**2018 Kansas City Crime Data Analysis**

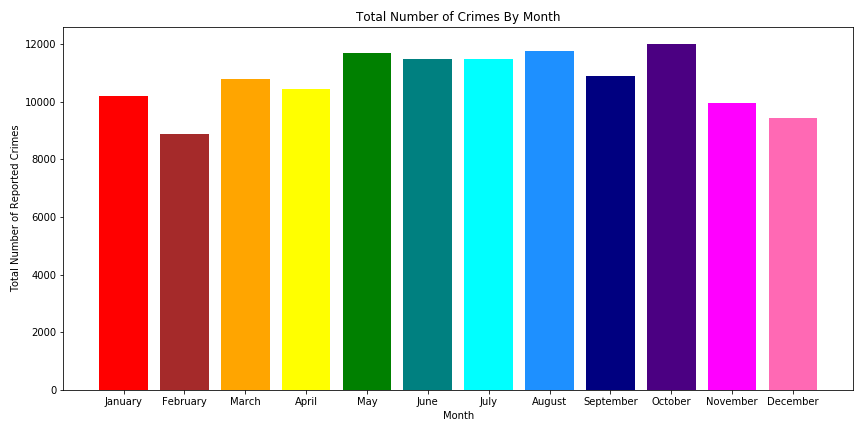
**What type of crime is the most common among different areas based on the greatest number of people involved in the crime?**

Our initial assumption was that stealing (larceny/theft offenses) would be a leading contribution to the crime rates. After cleaning the data by removing all of the default inputs when no data was present and or incorrect data entries in the dataset for ex, (99, ?, #); we were able to support our assumption as stealing (larceny/theft offenses) was the major reported crime type involving the greatest number of people. Following behind was assault offenses that involved the greatest number of people for the entire year of 2018 in Kansas City, MO. These types of crimes were the most common in the Urban areas of the city in the CPD- Central and EPD, - East police divisions.

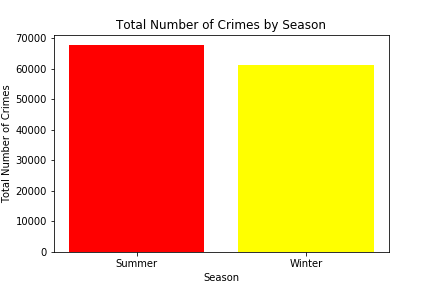
**Do different times of the year affect the number or types of crime committed? What is the total number of crimes by month?**

Initial assumption states that winter season has higher crime rate as compared to the summer season. By looking at the dataset I needed to figure out how to generate results from it. The data did not include the parameters that were required. Had to create a new column with named months from the reported date.

Used that to generate the plot shown below to see if winter does have a higher crime rate.



The graph shows a fluctuation in the total’s over the months. But still does help us determine if winter has a higher crime rate. Had to simplify the data more by grouping months into summer – warmer months (Apr, May, Jun, Jul, Aug, Sep), and winter – cooler months (Oct, Nov, Dec, Jan, Feb, Mar). But the following findings negated our assumption that winter does not have a higher crime rate.



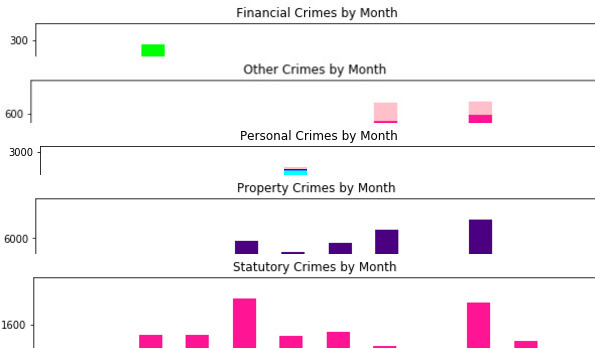
Summer has a higher rate in total for the crimes as to winter, but the difference is small about 500 in total, then rose the question if this small difference does hold any significance or not. So, I ran the chi-square analysis, with a critical value of 5.99 and a statistical value of 326.3, and p-value 6.25 \*10^-73. With p-value < 0.05, it does hold true that this small difference does matter a lot.

Then ran the chi-square test by quarter and month to see if the change in month matters as well or not, and the findings show that the p-value < 0.05 for both.

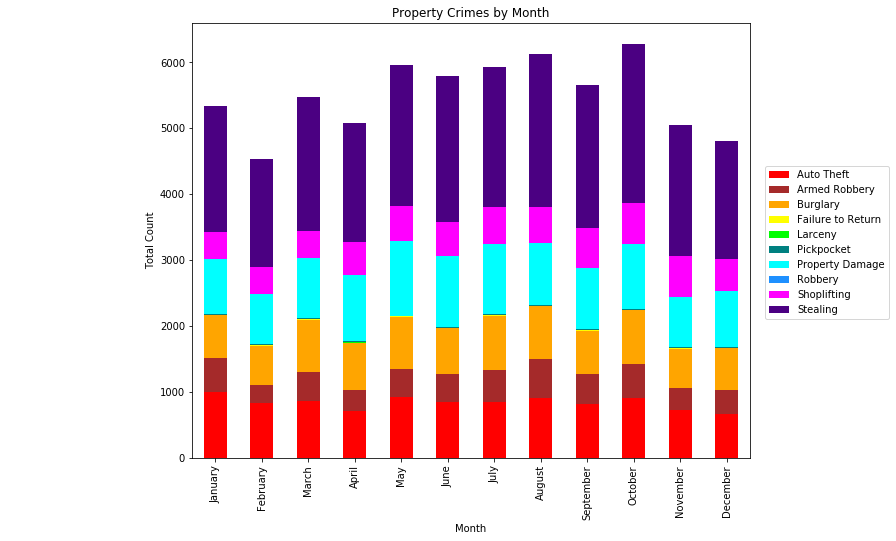
**What is the total of different categories of crimes by month?**

Next, we wanted to look at the total of each crime month. Took a lot of time to sort out the data because it had a total of 122 crime types, with the same crime described in different ways. Condensed the types to about 66 in total. Still it was hard to present so many types on one visual, so had to break down those types into categories; personal, property, statutory, financial, and other.

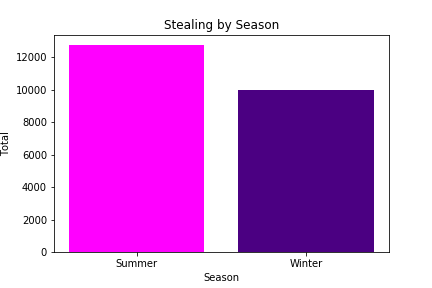
Among these categories’ property crimes has the highest total.



Property Crime is defined by the following type of crime:



In the above plot we can see that auto theft, property damage, and stealing are the most common crimes committed. Stealing is the top most committed crime, which holds our assumption true. But it is hard to determine by looking at the visual above that it was in higher in winter or not. So, to determine if stealing rate was higher in winter summer, the total for stealing had to broken down further by season.



The above figure further negates the assumption that crime rate for stealing was higher in winter. Further statistical analysis still has to be run in order to run if this difference does hold any significance but, because it is irrelevant to the assumption which was just to determine if the ‘stealing(larceny/theft) by young males during the “Winter” months’ holds true.

**Does the age or gender of the people involved in crime play a role in determining the greatest participants in criminal activities?**

Yes, we discovered after investigating and plotting the data that our assumption on Gender/Age was proven correct in that the majority of the crimes being committed were by young males in the 25-35 year old age bracket. The large unknown amount of data in the unknown column was that due to the nature of data being recorded. A victim of crime may not have known the criminals age or gender who committed a crime against them when the crime was reported and recorded by the police. Resulting in that there is a sizeble amount of data that cannot be quantified to add in support or disproving our assumption.

**Where is crime most likely to occur in the city limits of Kansas City, MO?**

Crime is most likely to occur in the CPD- Central and EPD, - East, Metro police division of the city which are all urban areas in the center of the Kansas City, MO metro area. After investigating the data of which police division reported the most crime and confirming those police divisions geographical locations from the ArcGIS : KCMO PD Division Stations data source map we concluded that part of our assumption was supported in that the majority of crimes would occur in urban areas.