Per the Blight Authority definition, Urban blight is the process by which a previously functioning city, or part of a city falls into disrepair and decrepitude. The removal of blight is critical to the health, safety and welfare of a community.

The questions we pondered at the beginning of the project include:

1. Are there economic indicators indicative of blight?
2. Do property values decline prior to blight?
3. Does the percentage of vacant properties increase in a community undergoing blight?
4. Does reinvestment in a community contribute to the lessening of blight?
5. Can you predict blight based upon an increase in vacant buildings?

A concentration of vacant properties can be an indicator of blight. We investigated potential data sources from the St. Louis City Open Data project to determine if there is evidence within the data that a concentration in vacant buildings results in a decline in property values.

Ultimately, we utilized residential tax data and parcel data from <https://www.stlouis-mo.gov/data/parcels.cfm>. The file was downloaded from the St. Louis City Open Data website (<https://www.stlouis-mo.gov/data/>) in an Access database format. The tables within the database were very large. Therefore, the team used PANDAS and a Jupyter Notebook to extract the parcel (PRCL) and residential data (PRCLREAR) tables out of the Access database into two separate CSV files. The CSV file outputs were named Prcl.csv and residential\_data.csv. Due to the size of the files, we limited each CSV file to include only fields necessary for our data analysis. All unnecessary fields were deleted from each file.

An ID was created in each file through concatenation of the city block, parcel and owner code data fields. The records in each file were validated to ensure they were unique. Duplication of the ID existed within the residential data file due to the fact that it contained multiple years of tax data. This was not an issue for the team as the group by function would be utilized to parse the data by year. Additionally, the team excluded all commercial data from the files to ensure the analysis only focused on residential data.

After all data cleansing activities were complete, the team imported the parcel and residential tax data into a final Jupyter Notebook. This notebook was utilized to read in both CSV files and ultimately create an inner join on the ID field. An inner join was utilized as the team intended to merge all data from both files where the ID matched. All merged data was put into a new dataframe variable called merg\_df.

After the merged dataframe was created, the team created individual dataframes for each tax year that was to be included in the analysis. Tax year information for 2015 through 2018 was individually extracted from the merged dataframe. The group by function was used to group all data by zip code for each tax year. 29 zip codes were represented in the dataset. This corresponds to the 29 zip code within the St. Louis City limits. The group by allowed the team to calculate the number of vacant properties by zip code in the data set.

Next the team created a scatterplot with vacant properties by zip code by year. 2018 had the greatest number of vacant buildings with 1,106 total vacant properties. From the scatterplot, it appears that nearly half of the zip codes had an increasing number of vacant properties in 2018 compared to prior years. From this scatterplot, there is a concentration of zip codes that have more than 40 vacant properties and a concentration that less than 20 vacant properties. It is plausible that the zip codes with the greatest number of vacant buildings have a lower tax base that is potentially declining.

The next scatterplot the team created was a comparison of average property tax values by zip code by year. From the scatterplot, it appears that many of the zip codes experienced increasing property values. Specifically, 19 properties had increasing values when comparing 2018 to 2015. There were 24 zip codes that had at least 1 vacant property. However, a significant portion of the zip codes had increasing property values. This could be due to a reinvestment in the community and the rehabbing of vacant properties. However, analysis of individual zip codes does not necessarily validate this fact. In the 5 zip codes with the highest concentration of vacant properties, the property tax value was increasing in 3 of the 5 zip codes.

Per the first bar chart in the study, more than half of the zip codes had a gain in the number of vacant properties of 2018 vs. 2015. However, from the second bar chart, it does not appear that all the zip codes with vacant property increases had declines in average property tax value. Actually, more than half of the zip codes with increases in vacant properties had gains in property tax value. Potentially this could be due to rehabbing activity or reinvestment in the properties to cause some gain in value.

A comparison of vacant properties versus non-vacant properties for the years 2018 and 2015 shows there was a slight increase (%0.18) in vacant properties in 2018. The surprising finding from the information in the study was that an overwhelming majority of residential properties in the dataset were non-vacant.

Based on the evidence within the study, increases in vacant properties alone does not necessarily indicate blight. This may be one factor that contributes to blight. However, there are certainly other variables that may more profoundly predict a community going into blight. Other variables may be increases in crime rates, the percentage of community members at or below the poverty line or potentially changes in unemployment rates.