

# **“A Tale of Three Cities”:**

## **Comparing Chicago, Phoenix, and San Diego**

*Jorge Bonilla, Kolby Devery, Victor Inyang, Kodzai Nyakurimwa*

W200 Project 2

[https://github.com/UC-Berkeley-I-School/Project2\\_Bonilla\\_Devery\\_Inyang\\_Nyakurimwa](https://github.com/UC-Berkeley-I-School/Project2_Bonilla_Devery_Inyang_Nyakurimwa)

### **Datasets:**

#### Housing Data:

[https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI\\_AT\\_metro.txt](https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_AT_metro.txt)

#### *Variables:*

- City
- Year
- Quarter
- House Price Index
- Median House Price

#### Crime Data (Accessed via REST API):

<https://crime-data-explorer.fr.cloud.gov/pages/home>

#### *Variables:*

- City
- Year
- Aggravated-assault
- Burglary
- Larceny
- Motor-vehicle-theft
- Homicide
- Rape
- Robbery
- Arson
- Violent-crime
- Property-crime
- Crimes Per One Hundred Residents

#### Homelessness Data:

[https://www.huduser.gov/portal/datasets/ahar/2020-ahar-part-1-pit-estimates-of-homelessness-in-the-us.html?utm\\_source=HUD+Exchange+Mailing+List&utm\\_campaign=ca0950288a-hud\\_releases\\_2020\\_annual\\_homeless\\_assessment&utm\\_medium=email&utm\\_term=0\\_f32b935a5f-ca0950288a-19646315](https://www.huduser.gov/portal/datasets/ahar/2020-ahar-part-1-pit-estimates-of-homelessness-in-the-us.html?utm_source=HUD+Exchange+Mailing+List&utm_campaign=ca0950288a-hud_releases_2020_annual_homeless_assessment&utm_medium=email&utm_term=0_f32b935a5f-ca0950288a-19646315)

#### *Variables:*

- City

- Year
- Gender
- Age
  - Race
  - Sheltered
  - Unsheltered

Weather Data(Accessed via REST API):

<https://www.ncdc.noaa.gov/cdo-web/webservices/v2#data>

*Variables:*

- City
- Year
- Quarter
- Average low temperature
- Average high temperature
- Average number of days with precipitation

## Supplemental Datasets:

Consumer Price Index:

<https://www.bls.gov/cpi/data.htm>

*Variables:*

- City
- Year
- Annual CPI

Detailed City-level Crime Data (Google Drive):

<https://drive.google.com/drive/folders/1gCmic-l6ofCMBgniFSJyj68WQgEI7-hj>

- Chicago
  - 7,442,165 rows
  - 22 columns
  - 1.64 GB
- Phoenix
  - 394,065 rows
  - 7 columns
  - 42.7 MB
- San Diego
  - 154,284 rows
  - 22 columns
  - 74 MB

US Census Median Household Income Data (Accessed via REST API):

<https://www.census.gov/data/developers/data-sets/acs-1year.html>

*Variables:*

- City
- Year

- Median Household Income

US Census Population Data (Accessed via REST API):

<https://www.census.gov/data/developers/data-sets/acs-1year.html>

<https://www.census.gov/data/developers/data-sets/decennial-census.html>

*Variables:*

- City
- Year
- Population

## Housekeeping:

Before starting the data analysis portion of our project, it is important that we perform a series of sanity and data completeness checks on our data sources. This is to ensure all data will be coherent and passes the gut check as well as ensuring the data is complete enough to offer compelling insight.

*Sanity Checks:*

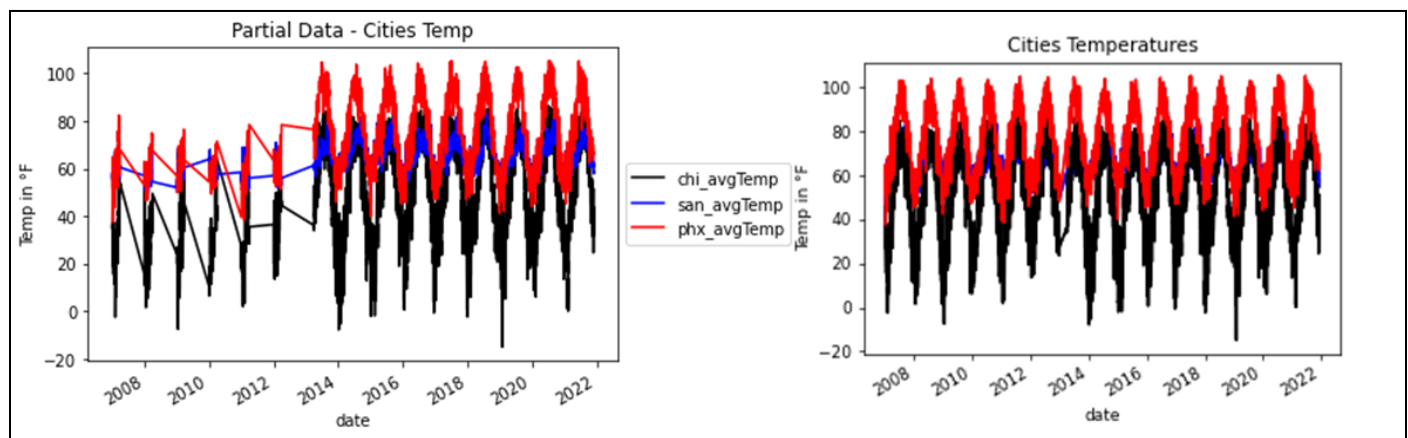
The FBI REST API provides us with annual data on standard crime categories across these three cities for every year since 2007. We found the summarized FBI data to be stable with minimal data cleaning required since there were almost no null or missing values for the major crime categories such as Property Crime, Burglary, Larceny, and Violent Crime. Additionally, the FBI data provides us with a uniform set of crime categories allowing us to compare crime rates per 100,000 individuals across three different cities. This also helps us to investigate the correlation to other metrics in the study such as weather, homelessness, and median home prices.

*Data Completeness checks:*

When pulling NOAA weather data from the REST API, we ran into a problem that we could not retrieve data between 2007 and 2013. After some research in the data structure, we were able to pull values from 2007 to 2013, but that data had a different format than the one for 2013 onwards. After plotting the data, we noticed that the data prior to 2013 was inconsistent, as shown on the trend line graph in Figure 1, titled, “*Partial Data - Cities Temp*” with the rest. See figure below.

After some investigation, we discovered that the NOAA REST API for weather was only providing the daily maximum and minimum weather values between 2007 and 2013. After 2013, the weather values were provided as daily averages so we had to transform data prior to 2013 from maximum and minimum values to average values. We did this by taking the average of the maximum and minimum daily values so that the data would be consistent with the weather values beginning in 2013. The results from our data cleaning efforts are shown in Figure 1 below, graph titled, “*All Data - Cities Temp*” The data is consistent with the US season, showing cold months with low temperatures and summer months with high temperatures, the graph shows consistency in data from 2007 to 2021.

Figure 1



Sanity checks were also performed on homelessness data to check the min and max values for data to make sure there aren't any negative or null values present.

Detailed crime data for each city contains null values and missing fields. For instance, the data for Chicago exhibits null values for the “Location Description,” “X-Coordinate,” “Y-Coordinates,” and “Location” fields. With null values we proceed to drop these rows as the complete address where the crime took place has been completely anonymized. Similarly, the crime data for San Diego also contains crimes for neighboring cities which fall outside of the city boundaries, or for crimes that took place in a different county. For these records, we filter the results from the dataframe so that we work with crimes that have been reported exclusively in the city of San Diego.

## Introduction:

We have seen massive social and economic changes in the United States over the last couple decades. From the recovery of the 2008 recession to the COVID pandemic, many states and cities have grappled with a variety of challenges and face many more to come. Our team chose 3 cities across the country - Chicago, Phoenix, and San Diego - and would like to explore variables such as housing markets, crime, homeless, and weather trends of these places since 2007. This information will be helpful to individuals who are contemplating relocating or investing in real estate.

With a population of nearly 3 million, Chicago is located off the coast of Lake Michigan and is currently listed as the third largest U.S. city. Phoenix is the capital city of Arizona with a population of roughly 1.6 million and located in the Southwest. It is currently listed as the fifth largest U.S. city. San Diego is a coastal city in Southern California with a prominent military base. It is currently the eighth largest U.S. city, having a population of roughly 1.3 million.

With this, we hope to shed light on:

1. How have median home prices in Chicago, San Diego, and Phoenix changed and compared over time?
  - a. How does the change in median home price compare to change in the Consumer Price Index?
    - i. Median Home prices
      1. CPI Price Index
  - b. How does the median home price compare to the median household income?
2. How do the populations of these 3 cities compare?
3. How has crime per one hundred thousand people in these three cities changed and compared over time?
  - a. In each city, what are the top 5 crime categories?
4. How much per capita homelessness exists in each city?
5. How closely correlated are median home prices across San Diego, Phoenix, and Chicago to per capita homelessness?
6. How does average annual weather correlate to per capita homelessness cities?
7. What is the correlation between crime and homelessness?
8. What is the correlation between crime and weather?
9. How does the median home price of a city compare to its average temperature?
10. How closely are annual crimes correlated to home prices? And what types of crimes have the biggest impact on home prices?
11. In each city, where are the crime hotspots and how do they compare against each other?

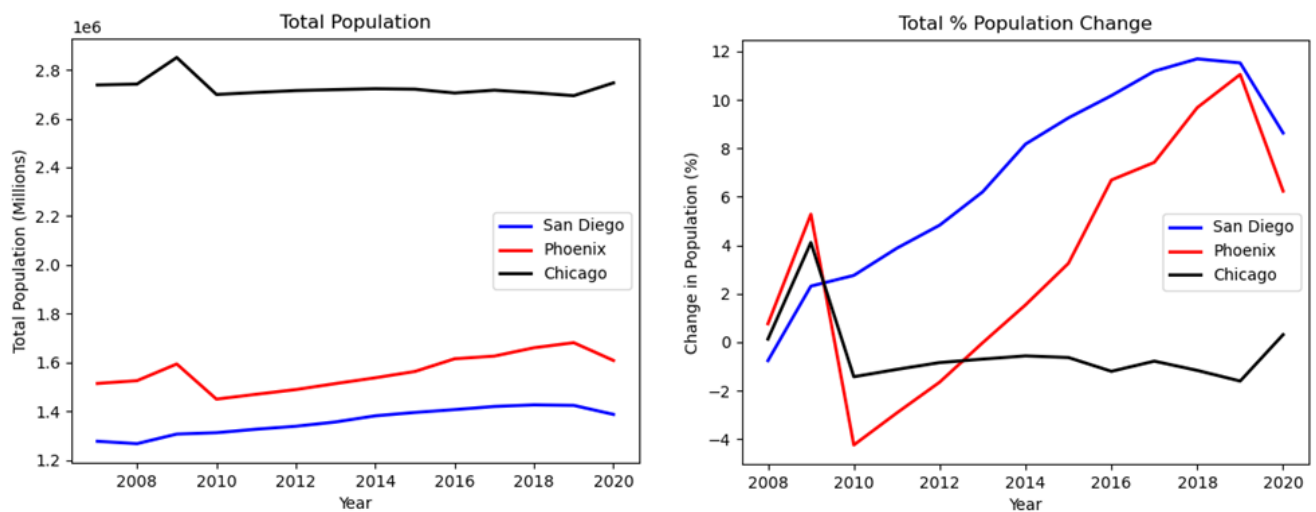
## Population and Life Indices:

In this section, we used data from the US census to better understand the population and life indices of these three cities. Life indices are calculated through the consumer price index (CPI), which is a weighted average of market goods over time.

### How do the populations of these 3 cities compare?

As seen in the figure below, Chicago has a significantly higher population than both San Diego and Phoenix (almost double). However, we can see that Chicago's population has remained stagnant, while Phoenix and San Diego have seen a net growth in population. The left plot of Figure 1 demonstrates the actual population values over time, while the right plot shows the relative population growth of each city since 2007. San Diego has seen the highest amount of growth in this time, with a peak of 12% growth since 2007. This is followed by Phoenix with a peak of 11% growth and Chicago with almost no net growth.

Figure 2



Interestingly, the Consumer Price Index (CPI) for San Diego is over 4 times higher than that of Chicago and almost 8 times higher than that of Phoenix. The higher the CPI, the more expensive market goods are to purchase, which means it is likely far more expensive to buy everyday goods in San Diego than either Phoenix or Chicago, as shown by the left plot in Figure 3. The right plot of Figure 3 demonstrates that San Diego's CPI is also increasing faster than either city. There is a correlation here between CPI and the total population of Phoenix and San Diego, but Chicago's CPI is increasing despite its population stagnation. This makes it difficult to discern what effect there is, if any, of population growth on the cities' CPI.

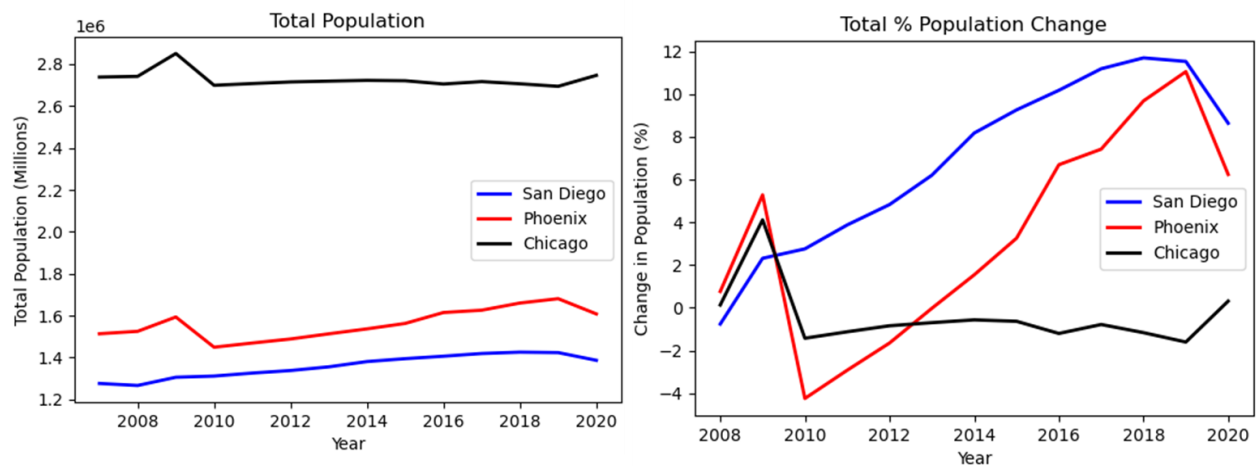
## Home Prices:

In this section, home price data was collected from the Federal Housing Finance Agency and compared with household income data collected by the US census.

### How have median home prices in Chicago, San Diego, and Phoenix changed and compared over time?

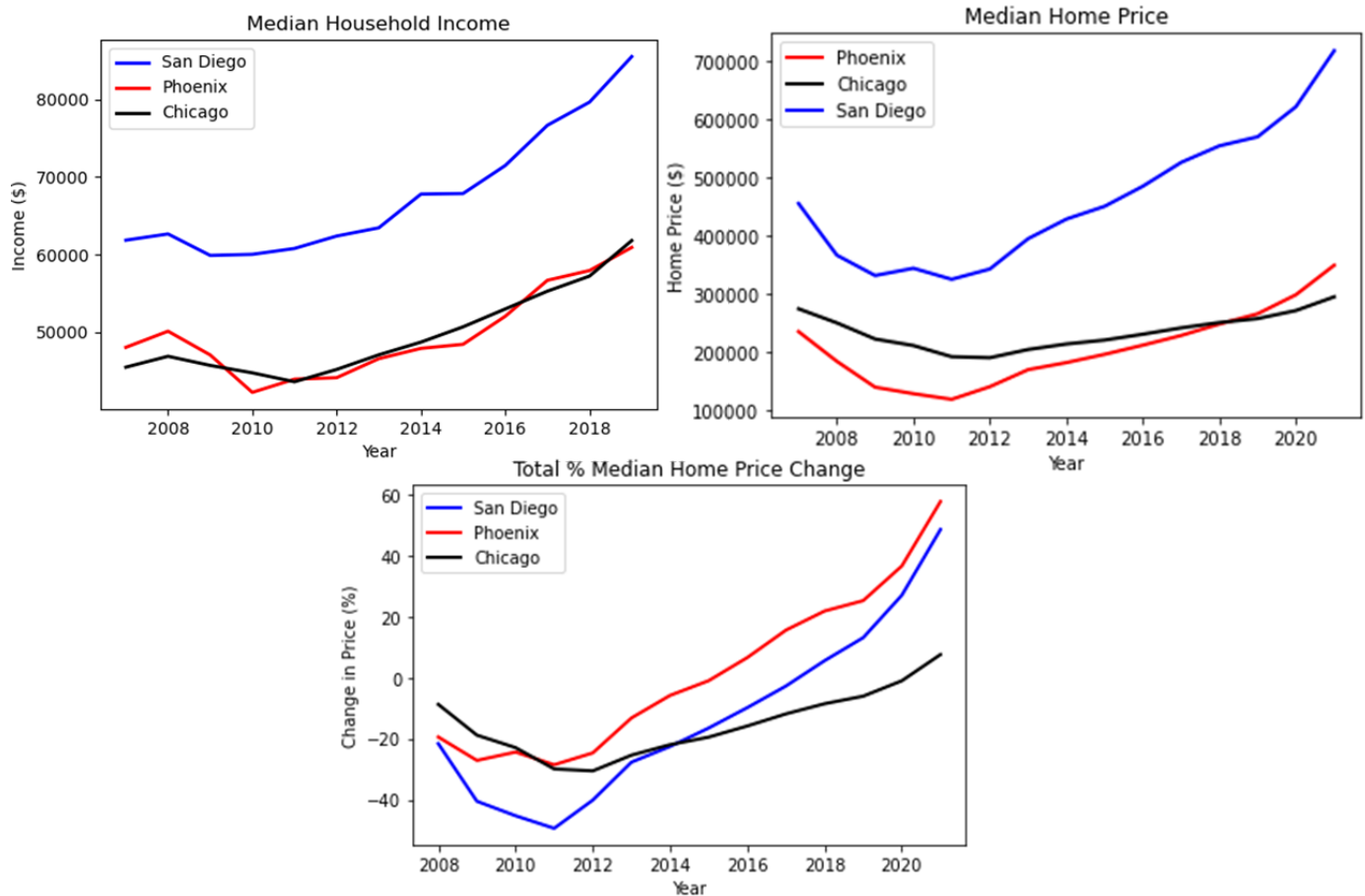
Like CPI, San Diego is also leading in median household income

Figure 3



while Phoenix and Chicago are relatively matched, as shown in Figure 4. One point of interest is the median home price graph compared to median household income. It seems that median household income and median home prices are very tightly correlated among these 3 cities. Visually, these graphs look incredibly similar. However, Phoenix in recent years is experiencing a housing boom where Phoenix's relative house price is increasing the fastest. As shown by the top right plot of Figure 4, Phoenix surpassed Chicago's median home price in 2018 and continues to rise.

Figure 4

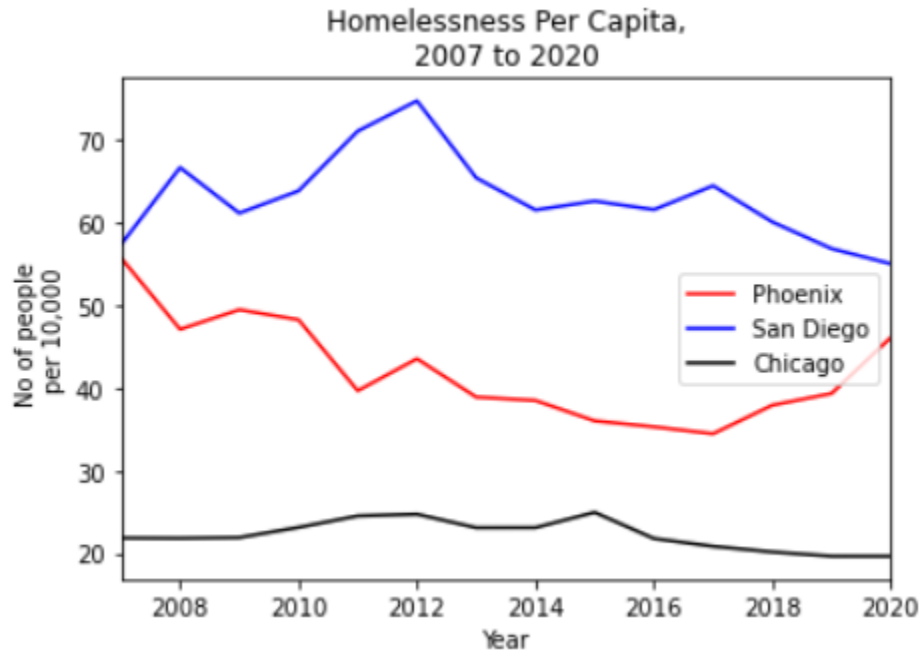


## Homelessness:

Data was collected from the U.S. Department of Housing and Urban Development, the estimated homelessness report provides national, state, and County-level estimates of homelessness, as well as estimates of chronically homeless persons, homeless veterans, and homeless children and youth. The data was compared to median home price and average weather to explore if there is a correlation between them



Figure 5



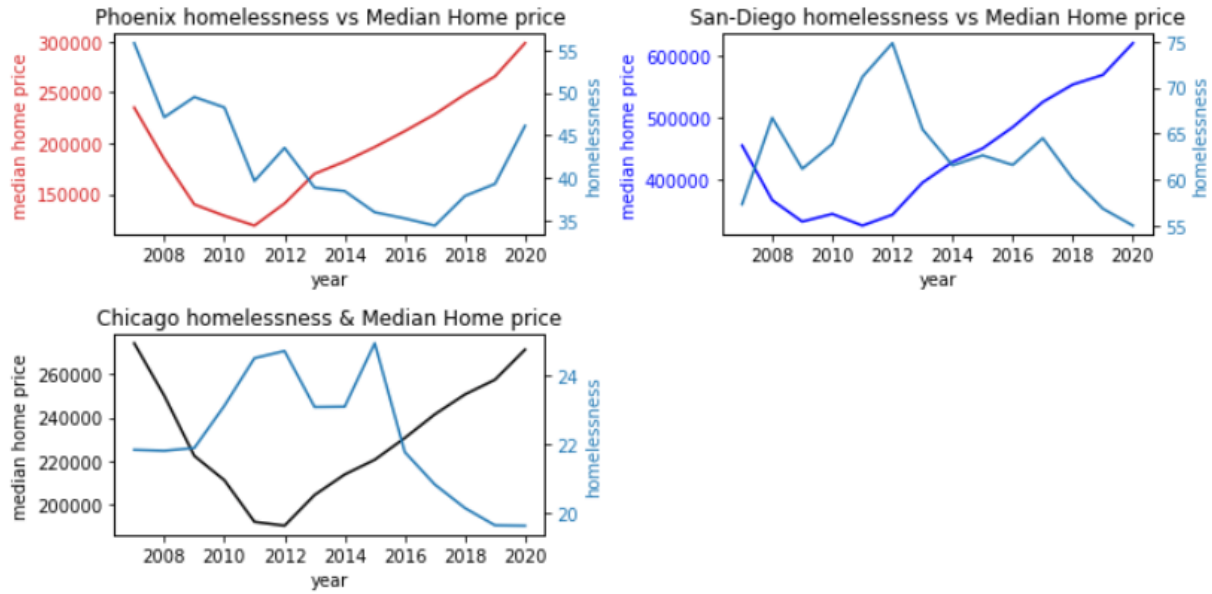
#### How much per capita homelessness do we have in each city?

As seen in the figure(5) above, San-diego has the highest number of homeless people per capita, followed by Phoenix and Chicago. California has notability ranked as one of the top states with the highest homelessness people. It's no surprise to see that san-diego tops the list here. According to Forbes, San-diego ranks 7th with the city with the most homeless people, followed by Phoenix. Phoenix over the last few years has seen an increase in the number of homeless people compared to the other two cities. A good assumption could be that due to the increase home prices coupled with the disruption caused by the COVID-19 pandemic.

#### How closely correlated are median home prices across San Diego, Phoenix, and Chicago to per capita homelessness?

In figure(6) and the correlation matrix below we see that there is a correlation between homelessness and median home prices, especially in phoenix, in 2017 the homeless population grew in phoenix. It is important to note that there are other factors that may have led to the homelessness such as job loss, domestic violence, mental health and divorce which was not explored in this analysis.

Figure 6



	Median Home Price in San-Diego	Homelessness per Capita in San-Diego
Median Home Price in San-Diego	1.000000	-0.722326
Homelessness per Capita in San-Diego	-0.722326	1.000000

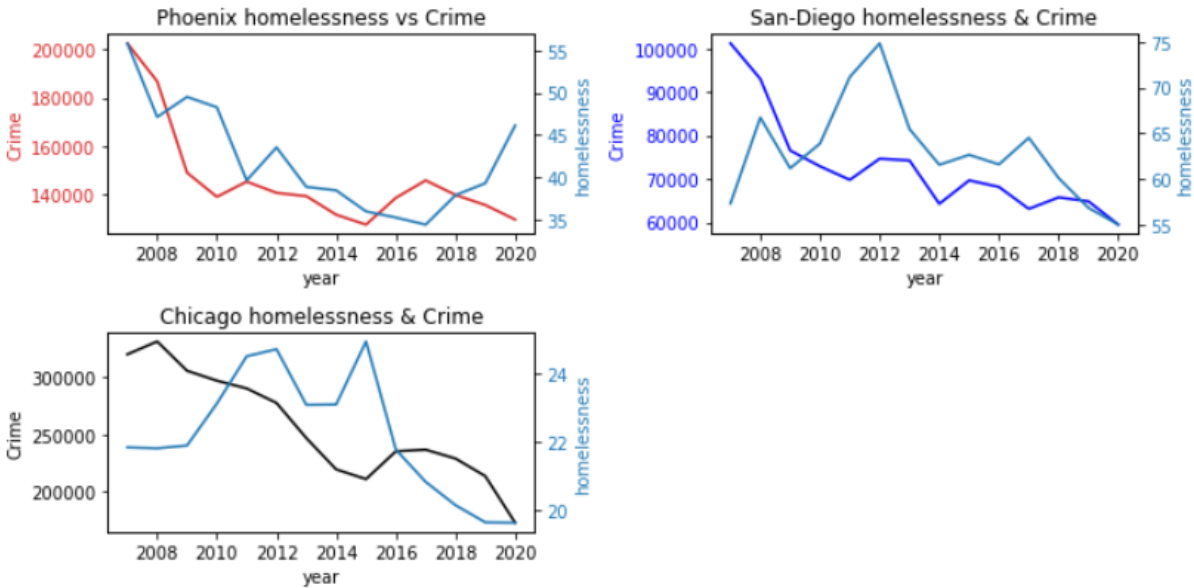
	Median Home Price in Phoenix	Homelessness per Capita in Phoenix
Median Home Price in Phoenix	1.000000	-0.105093
Homelessness per Capita in Phoenix	-0.105093	1.000000

	Median Home Price in Chicago	Homelessness per Capita in Chicago
Median Home Price in Chicago	1.000000	-0.822553
Homelessness per Capita in Chicago	-0.822553	1.000000

### ***What is the correlation between crime to homelessness?***

There is a negative correlation between total crime and homeless per capita, this indicates that the lower the number of homeless people the lower the crime in the city and this. Figure 11 on san-diego shows a really strong correlation between total crime and homeleness. This is a good example to use as San-diego has the highest number of homelessness people per capita.

Figure 7



	Crime in Phoenix	Homelessness per Capita in Phoenix
Crime in Phoenix	1.000000	0.666815
Homelessness per Capita in Phoenix	0.666815	1.000000

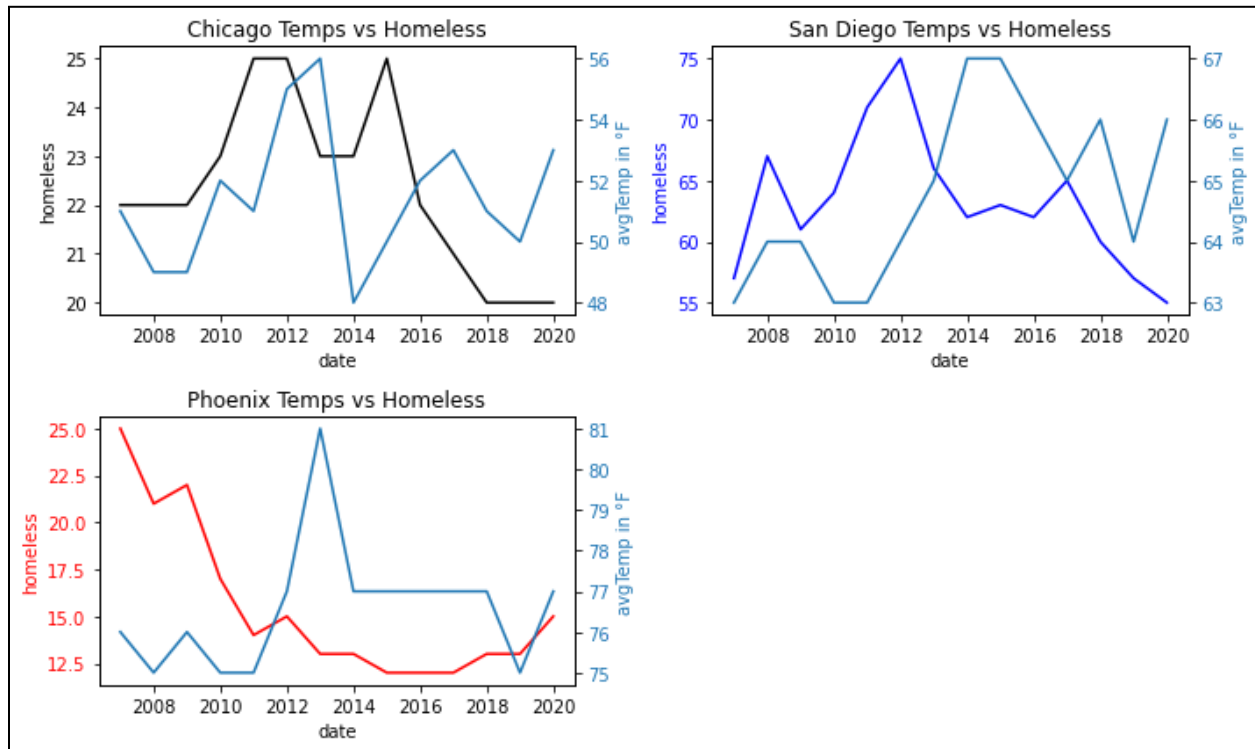
	Crime in San-Diego	Homelessness per Capita in San-Diego
Crime in San-Diego	1.000000	0.11045
Homelessness per Capita in San-Diego	0.11045	1.000000

	Crime in Chicago	Homelessness per Capita in Chicago
Crime in Chicago	1.000000	0.320014
Homelessness per Capita in Chicago	0.320014	1.000000

### ***How does average annual weather correlate to per capita homelessness?***

After going through the data on homelessness and weather, the following insights were illuminated. In San Diego and Phoenix, the numbers of homeless people tend to increase when temperatures are on the cool side. Chicago is the opposite when the temperature dips, the number of homeless people decreases. See the city vs temperature graphs below.

Figure 8

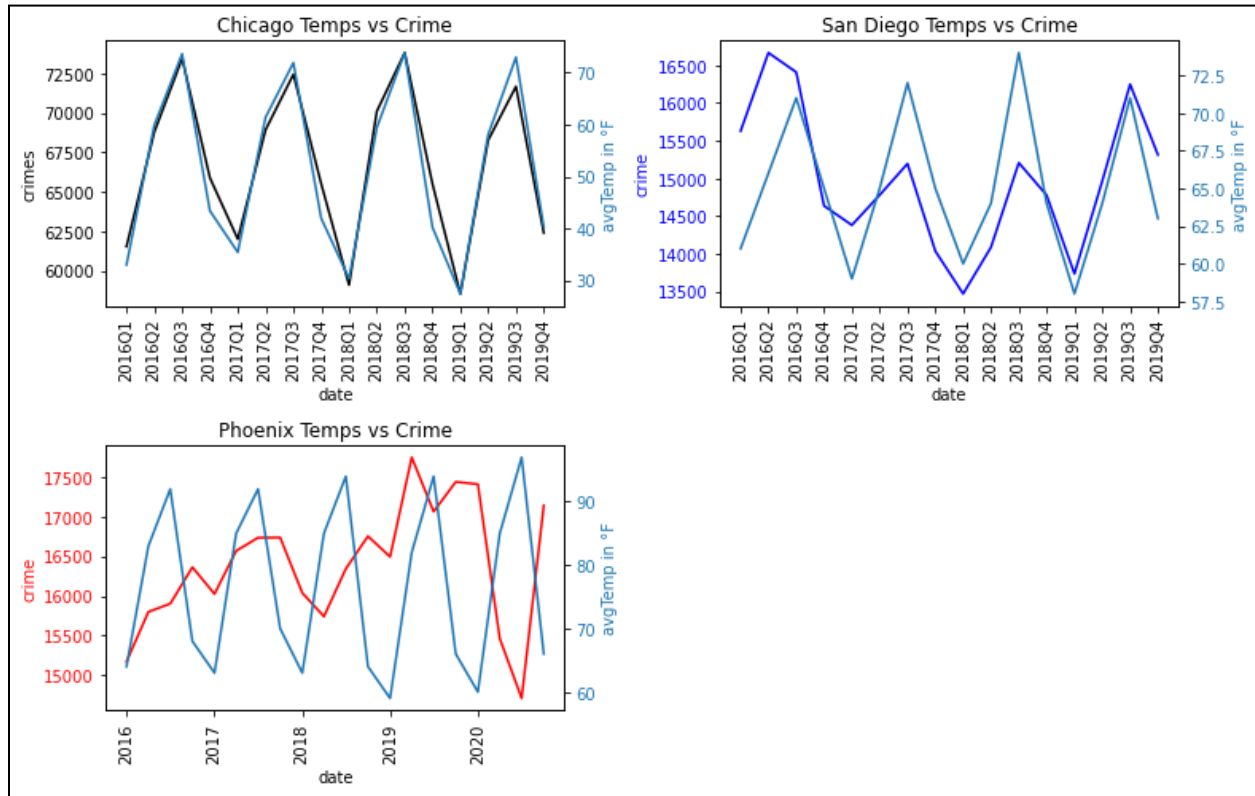


## Weather:

### ***What is the correlation between crime to weather?***

There is a strong correlation between crime and weather. As the temperature gets warm, there is also a rise in the number of crimes committed. In figure 8, Chicago and San Diego graphically show the correlation. On the other hand, Phoenix data, in a way, support this narrative. When it's cold in Phoenix, there are fewer crimes, and also, when it gets scorching hot above 95 degrees, the number of crimes decreases. Most of the crimes are committed between 70 degrees and 90 degrees.

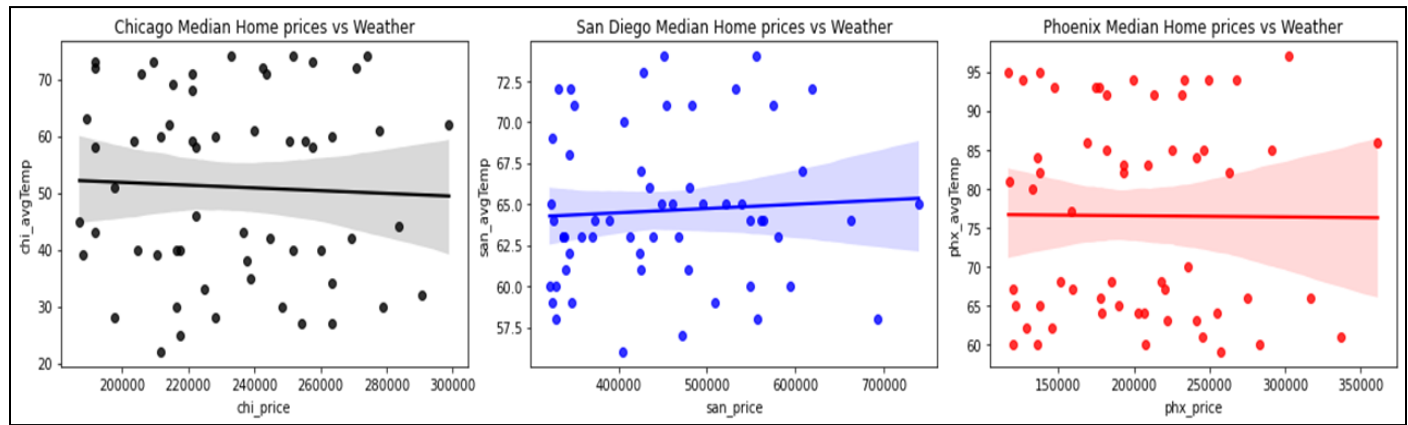
Figure 9



**How does median home price compare to average temperature?**

There is no correlation between median home prices and weather. Which makes sense because the price of the house will not go up or down due to seasonal changes. See figure 12.

Figure 10

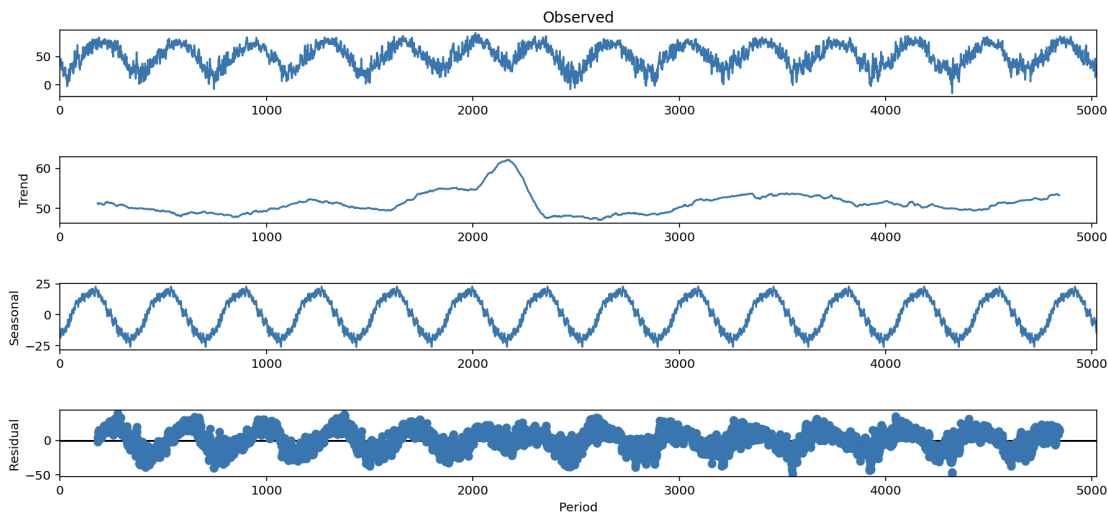


### ***How can the weather data be used for further analysis?***

We can use the weather data to make forecasts about home sales since home sales tend to be correlated with annual seasonal weather. That is, people tend to buy homes during the warmer months than the colder months. First, we would need to remove the seasonality to better understand the cycles and moving averages associated with the weather time series data. These are our results for the weather decomposition analysis for the cities of Chicago, Phoenix, and San Diego:

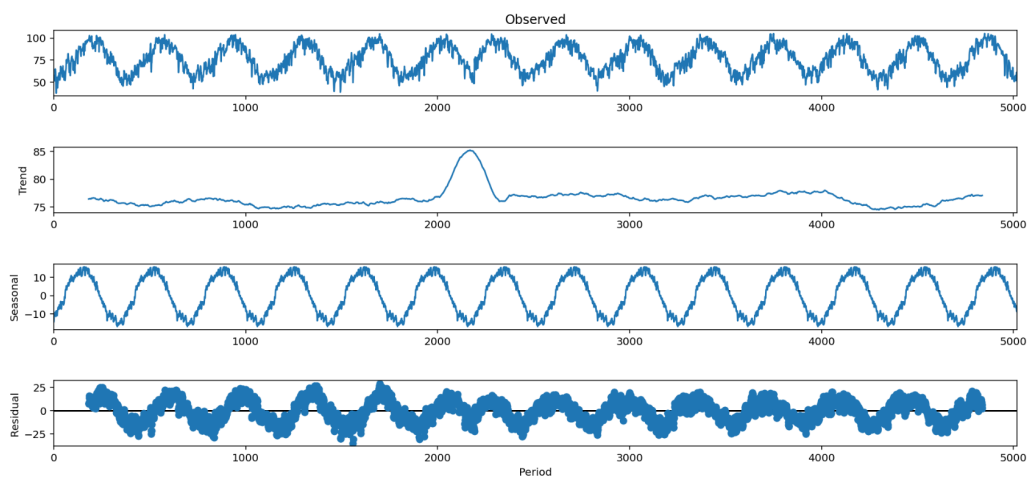
#### ***Chicago Weather:***

Figure 11



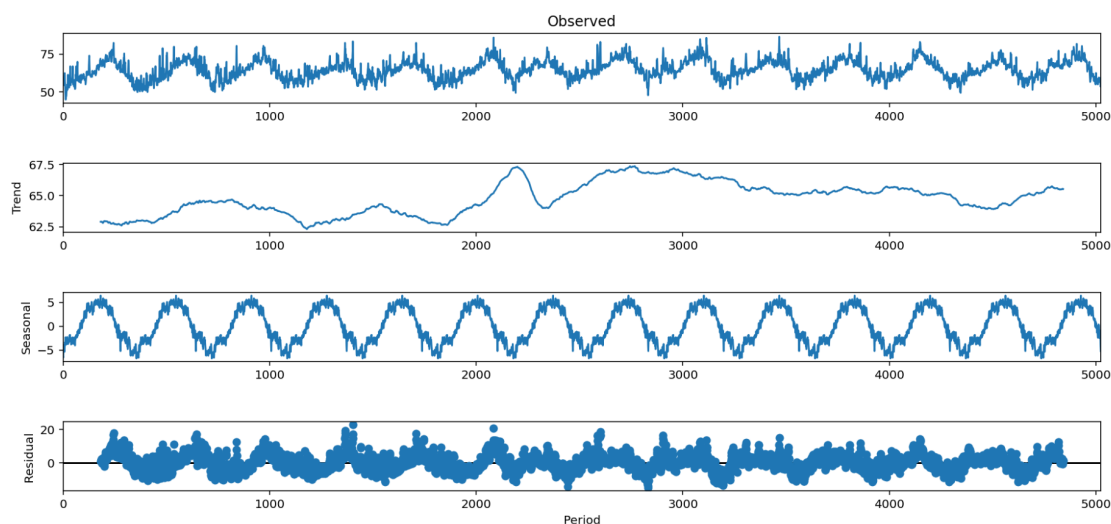
#### ***Phoenix Weather:***

Figure 12



## San Diego Weather:

Figure 13

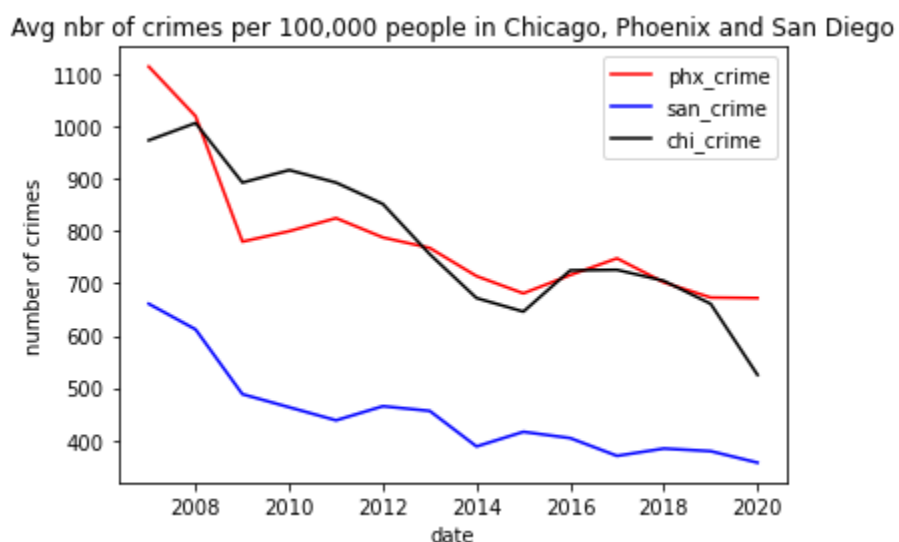


## Crime:

In this section we will present our results from the crime data for each of the three cities. We extracted data from the FBI REST API because it provided a direct comparison of values across equal crime categories for the three cities. The FBI REST API provides this information by allowing us to pull data from the primary police stations in each city: Chicago Police Department, Phoenix Police Department, and San Diego Police Department. We found that some cities may have similar police departments such as the West Chicago Police Department, but such stations are excluded from our data extraction processes. We remain focused on the city's primary police department as our source of crime data. Additionally, we also downloaded detailed crime data sets for each city presenting information for the crimes committed such as the crime classification, the date it took place, and the location.

The FBI crime data allows us to make a direct comparison across three cities by normalizing the crime values per one hundred thousand people across the same crime categories. This gives us insights into how crimes per one hundred thousand residents compare to one another in these three cities since 2007.

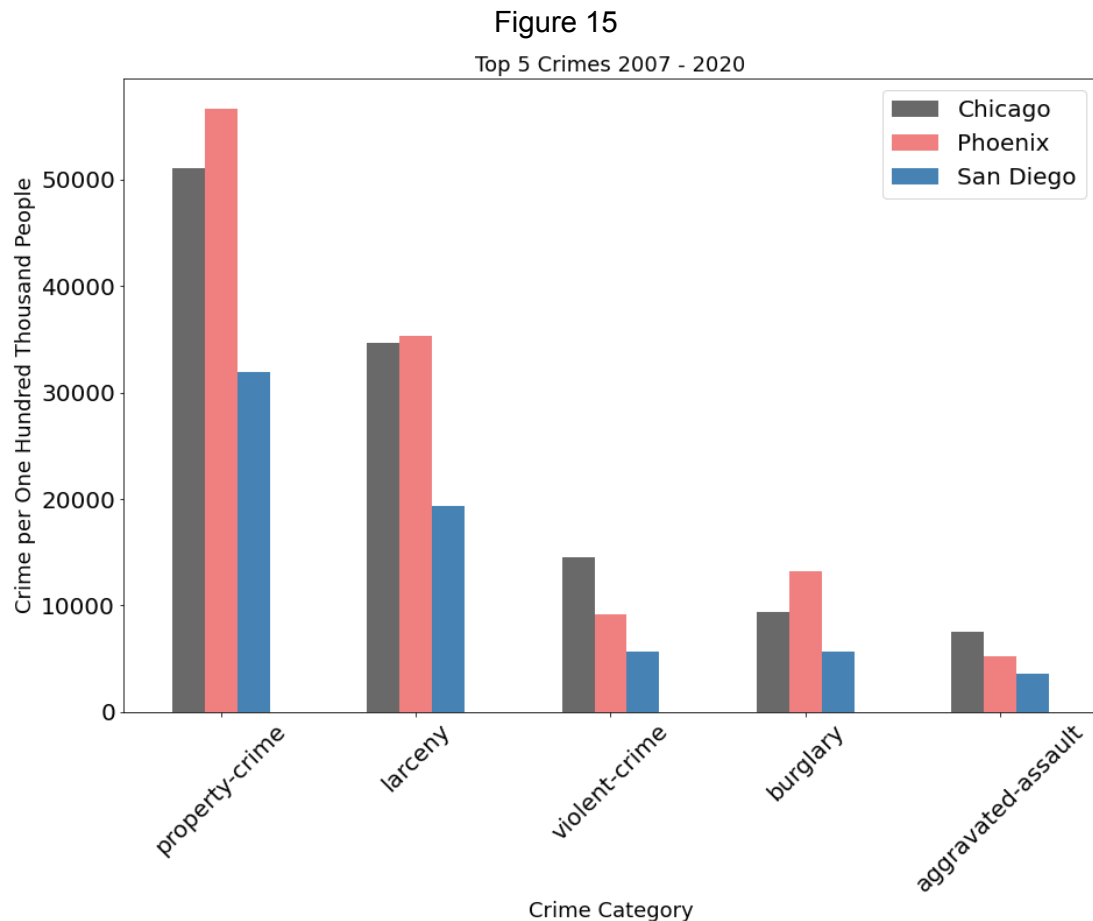
Figure 14



In Figure 14 we aggregate the average number of crimes per 100,000 people across three cities. Chicago and Phoenix have approximately the same annual crime rate with equal annual trends. Since 2007 crimes reported to each police station have been on a decreasing trend which could potentially be attributed to people moving away from these urban centers. Additionally, the population data shows a decline in residents for Phoenix and San Diego during the past two years which could also explain the drop in the number of crimes. The data shows San Diego has the lowest overall crime rate as of 2020 with approximately 400 crimes per 100,000 people while Chicago and Phoenix have crime rates of roughly 600 and 700 crimes per 100,000 people, respectively. Next, we will answer the questions initially posed in this paper:



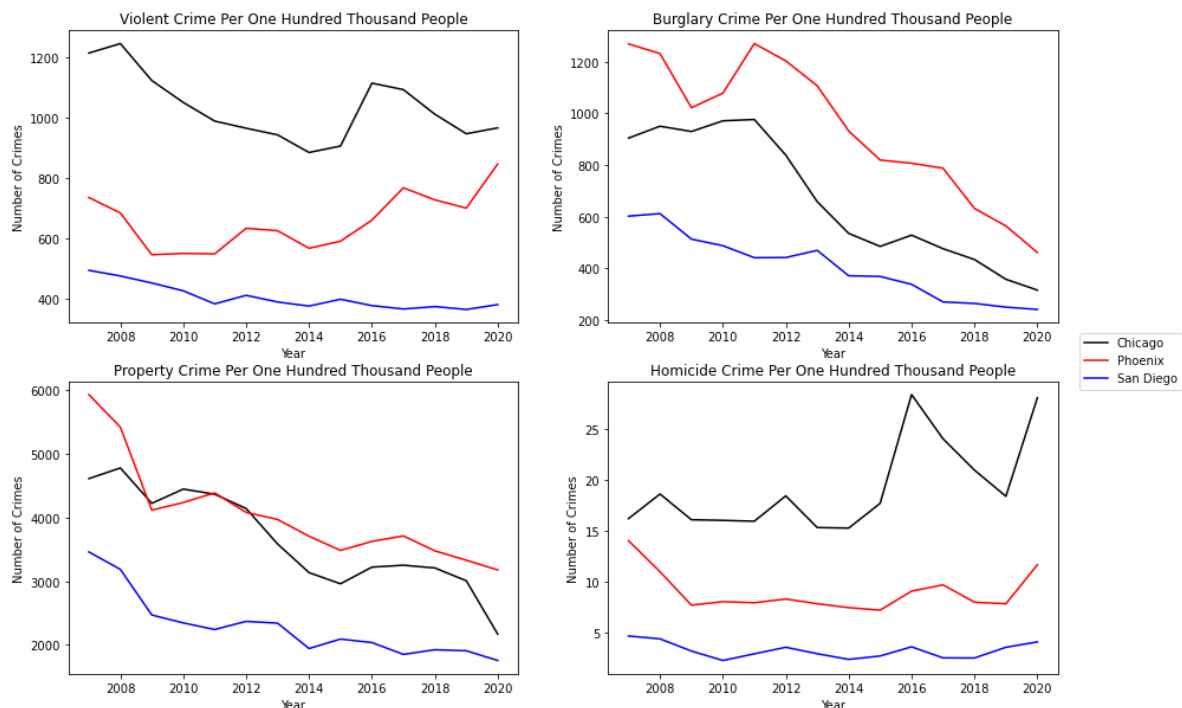
***In each city, what are the top 5 crime categories?***



Among the three cities, property crime and larceny are the most common number of offenses according to FBI data shown in Figure 15. Between 2007 and 2020, Phoenix experienced the most property crimes at 56,662 offenses per 100,00 people, the highest of all cities. Notably, San Diego has the lowest crime as reported to the San Diego Police Department. It is worth noting that San Diego hosts several military bases for the US Air Force, Navy, and Coast Guard with service members commonly seen in the streets among civilians. This raises the question if military presence in a city affects crime rates. The third most common type of crime varies among cities which could be attributed to weather, location, geography, and other factors.

**How has crime per one hundred thousand people in these three cities changed and compared over time?**

Figure 16

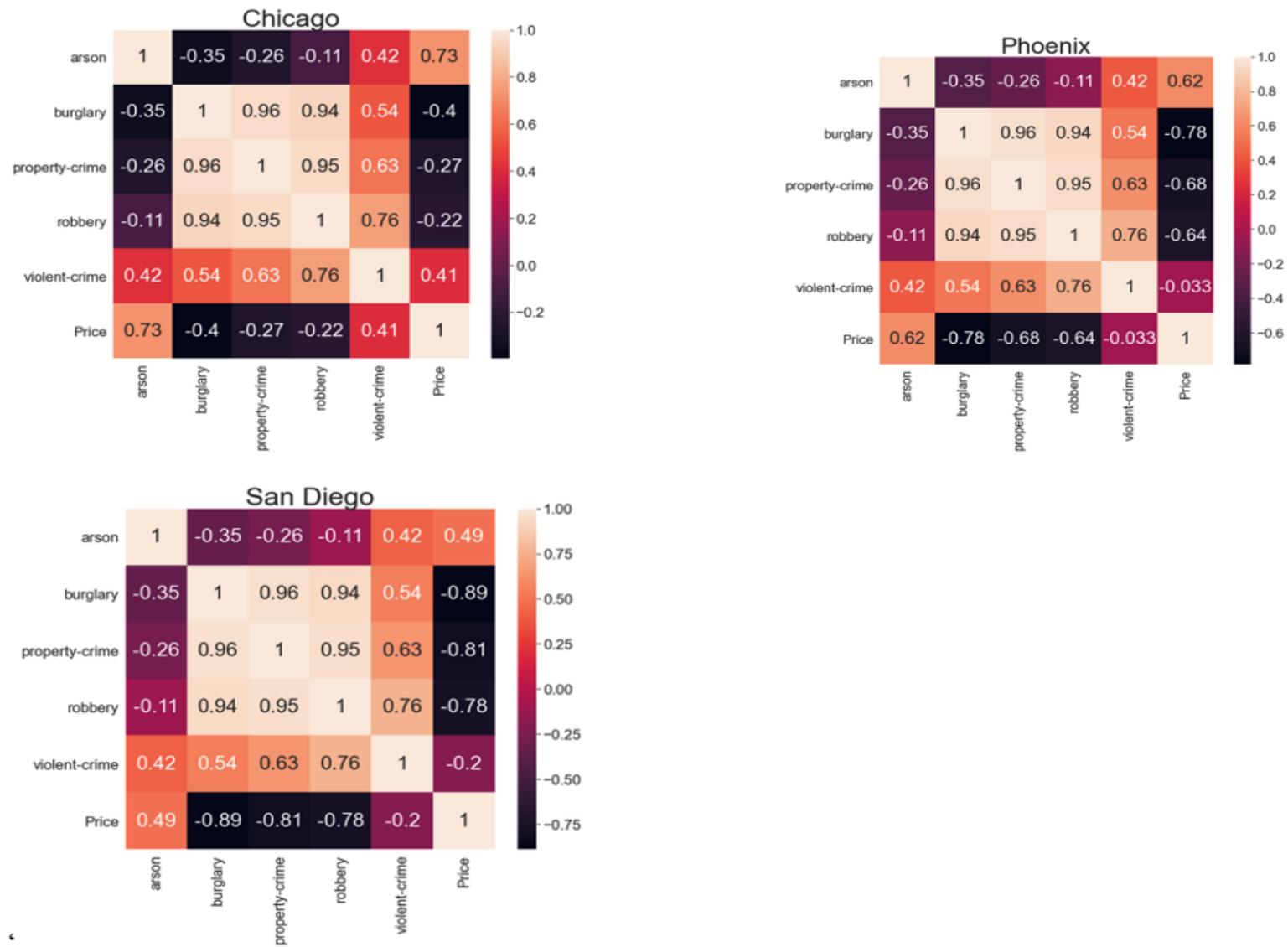


From the FBI data we can discern interesting patterns in the crime dataset and how this is affecting each city. Figure 16 shows that among the four main categories San Diego consistently records the lowest crime levels per 100,000 residents in categories such as burglary, homicide property crime, and violent crime. For violent crimes and homicides the FBI data shows that Chicago consistently registers more of these types of crimes than Phoenix and San Diego. Chicago experienced 1100 violent crimes per 100,000 residents in 2020, while Phoenix and San Diego reported 900 and 400 crimes, respectively. In the FBI's Uniform Crime Reporting (UCR) Program, violent crime is composed of four offenses: murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault. Sadly, we see a recent 25% spike in homicides per 100,000 residents for Chicago over the past two years as compared to Phoenix and San Diego, although previous trends in 2008, 2012, and 2016 show that short-term spikes in homicides appear to be cyclical.

**How closely are annual crimes correlated to home prices? And what types of crime have the largest impact on home prices?**

The correlation matrices presented in Figure 17 in page 19 present evidence that there may be an inverse relationship between home prices and crimes, specifically crimes that involve real estate such as burglary and property crime. These matrices show the strongest correlations between burglary and property crimes and median home prices across all three cities. This information may tell us that as the price of a house increases, the number of crimes decreases, presumably because the property is in a safer neighborhood that experiences less crimes.

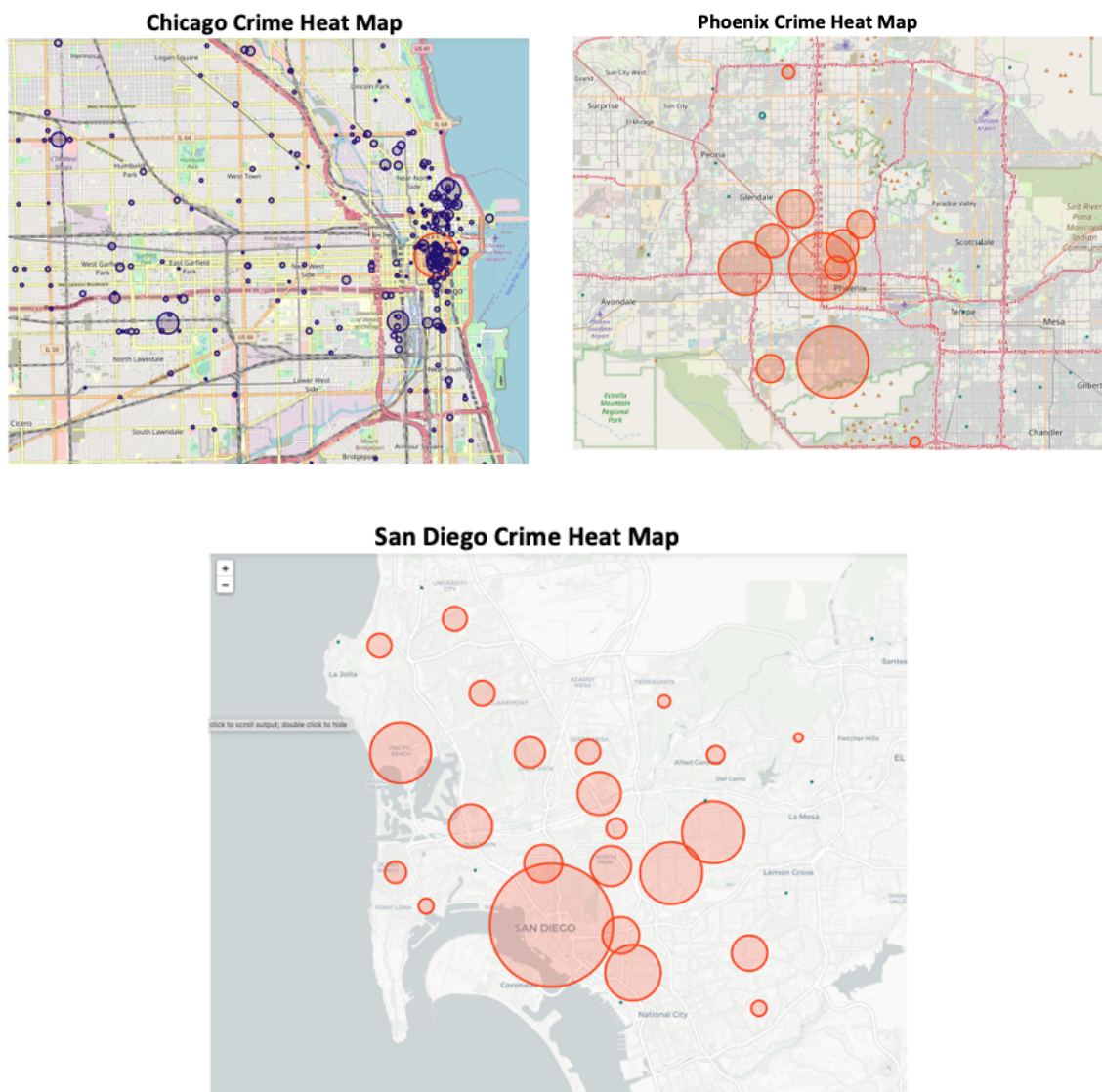
Figure 17



***In each city, where are the crime hotspots and how do they compare against each other?***

As expected, the crime heat maps for each city in Figure 18 show that most of the crimes are clustered around heavily urbanized parts of the city such as the downtown areas as in the case of Chicago, Phoenix and San Diego. This can also be the result of the police stations strategically placed in those areas which report the most crimes due to their proximity to heavily populated areas. This information can be useful to prospective residents who need to have a general understanding of where the high-crime areas are in each of these cities because we present other clusters of crimes that take place throughout the city.

Figure 18



## Conclusions:

- Population and Living Index:
  - Chicago has a very high population, but has remained stagnant compared to Phoenix/San Diego
  - San Diego's CPI is significantly higher than both cities and is increasing faster than both as well
  - There may be a slight correlation between population and CPI - Chicago showed little correlation
- Home Prices:
  - Median home price is strongly correlated to median household income.
- Weather:
  - There is a strong correlation of crime and weather.
  - There is a moderate correlation of homelessness and weather.
  - There is no correlation between home medium prices and weather.
- Homelessness
  - There is a moderate correlation between median home price and homelessness.
  - There is a strong correlation between total crime and homelessness.
  - San Diego has the highest number of homelessnes people per capita followed by Phoenix and Chicago.
  - The past 3 years Phoenix homelessness has been on the rise.
- Crime:
  - Crime rates are similar between Chicago and Phoenix while San Diego has the lowest crime rates as reported by the city's main police stations.
  - Evidence of correlation between crimes and home prices suggests further research is needed in this area.
  - There are clusters of crimes concentrated in urbanized zones as presented in the crime heat maps.