When and Where to Worry:

an Analysis of Crime in Salt Lake City

By Jared Harmer and Bryce Fairbanks

Introduction

Salt Lake City, a bustling metropolis full of enticing restaurants, magnificent sights, and a bustling nightlife; a city which people from many walks of life call home. Salt Lake City is culturally and economically diverse, a fact which is desirable, but also brings adverse consequences, one of which is a high crime rate. There were 612,000 crimes committed in Salt Lake City from 2007-2016 [citation needed]. Our goal in this project is to use linear algebra to calculate linear regression lines, comparing crime rates to different variables, such as time of day or the homeless population in the area. We will also set out to explain the linear algebraic principles behind linear regression. Next, we will use the linear regression lines to find correlations between the data points and hypothesize different causes to the high crime rates and ways to lower the probability of having a crime committed against you in Salt Lake City.

The Math Behind Linear Regression

Linear regression is a very useful statistical tool, used by data scientists in order to find the extent that two variables are correlated; that is to say, whether an increase or decrease in one variable is mirrored by a similar increase or decrease in the other variable. There are multiple ways to calculate a linear regression line, but an efficient and simple way to calculate it is through using linear algebra.

To start out, imagine you have three instances of two variables, plotted onto the x-y plane. For an example, let's label these data entries (0, 1), (1, 2), and (2, 4). You would like to find an equation of a line which goes through each of these points, but they don't lie on a straight line. The next step would be to find the "best fit" line, or the line which deviates from the data points the least. We want a line equation of the form

$$y = ax + b. y = ax + b.$$

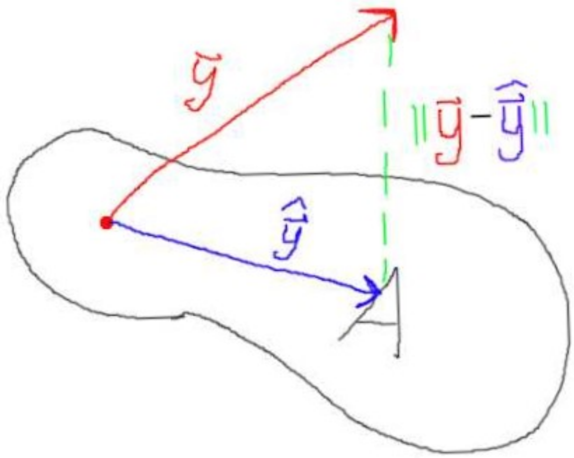
Plugging each point into the equation yields the results:

$$1 = 0a + b$$
$$2 = 1a + b$$
$$4 = 2a + b.$$

This then becomes a system of linear equations, which can be solved with matrix algebra. The matrix equation of $A\vec{x} = \vec{b}$ is

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Obviously, this doesn't have a solution, as the points do not lie on a straight line. In order to adjust for this, we need to find a $\hat{\vec{y}}$ that corresponds to the columns of A. In other words, we need to manipulate $\vec{y}$ so that it lies in the column space of A.$\hat{\vec{y}}$ that corresponds to the columns of A. In other words, we need to manipulate $\vec{y}$ so that it lies in the column space of A.



We can't just choose any point in A and call that $\hat{\vec{y}}$; we want the solution to be as close as possible to $\vec{y}$. Thus, $||\vec{y}-\hat{\vec{y}}|| < ||\vec{y}-\vec{b}||$ should be true for all $\vec{b}$ which satisfy $A\vec{x}=\vec{b}$. Apply the Best Approximation Theorem to find $\hat{\vec{y}}$, using $\hat{\vec{y}}=\ proj_{\text{Col A}}\vec{y}$.$\hat{\vec{y}}$; we want the solution to be as close as possible to $\vec{y}$. Thus, $||\vec{y}-\hat{\vec{y}}|| < ||\vec{y}-\vec{b}||$ should be true for all $\vec{b}$ which satisfy $A\vec{x}=\vec{b}$. Apply the Best Approximation Theorem to find $\hat{\vec{y}}$, using $\hat{\vec{y}}=\ proj_{\text{Col A}}\vec{y}$.

$\hat{\vec{y}}$ is in the column space of A, so there is some $\hat{\vec{x}}$ such that $A\hat{\vec{x}}=\hat{\vec{y}}$. Due to the Orthogonal Decomposition Theorem, we know that $\vec{y}-\hat{\vec{y}}$ is orthogonal to $Col(A)$, and $\vec{y}-A\hat{\vec{x}}$ is also orthogonal to $Col(A)$. Let $\vec{v_i}$ be a column in A. $\vec{v_i}\cdot(\vec{y}-A\hat{\vec{x}})=0$ and $\vec{v_i}^T(\vec{y}-A\hat{\vec{x}})=0$. We can then write that $\vec{v_i}^T=A^T$, so $\hat{\vec{y}}$ is in the column space of A, so there is some $\hat{\vec{x}}$ such that $A\hat{\vec{x}}=\hat{\vec{y}}$. Due to the Orthogonal Decomposition Theorem, we know that $\vec{y}-\hat{\vec{y}}$ is orthogonal to $Col(A)$, and $\vec{y}-A\hat{\vec{x}}$ is also orthogonal to $Col(A)$. Let $\vec{v_i}$ be a column in A. $\vec{v_i}\cdot(\vec{y}-A\hat{\vec{x}})=0$ and $\vec{v_i}^T(\vec{y}-A\hat{\vec{x}})=0$. We can then write that $\vec{v_i}^T=A^T$, so

$$A^T(\vec{y}-A\hat{\vec{x}})=0\,A^T(\vec{y}-A\hat{\vec{x}})=0$$

$A^T\vec{y}-A^TA\hat{\vec{x}}=0$, and finally $A^T\vec{y}-A^TA\hat{\vec{x}}=0$, and finally

$$A^T A\hat{\vec{x}} = A^T\vec{y} \; A^T A\hat{\vec{x}} = A^T\vec{y}$$

We now have a formula to find $\hat{\vec{x}}$. Applying this to our example above, we have the equation$\hat{\vec{x}}$. Applying this to our example above, we have the equation

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \hat{\vec{x}} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix} \widehat{\vec{x}} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

$$\widehat{\vec{x}} = \begin{bmatrix} 3/2 \\ 5/6 \end{bmatrix}.$$

Finally, plugging in $\hat{\vec{x}}$, we get

$$\widehat{\vec{y}} = A\widehat{\vec{x}} = \begin{bmatrix} 5/6 \\ 7/3 \\ 23/6 \end{bmatrix}$$

Thus, our least squares regression line is $y = 3/2x + 5/6$, and our approximate data point is (5/6, 7/3, 23/6).

Methodology

We wanted to calculate the number of crimes committed against individuals in Salt Lake City. In order to properly analyze the crime data, we removed any crimes that did not affect or endanger individuals.[1] We also removed any crimes that were responded to by Salt Lake City police officers but were not located inside the borders of Salt Lake City. Our data is from the 2016 Salt Lake City Crime Database, which recorded around 50,000 crimes that occurred during the year. They recorded the location of the crime as GPS coordinates, so we converted them to street addresses.

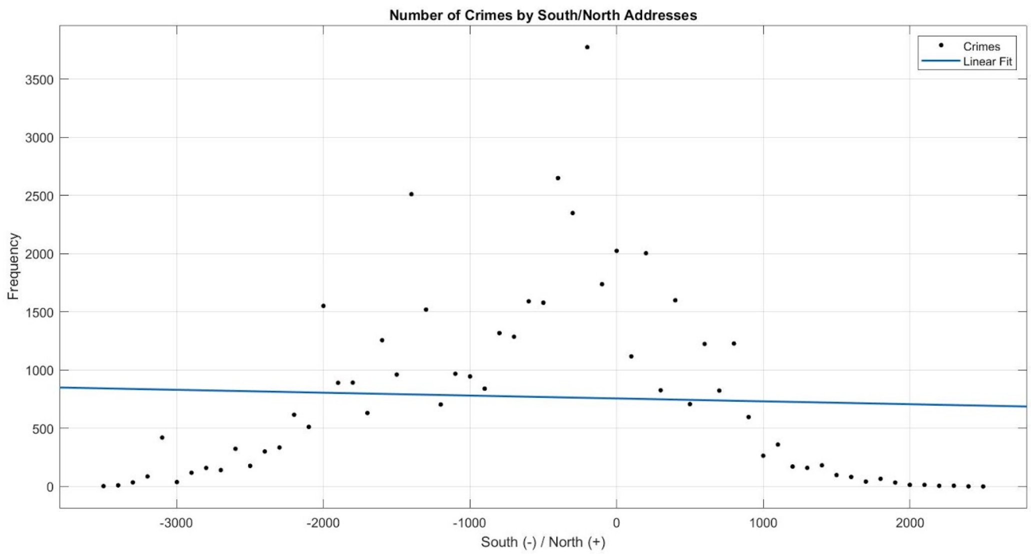
We used different types of regression lines, the first being a simple linear association (as described earlier) and the second being a more complex fit (the derivation for which goes beyond the scope of this project).

The factors we looked at were location, time of day, day of week, and time of year in relation to crime.

[1] The crimes that were included in this analysis were arson, assault, burglary, damaged property, DUIs, escape, exploitation, extortion, fleeing, fraud, hit and run, homicide, invasion of privacy, kidnapping, larceny, property crime, against public order or peace, robbery, stolen property, stolen vehicle, drug or alcohol-related traffic injury, threats, and weapon offenses.

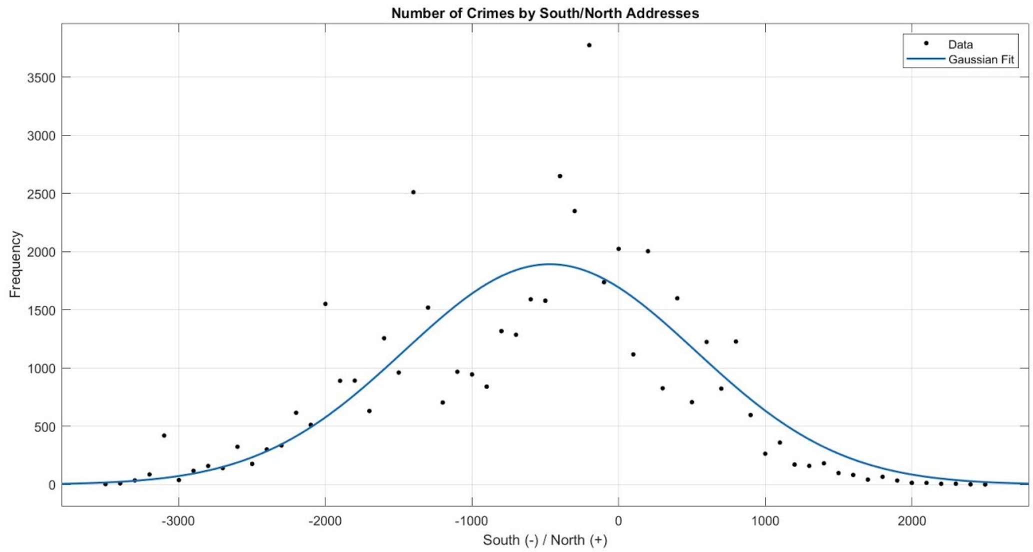
Crime by Location in 2016

North-South Addresses:



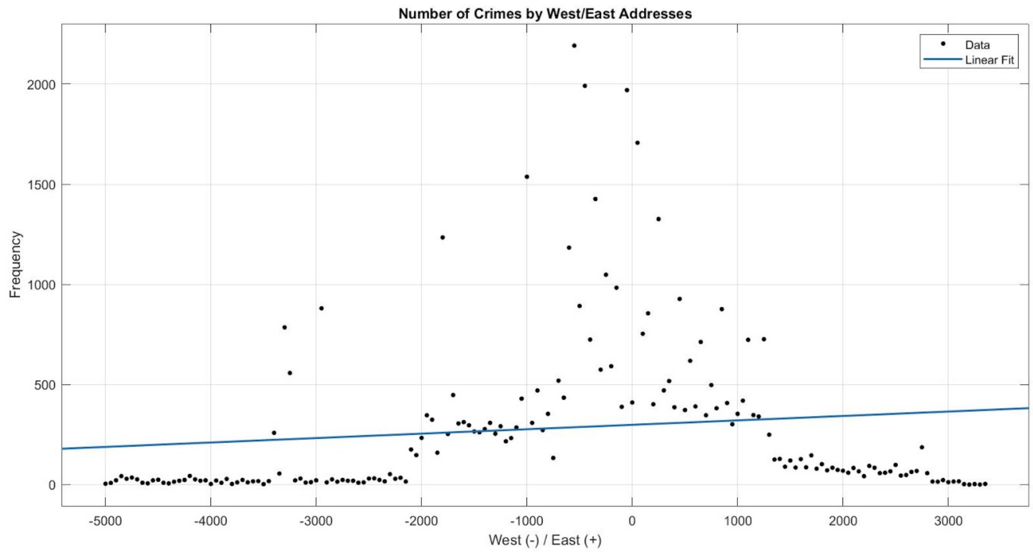
Salt Lake City is set up in a grid system. The center of the city is at 0N, 0E, and from the center, the numbers increment by 100. Our graphs reflect this system: a point on the x-axis at -500 means that the crime occurred around 500 South Street. 1,000 indicates 1000 North Street. Negative numbers indicates south, positive indicates north. The y-axis shows the total number of crimes committed during 2016.

Unfortunately, linear regression does not give us a clear picture of the data. Luckily there are other regression lines which can better approximate the data. The Gaussian regression line is a much better fit for the data.

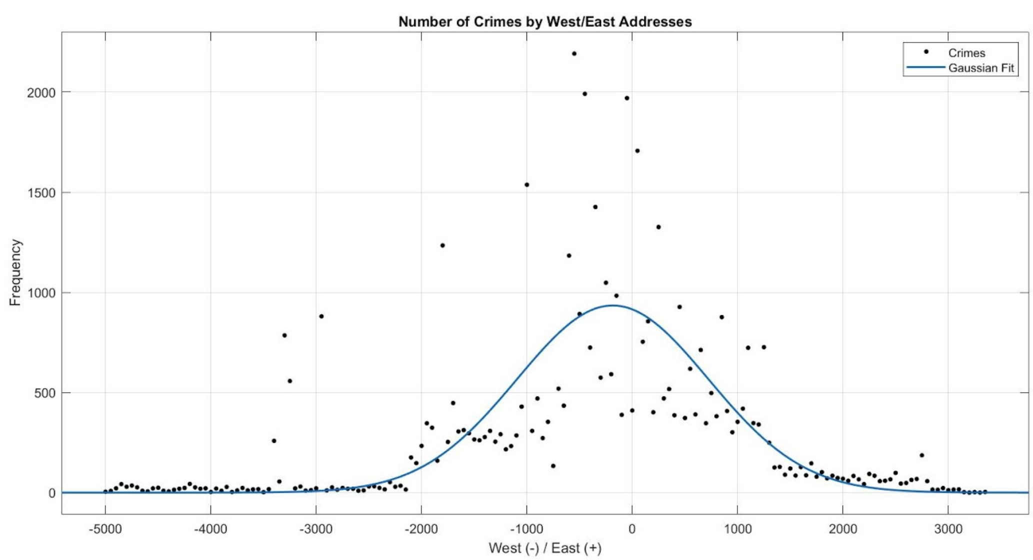


Crimes are mostly centered around 500 S, with the highest data point being at 200 S, 3800 crimes having occurred there over 2016. The amount of crimes decreases as you move towards the edges of the city; likely because there are fewer people in the outskirts of the city than in the center.

East-West Addresses



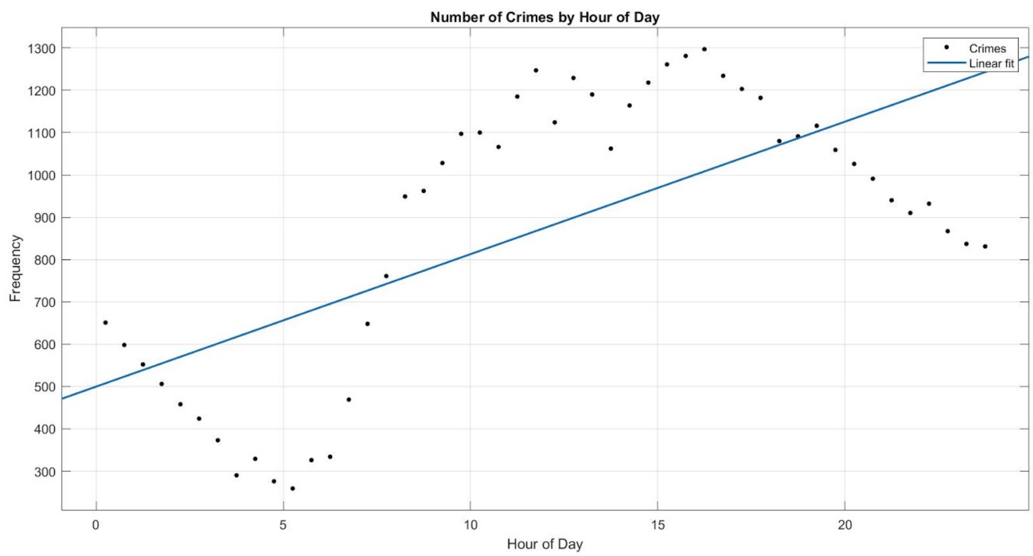
Again, the linear regression line does not do the best job in approximating the data. However, we can see a slight positive trend in the line. This is probably due to the larger number of West addresses compared to East addresses, rather than a higher amount of crimes being committed on the East side of town.



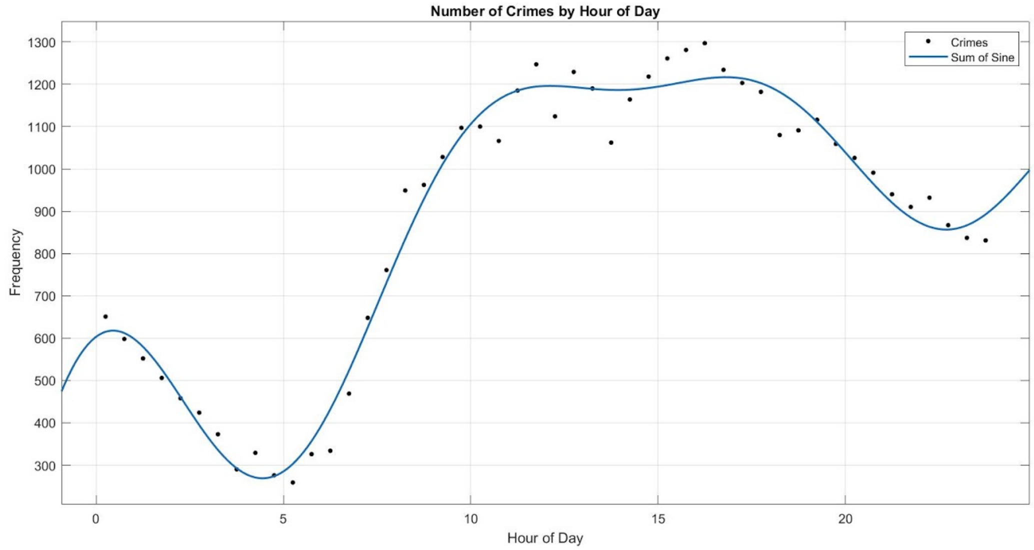
From the Gaussian curve, we can see that crimes are mostly centered around 200 West Street, decreasing as the streets increase in number. The greatest amount of crimes occurred at 500 W., where there were around 2,300 crimes which occurred during 2016. This averages out to about 6 crimes a day. There is more noise in this graph than in the South/North graph, which suggests more crimes being committed away from the center of the city in the West/East direction than in the South/North direction.

Crimes by Time of Day in 2016

Next, we compared the crimes which occurred at a certain time of day. For the graph, a point on the x-axis at 5 indicates a time around 5:00 A.M. A point around 20 indicates a time around 8:00 P.M. The points are separated into 30-minute increments, the first one starting at 0:00 and ending at 0:29. The y-axis again is the total number of crimes which occurred at a specific time during 2016.



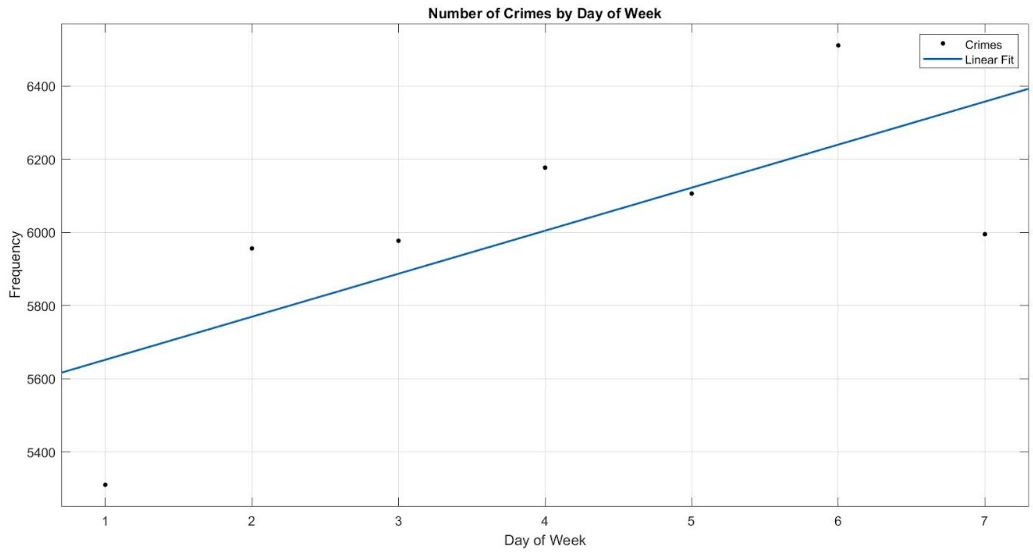
The Linear Regression Line does a better job at approximating the time of day data, but it still is not incredibly accurate. We can see an increase in crimes committed as the day progresses.



For a better approximation of the data, we used a sum of sine approximation. This gives us a better view of the times with the most crimes committed and the time with the least crimes. We can see that most crimes occurred in the middle of the day, between 12:00 and 6:00, and tapered off in the early morning. The least amount of crimes occurred just after 5:00 A.M., and the most crimes were committed around 4:00 P.M.

Crimes by Day of Week in 2016

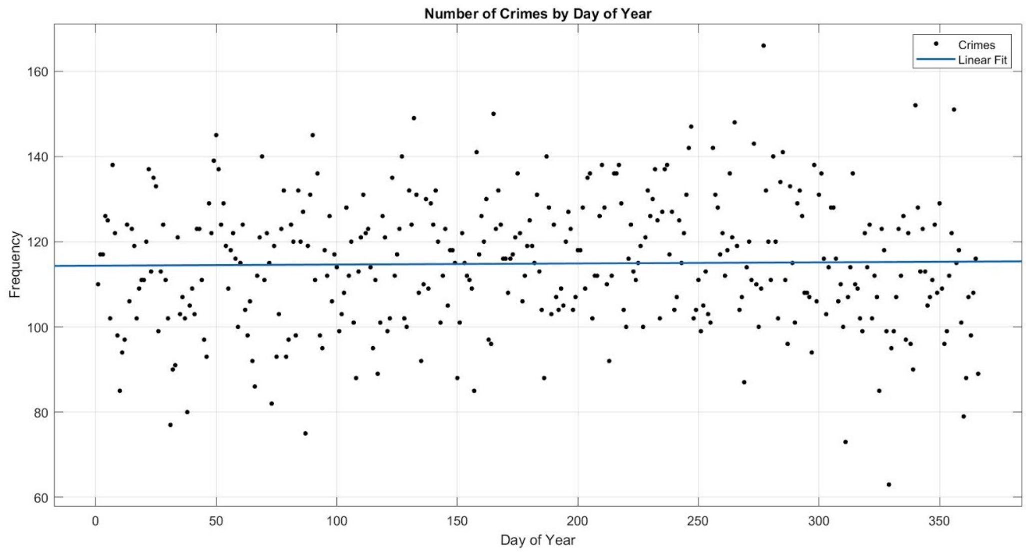
The x-axis for this graph is measured from 1 - 7, 1 being Sunday and 7 being Saturday.



The Linear Regression Line does an impressive job of modeling these points. We see that crimes generally increase during the work week, and taper off during the weekends. This is interesting, as we saw in the above graph that most crimes occurred during the time period from 12:00 to 6:00, which is during the workday for most people. Friday seems to have the highest amount of crimes during 2016, coming in at about 6,500 crimes over 2016. Sunday only had approximately 5,300 crimes

Crimes by Day of Year

The x-axis is scaled from 1 to 365, with each number cooresponding to the number of day, starting at January 1st and ending at December 31st.



Unlike the previous graphs, there doesn't appear to be a clear pattern in this data. There is a slight decrease in crime rates toward the end of the year, but overall the crime rate seems to be constant throughout the year.

Conclusion

From the data, we can construct the most dangerous place and time to be in Salt Lake City, where you are most likely to have a crime committed against your person. Combining the data, we see that the most crimes occurred at 200 S 500 W at 4:00 P.M. on Friday. This location is understandable, as there is a homeless shelter, soup kitchen, and men’s night club in the immediate vicinity. There are likely more factors contributing to this number, but these places are likely large influences on this number. So if we learned anything from this analysis, it is to avoid 200 S 500 W at 4:00 P.M. on Friday. Or, if you’re looking for trouble, you know the place and time to go.

Bibliography

Brownlee, Jason. “How to Solve Linear Regression Using Linear Algebra.” *Machine Learning Mastery* (blog), March 4, 2018. https://machinelearningmastery.com/solve-linear-regression-using-linear-algebra/.

Lay, David C., Steven R. Lay, and Jude J. McDonald. *Linear Algebra and Its Applications*. 5th ed. Pearson, 2015.

“Linear Regression | Statistically Significant Consulting.” Accessed May 7, 2019. https://www.statisticallysignificantconsulting.com/RegressionAnalysis.htm.

“Mortgage Tips & Advice| GOBankingRates | GOBankingRates.” Accessed May 7, 2019. https://www.gobankingrates.com/category/loans/mortgage/.

Ph.D, Andrew Chamberlain. “The Linear Algebra View of Least-Squares Regression.” *Medium* (blog), December 10, 2016. https://medium.com/@andrew.chamberlain/the-linear-algebra-view-of-least-squares-regression-f67044b7f39b.

“SALT LAKE CITY - POLICE CASES -2016 | Open Data | Socrata.” Accessed May 7, 2019. https://opendata.utah.gov/Public-Safety/SALT-LAKE-CITY-POLICE-CASES-2016/trgz-4r9d.