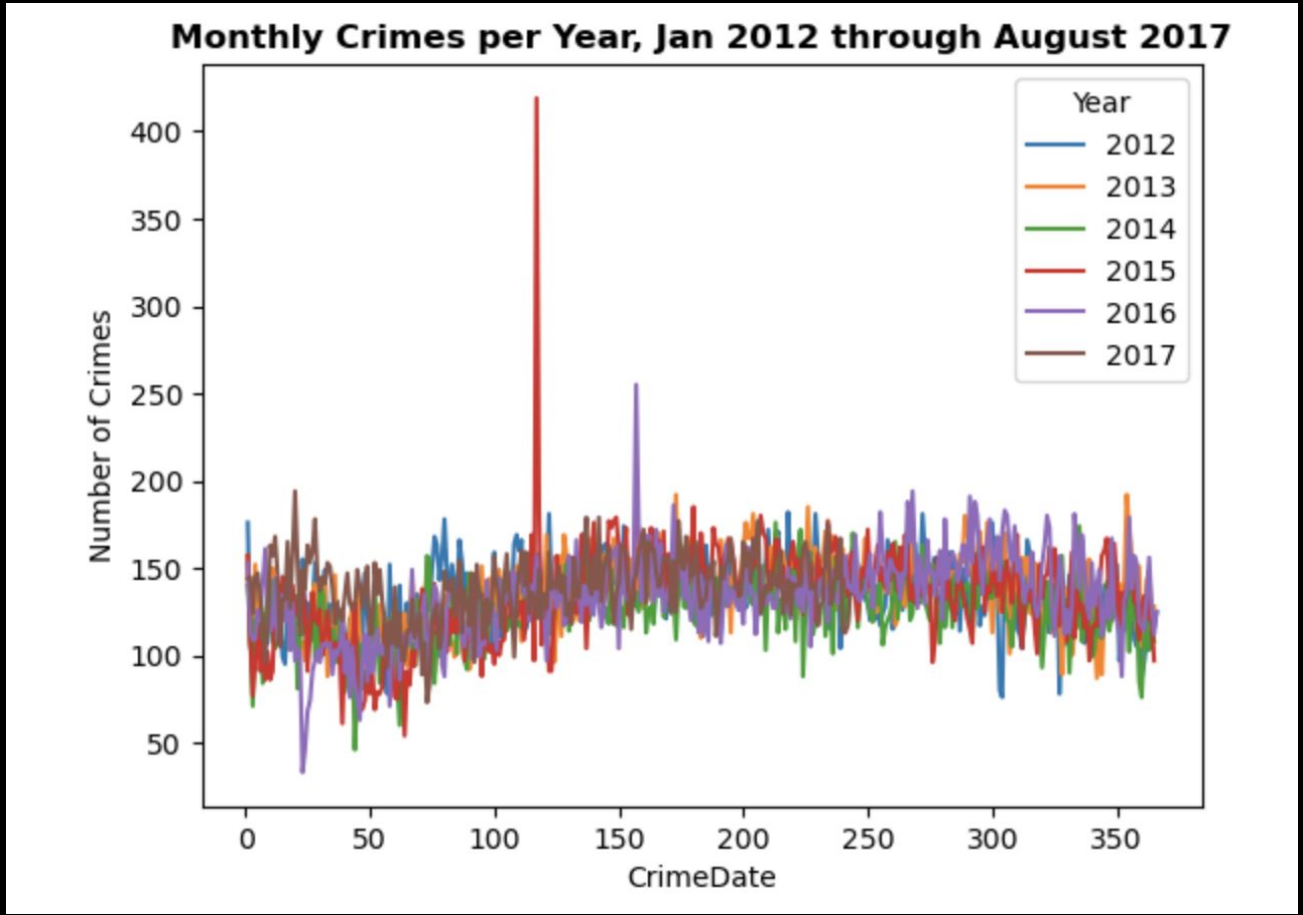
GW-Project 1:

Siobhan Byrne
Bryan Johns
Katie Loosararian
Robert Takan

Overview

- Does the Hour of Day Impact Crime Rates?
- Does the Day of the Week Impact Crime Rates?
- Does the Month of the Year Impact Crime Rates?

Outliers


Monthly Crimes per Year, Jan 2012 through August 2017
Year
2012
2013
2014
2015
2016
2017
Number of Crimes
400
350
300
250
200
150
100
50
0
50
100
150
200
250
300
350
CrimeDate

Does the Hour of Day Impact Crime Rates?

Highest crime rate per hour of day:

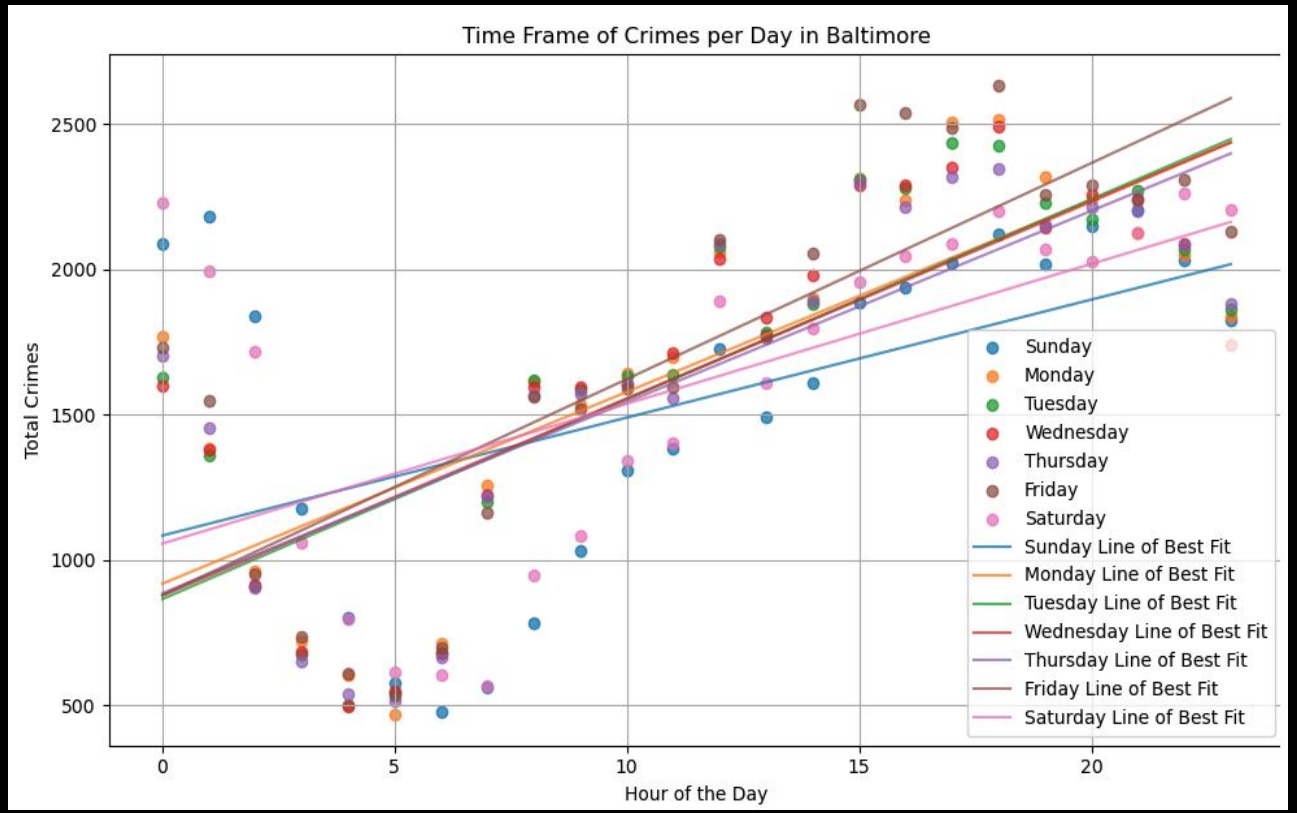
- Friday at 6pm with 2630 total crimes.

Lowest Crime rate per hour day of day:

- Monday at 5am with 469 total crimes

Line of best fit for all the days shows a slight raise in crime for all the days

- Sunday is the slowest rise in crime per hour



Linear Regression:

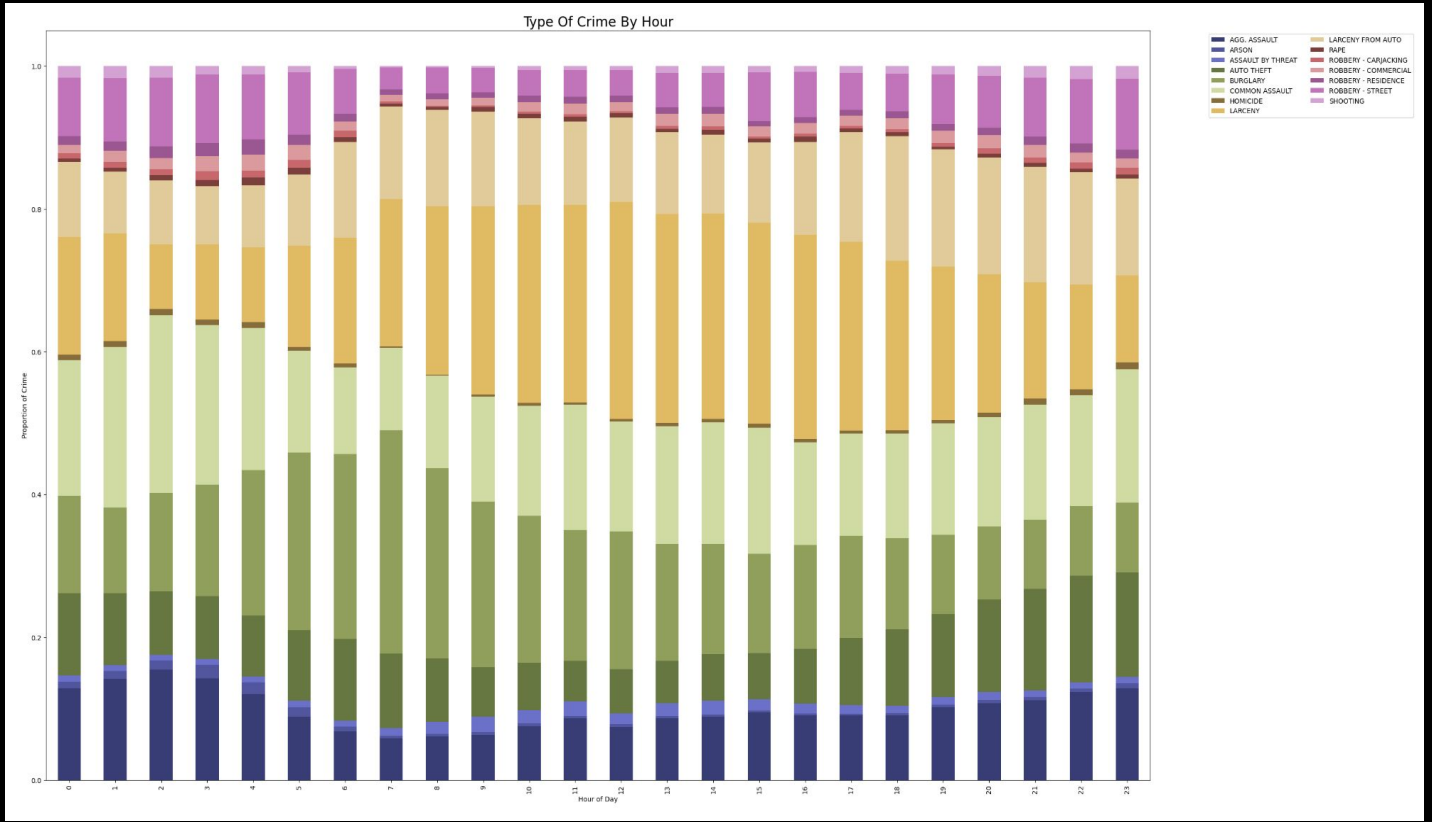
The pvalue is well below .05, signifying a statistical correlation between time of day and the number of victim-based crimes.

The rvalue is also high, at least .5, and often close to .8, meaning the scatter points fit the line relatively well.

```
The slope for Sunday is 40.587826086956525
The r-value is 0.4977664036358951
The pvalue is 0.013317812491476167
-----
The slope for Monday is 65.98130434782608
The r-value is 0.7599378235525983
The pvalue is 1.64674638973244e-05
-----
The slope for Tuesday is 68.78478260869565
The r-value is 0.7868991613438004
The pvalue is 5.09180748886111e-06
-----
The slope for Wednesday is 67.75999999999999
The r-value is 0.7783972560799419
The pvalue is 7.501013011413863e-06
-----
The slope for Thursday is 65.79434782608695
The r-value is 0.7730847128748713
The pvalue is 9.475650440040368e-06
-----
The slope for Friday is 74.30956521739131
The r-value is 0.7913580178198913
The pvalue is 4.126687464296085e-06
-----
The slope for Saturday is 48.11217391304348
The r-value is 0.581917220173801
The pvalue is 0.002854172971297055
-----
The pvalue is well below .05, signifying a statistical correlation between time of day and the number of victim-bas
ed crimes.
The rvalue is also high, at least .5, and often close to .8, meaning the scatter points fit the line relatively wel
l.
```

Type of Crime by Hour

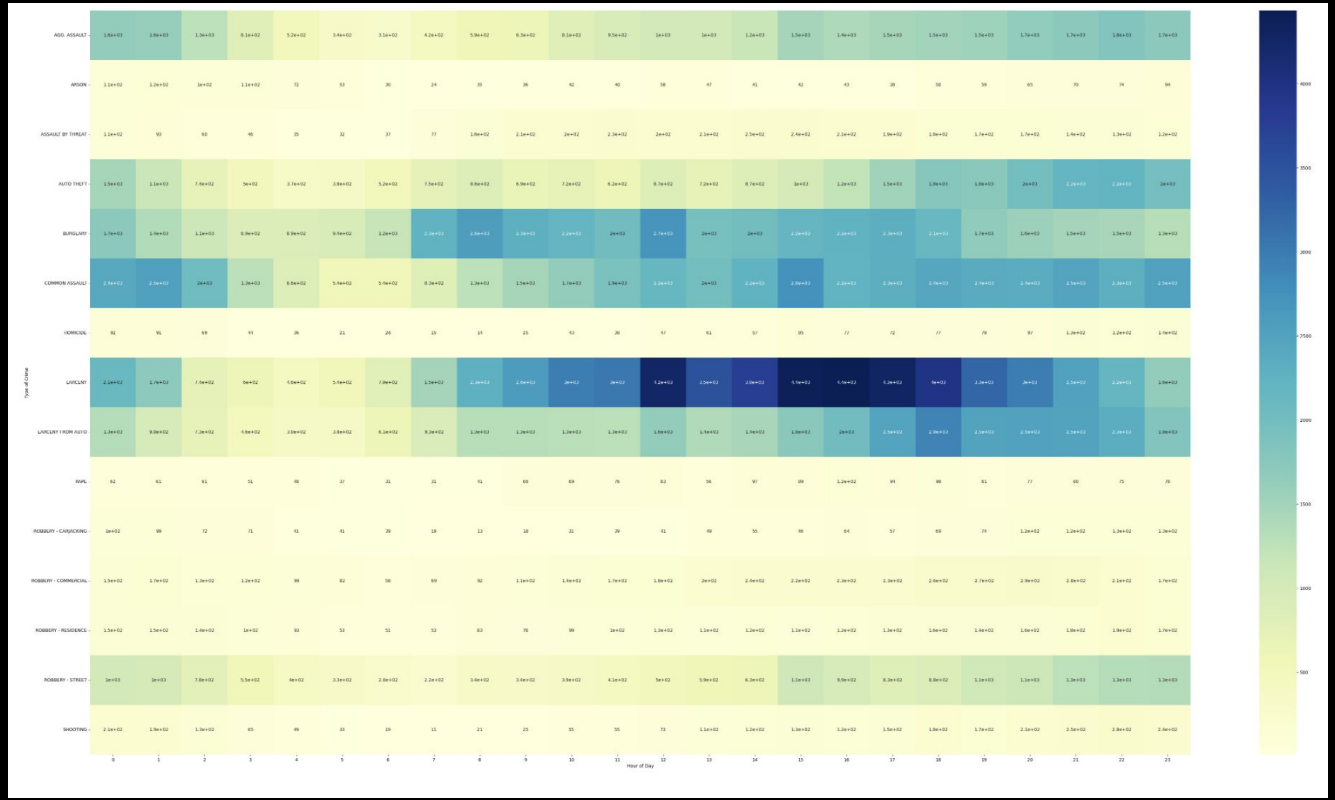
- Largest proportion for a crime per hour of day is Larceny at 4pm



Type of Crime by Hour

The chi square statistics indicate there is a strong statistical association between the frequency of crimes, and the two categorical variables of Type of Crime and Hour of the Day.

p-value = 0.0



Does the Day of the Week Impact Crime Rates?

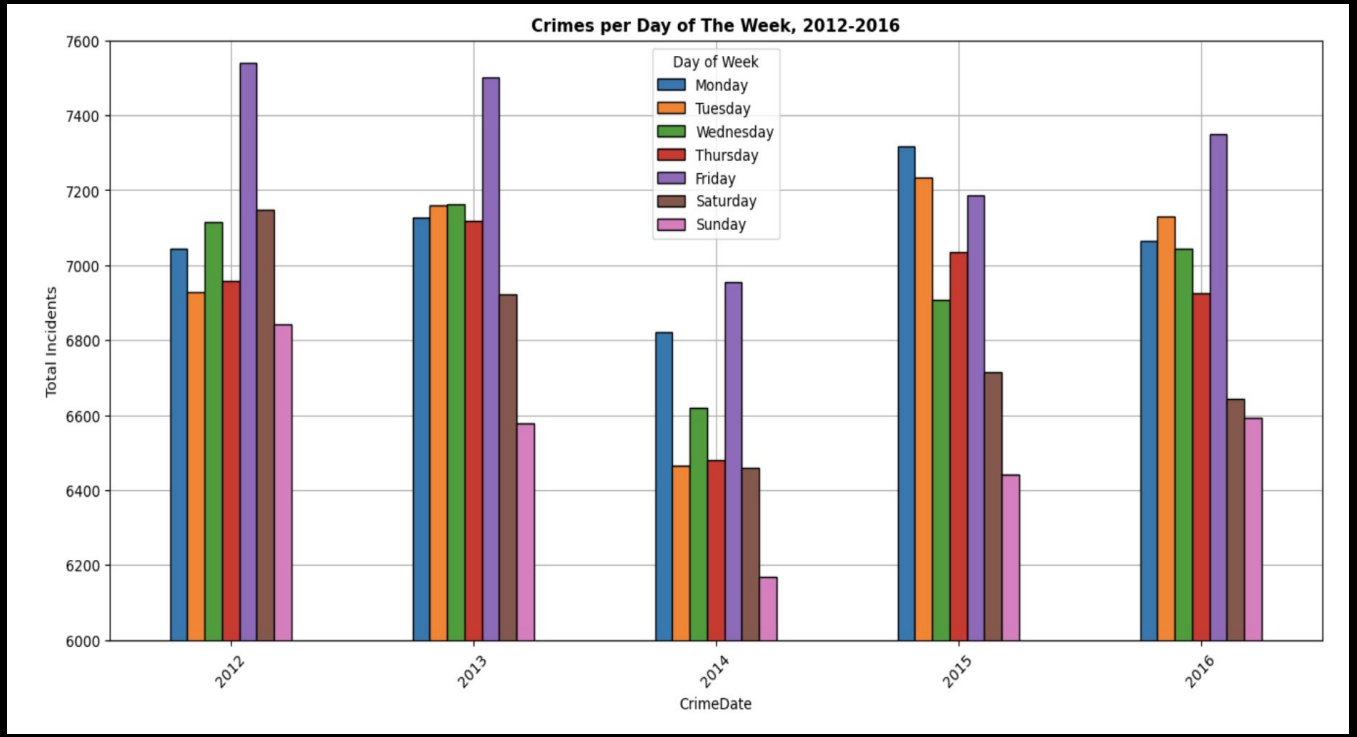
Highest crime rate per day of week:

- Friday between 7,400 and 7,500 total crimes.

Lowest Crime rate per day of week:

- Sunday between 6,400 and just over 6,800 total crimes

Comparing year to year, you can see a large drop across all days of the week in 2014. But afterwards an increase in 2015 which remains somewhat consistent in 2016.



Crime per Time of Day

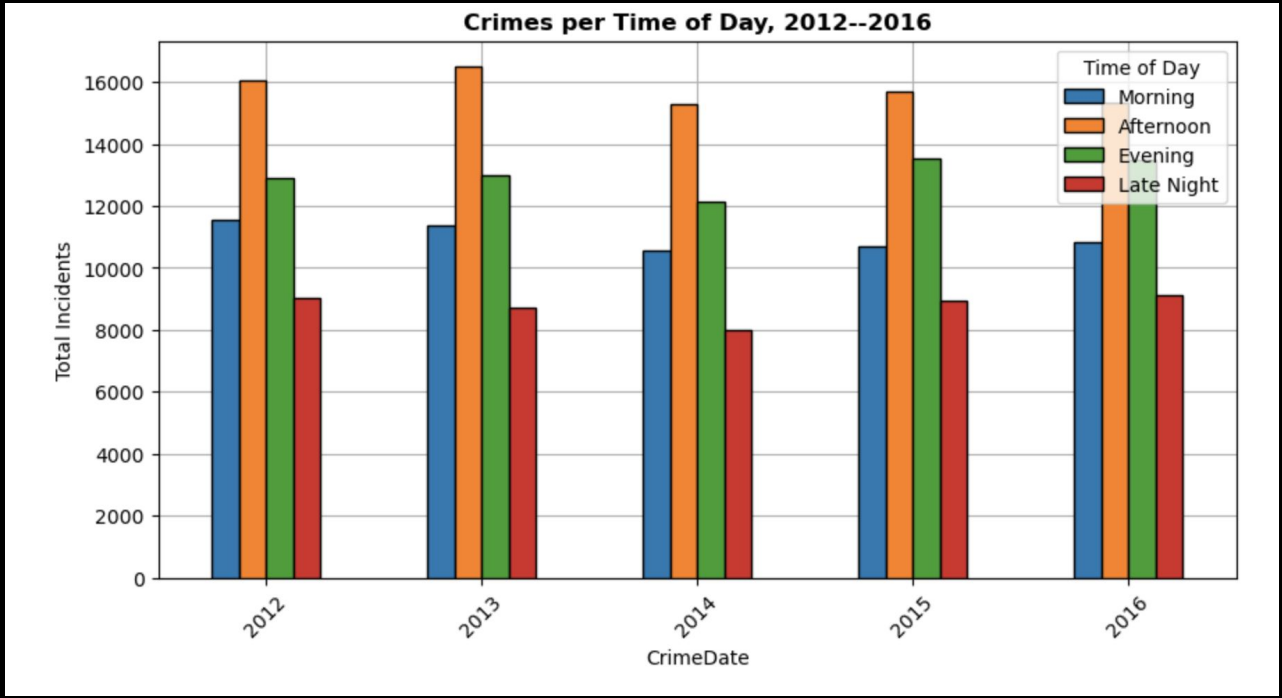
Highest crime rate per time of day:

- Afternoon with close to or above 16,000 total crimes.

Lowest crime rate per time of day:

- Late night with around 9,00 total crimes

There appears to be consistency year to year(2012-2016)of total crimes committed. So clearly no improvement.



Chi- squared analysis for crime per day of week

The chi square statistic(s) is(are) quite large, and the pvalue(s) effectively zero,

indicating there is a strong statistical association between the frequency of crimes

and the two categorical variables of time of day and day of the week.

"""

Day of Week	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
Time of Day							
Late Night	5842	5768	7832	8039	5501	5383	5399
Morning	8433	8605	6339	5973	8450	8636	8623
Afternoon	12393	11708	10356	9707	11405	11599	11708
Evening	9865	9296	9362	8902	9163	9301	9117

Chi square test:

```
Power_divergenceResult(statistic=array([2461.78575534, 2026.81931764, 1097.80102688,  949.24039729,
       2063.53078015, 2264.49173802, 2309.97569375]), pvalue=array([0.00000000e+000, 0.00000000e+000, 1.09177424e-237, 1.84
565997e-205,
       0.00000000e+000, 0.00000000e+000, 0.00000000e+000]))
```

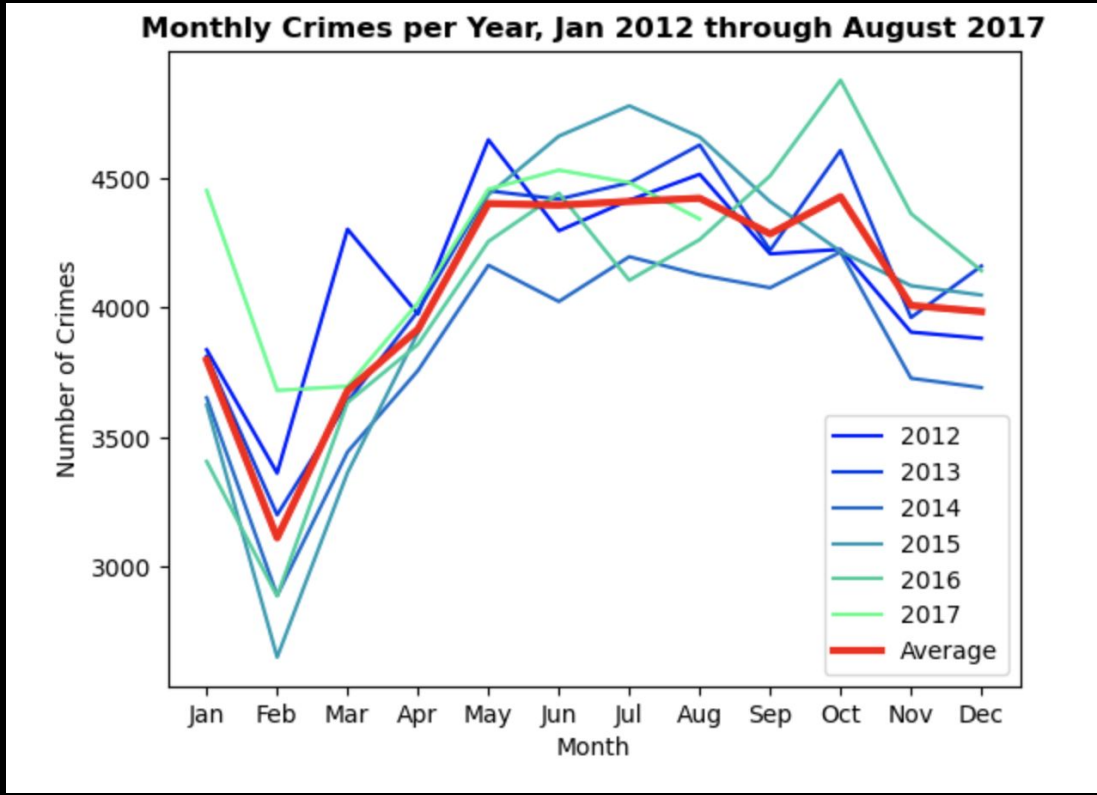
Chisquare contingency:

```
Chi2ContingencyResult(statistic=2679.7751310903027, pvalue=0.0, dof=18, expected_freq=array([[ 6587.54542346,  6379.0981974
,  6110.78550504,  5882.14269999,
         6224.38563688,  6296.51270472,  6283.52983251],
       [ 8287.71738118,  8025.47225232,  7687.91104839,  7400.25808698,
         7830.83010651,  7921.57236563,  7905.23875899],
       [11872.75461157, 11497.06949589, 11013.48865495, 10601.40498136,
        11218.23054325, 11348.22539297, 11324.82632002],
       [ 9784.9825838 ,  9475.36005439,  9076.81479162,  8737.19423168,
         9245.55371336,  9352.68953668,  9333.40508848]]))
```

Does the Month of the Year Impact Crime Rates?

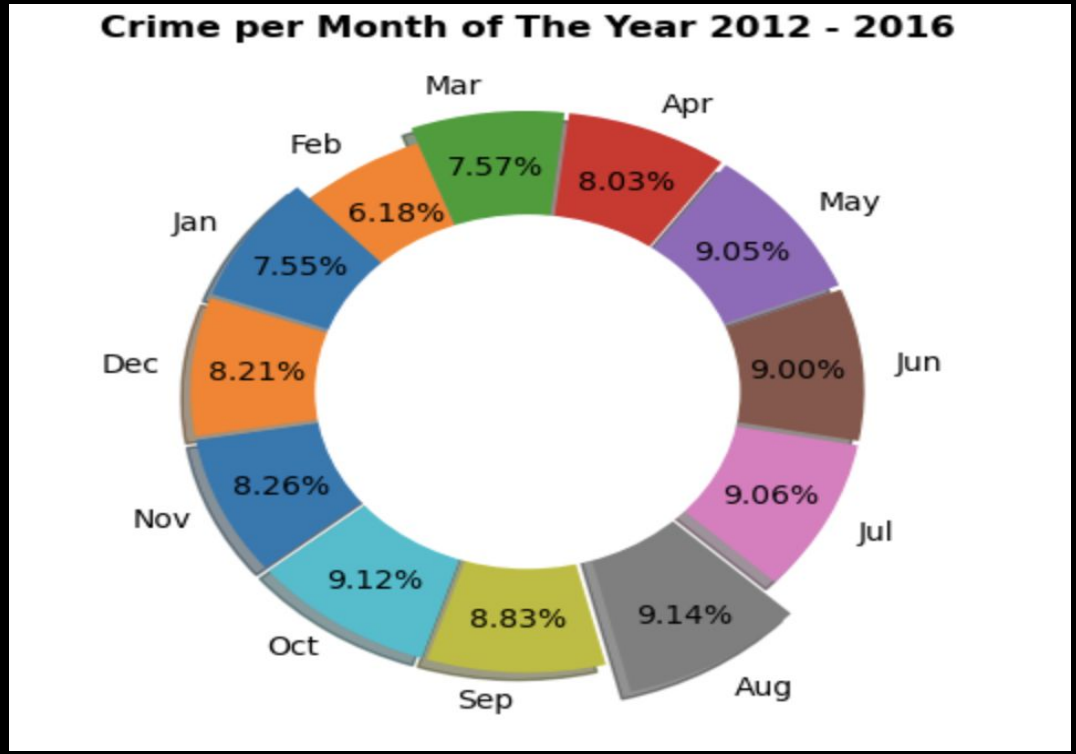
On average, summer months see the most Incidents of crime.

Winter months see the greatest reduction in crime rates.



Crime per Month of the Year

We performed statistical analysis on this data using ANOVA. The resulting p-value of 0.001 supports our hypothesis that month of the year does impact the rate of crime. The p-value is very small, so we reject the null hypothesis.



ANOVA

The number of victim-based crimes occuring is impacted by the month of the year, according to the extremely small pvalue.

```
month_df = df.copy()
month_df["Month"] = month_df.CrimeDate.dt.month
month_df["Year"] = month_df.CrimeDate.dt.year
# month_df = year_df.groupby("Month")["Year"].value_counts().unstack()
# month_df = month_df.reset_index()

jan = month_df[month_df.Month == 1].groupby(month_df.Year).size()
feb = month_df[month_df.Month == 2].groupby(month_df.Year).size()
mar = month_df[month_df.Month == 3].groupby(month_df.Year).size()
apr = month_df[month_df.Month == 4].groupby(month_df.Year).size()
may = month_df[month_df.Month == 5].groupby(month_df.Year).size()
jun = month_df[month_df.Month == 6].groupby(month_df.Year).size()
jul = month_df[month_df.Month == 7].groupby(month_df.Year).size()
aug = month_df[month_df.Month == 8].groupby(month_df.Year).size()
sep = month_df[month_df.Month == 9].groupby(month_df.Year).size()
oct = month_df[month_df.Month == 10].groupby(month_df.Year).size()
nov = month_df[month_df.Month == 11].groupby(month_df.Year).size()
dec = month_df[month_df.Month == 12].groupby(month_df.Year).size()

print(
    f"""{st.f_oneway(jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec)}
Month has a statistical impact on the frequency of crime over a year, as evidenced by the extremely low pvalue.""
)
```

```
F_onewayResult(statistic=3.337649313618808, pvalue=0.0013481654466416175)
Month has a statistical impact on the frequency of crime over a year, as evidenced by the extremely low pvalue.
```

Conclusion

Further analysis:

- This data can potentially be useful from an allocation of resources perspective-deciding what days of the week, times of day, and months of the year might benefit from increased coverage by law enforcement.
- Deeper dive into various types of crime depending on hour of week, and month.
- Crimes for inside vs outside depending on hour of week, and month.
- Crime types that are more prevalent in different neighborhood depending on hour of week, and month.


Questions?