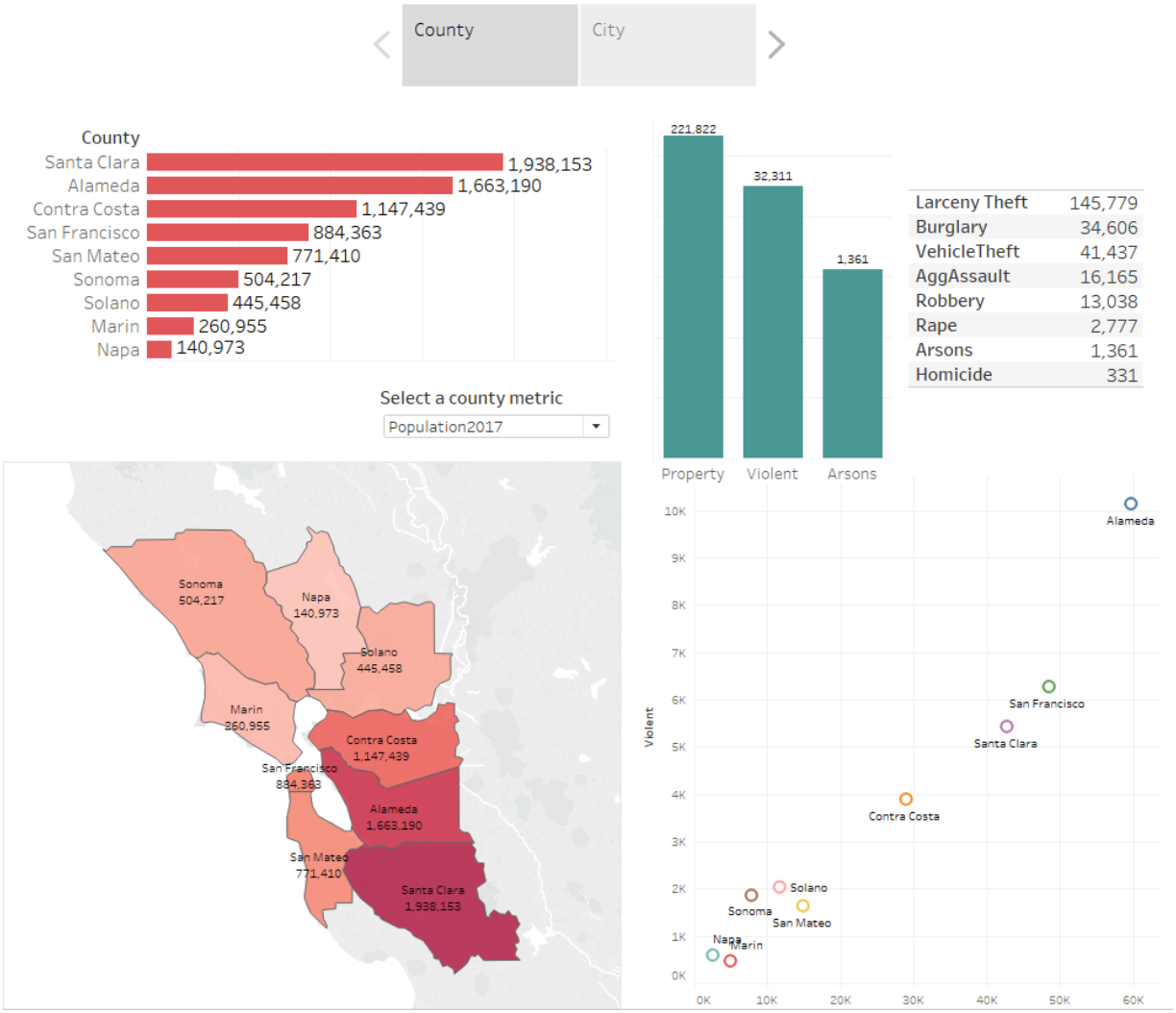
Explore San Francisco Bay Area

Name: Zixin Ouyang

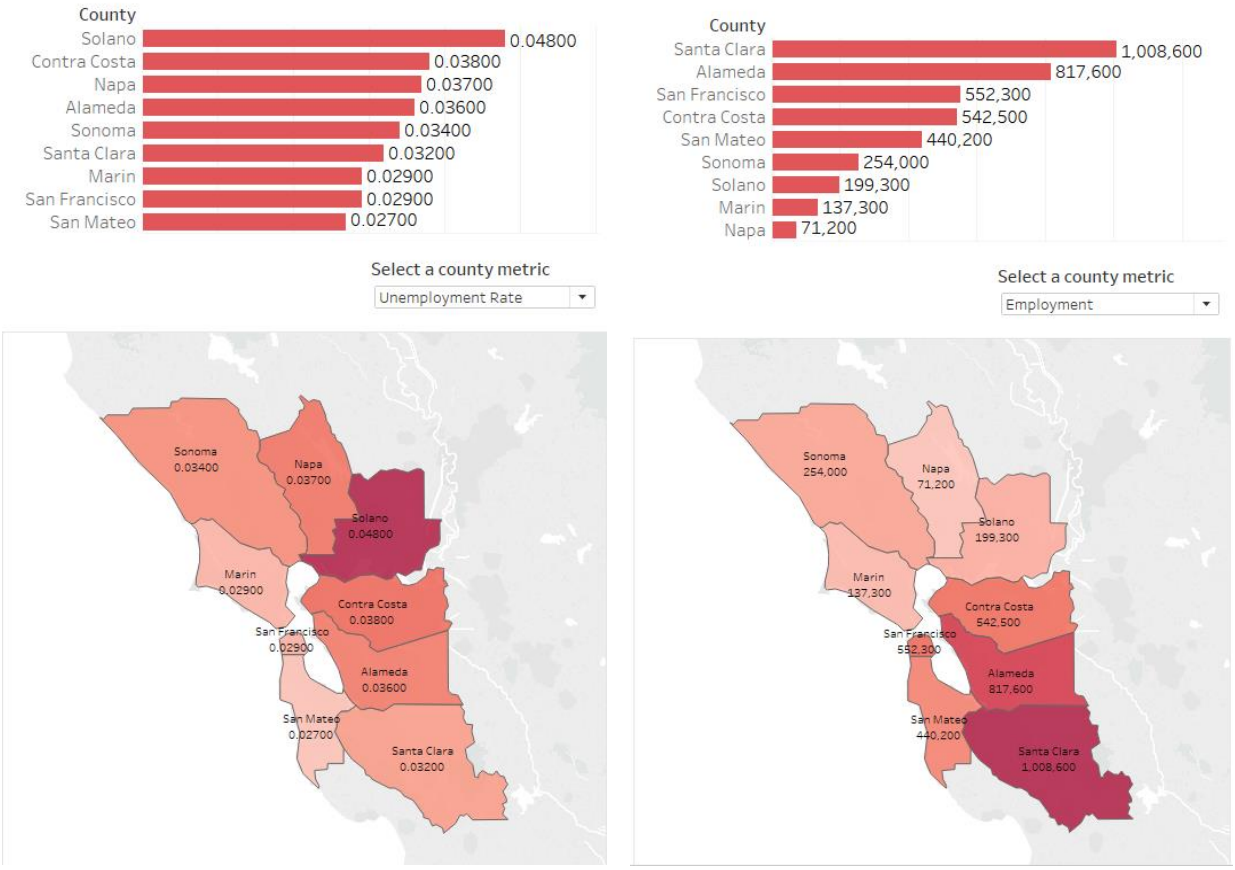
I want to explore San Francisco Bay Area in the project. I know many students relocate to Bay Area after graduation, so I hope my project could provide some insights.

I analyzed data from there perspectives: Employment, Crime, and Rental Cost. The population data comes from the website of United State Census Bureau, the employment data come from the website of State of California Employment Development Department, the crime data comes from the website of State of California Department of Justice, and the rental data comes from the website of Zillow, an online rental estate database company. The tools I used were Microsoft Excel, R, and Tableau. I used Microsoft Excel, including one of its add-in Power Query, to prepare the data: like filtering states, counties, and years, and deleting columns not needed. In the next step, I used R to further clean and re-organize data, like merging several datasets, and splitting data by different level of regions (County, City). You can see the R codes for data manipulation in the project folder.

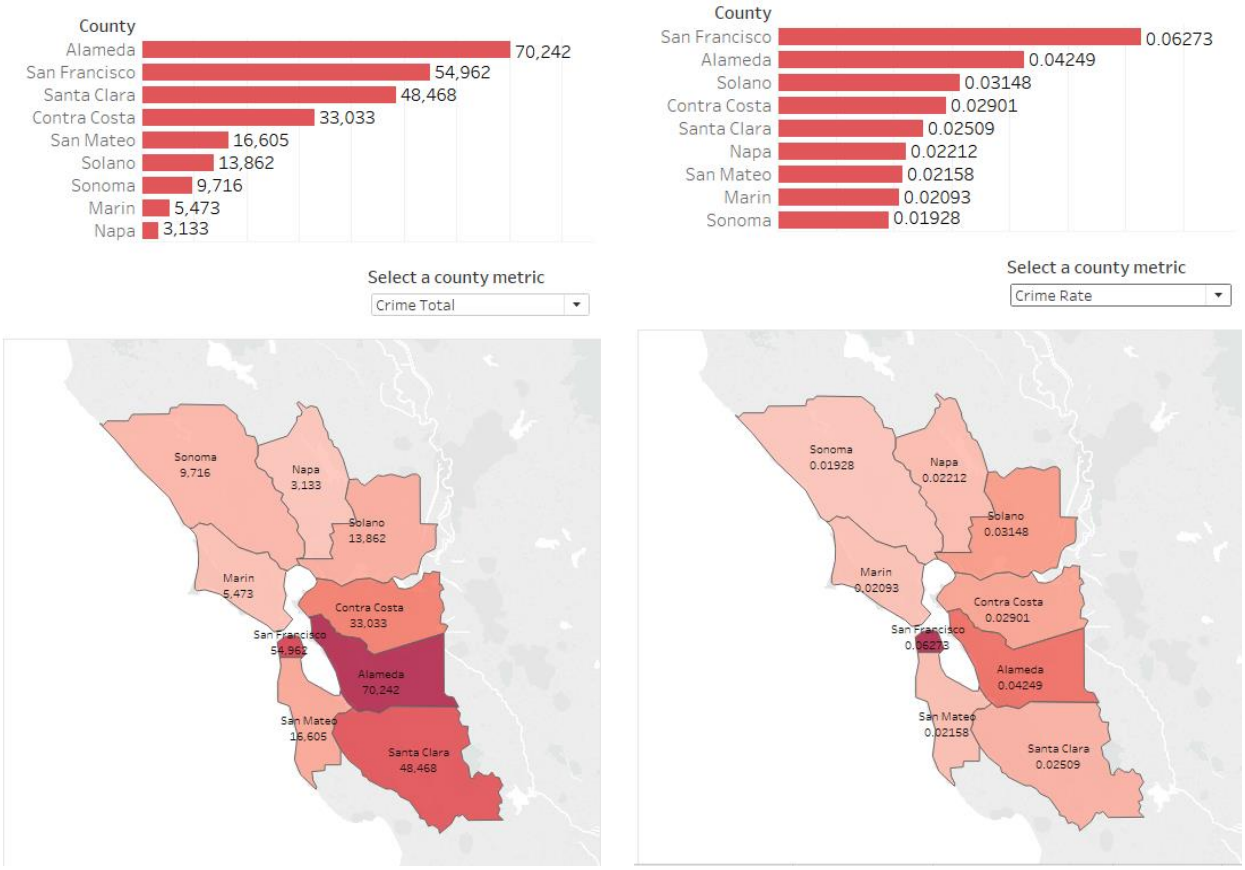
Bay Area consists of nine counties: Santa Clara, Alameda, Contra Costa, San Francisco, San Mateo, Sonoma, Solano, Marin, and Napa. The nine-county Bay Area can be further divided into five sub-regions: East Bay, North Bay, South Bay, Peninsula, and the city of San Francisco. The East Bay includes Alameda and Contra Costa. The Peninsula includes the cities and towns on the San Francisco Peninsula, excluding the city of San Francisco. It includes San Mateo county, and also includes the northwestern Santa Clara county cities of Palo Alto, Mountain View, and Los Altos. The South Bay includes all of the rest of the cities in Santa Clara county. The North Bay includes Marin, Sonoma, Napa, and Solano counties.



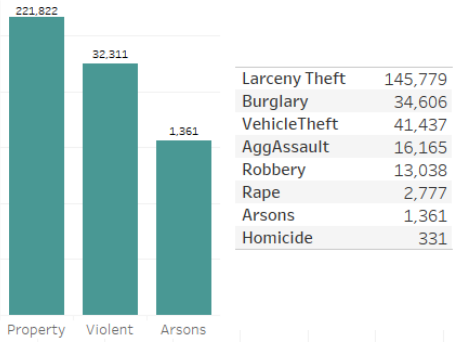
First, I analyzed data by county. I filtered the nine counties in Bay Area when visualizing the data. If you are interested in population, employment, crime, and housing unit of all counties in the state of California, you only need to remove the county filter in the County_Map and County_Bar Chart sheet. The bar chart in the top left shows the ranking, while the map under it shows the geographical distribution. You can select a metric using the drop-down menu. The three largest counties ranked by population are Santa Clara, Alameda, and Contra Costa, and their population are all over one million. The North Bay is the largest and least populated sub-region.



The left plot shows unemployment rate, while the right plot shows number of employment. Solano has the highest unemployment rate, while San Mateo has the lowest unemployment rate, and it is much lower than all other counties. The unemployment rate of Marin and San Francisco is the same. However, San Francisco has much larger employment number than Marin. The employment number is almost proportional to population, except San Francisco and Contra Costa. San Francisco has the fourth largest population, but its employment number is the third largest.

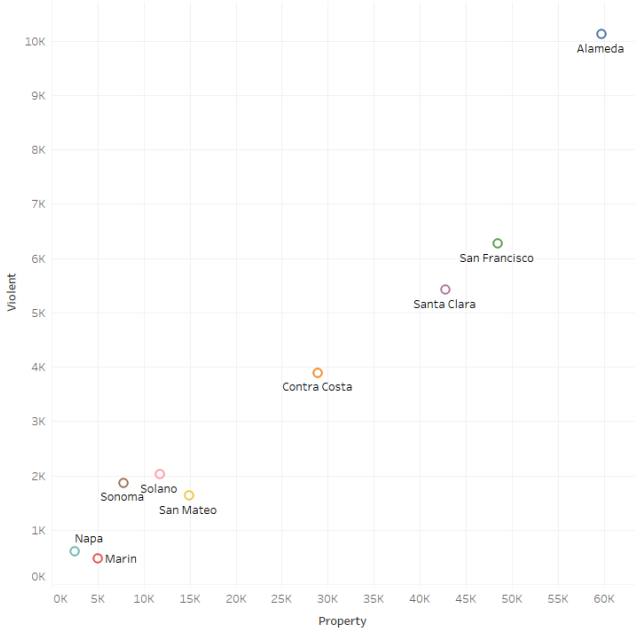


I analyzed crime data in the next part. Crime numbers of Alameda is about seventy thousand, San Francisco is fifty-five thousand, Santa Clara is forty-eight thousand, and Contra Costa is thirty-three thousand. Again, the four counties in the North Bay have small crime numbers. Crime number is related to population. However, Alameda has the second largest population, while it has the largest crime number. San Francisco has the fourth largest population, while it has the second largest number of crime. Therefore, analyzing crime rate might be more meaningful. Crime rate is calculated using total crime number, divided by population. San Francisco has the largest crime rate, followed by Alameda, Solano, Contra Costa, and Santa Clara. Even though Solano has small number of crime, its crime rate is high. San Mateo has low unemployment rate and low crime rate.

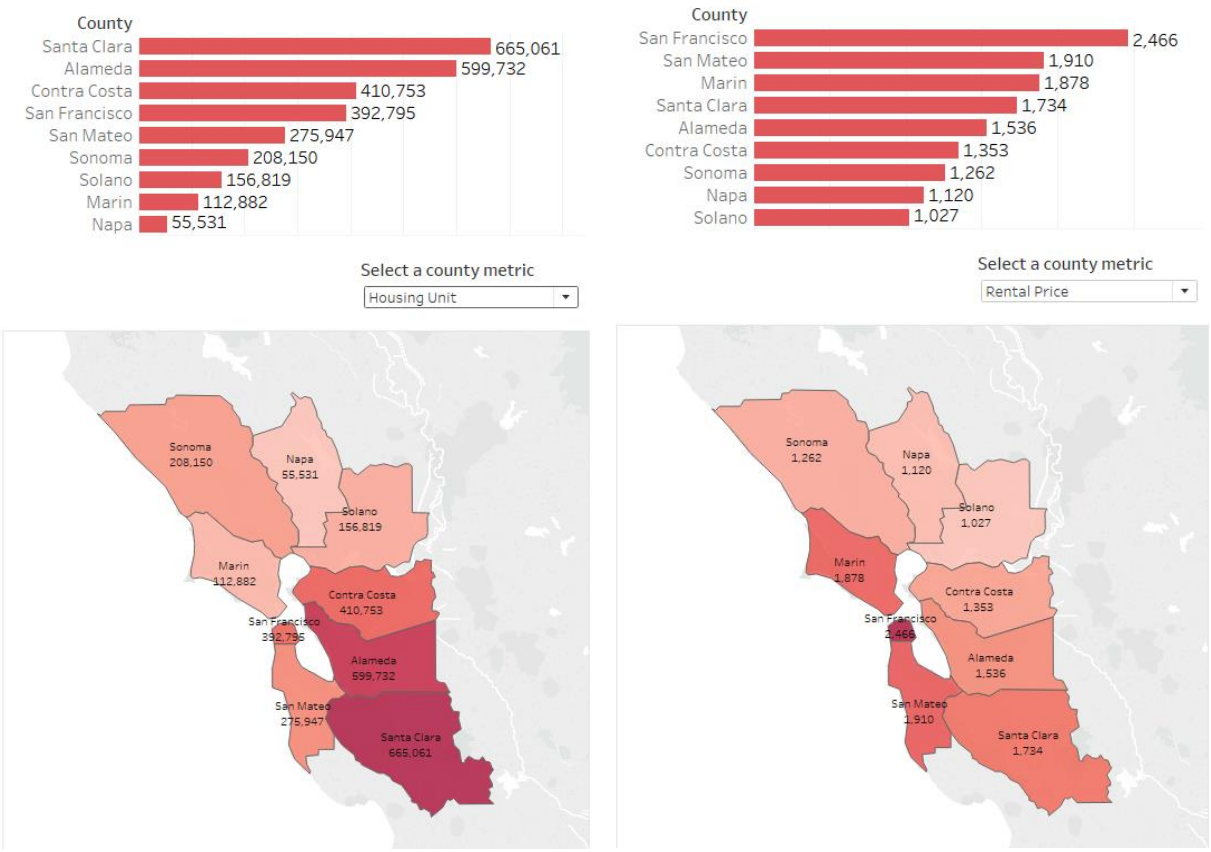


Larceny Theft	145,779
Burglary	34,606
VehicleTheft	41,437
AggAssault	16,165
Robbery	13,038
Rape	2,777
Arsons	1,361
Homicide	331

There are three general crime types: violent crime, property crime, and arson. Violent crime can be further divided into homicide, rape robbery, and aggravated assault. Property crime can be further divided into burglary, larceny theft, and vehicle theft. The left plot shows the three general types of crime in the entire bay area. The number of property crime is about two hundred and twenty-two thousand, the number of violent crime is about thirty-two thousand, and the number of arson is about one thousand. The number of property crime is about seven times of violent crime, and arson is very rare. The right plot shows the sub-categories of crime type. The top three crime types are all property crime. Larceny theft is the most common crime type, and it is more than the total of all other crime types. Homicide is the least common crime type.

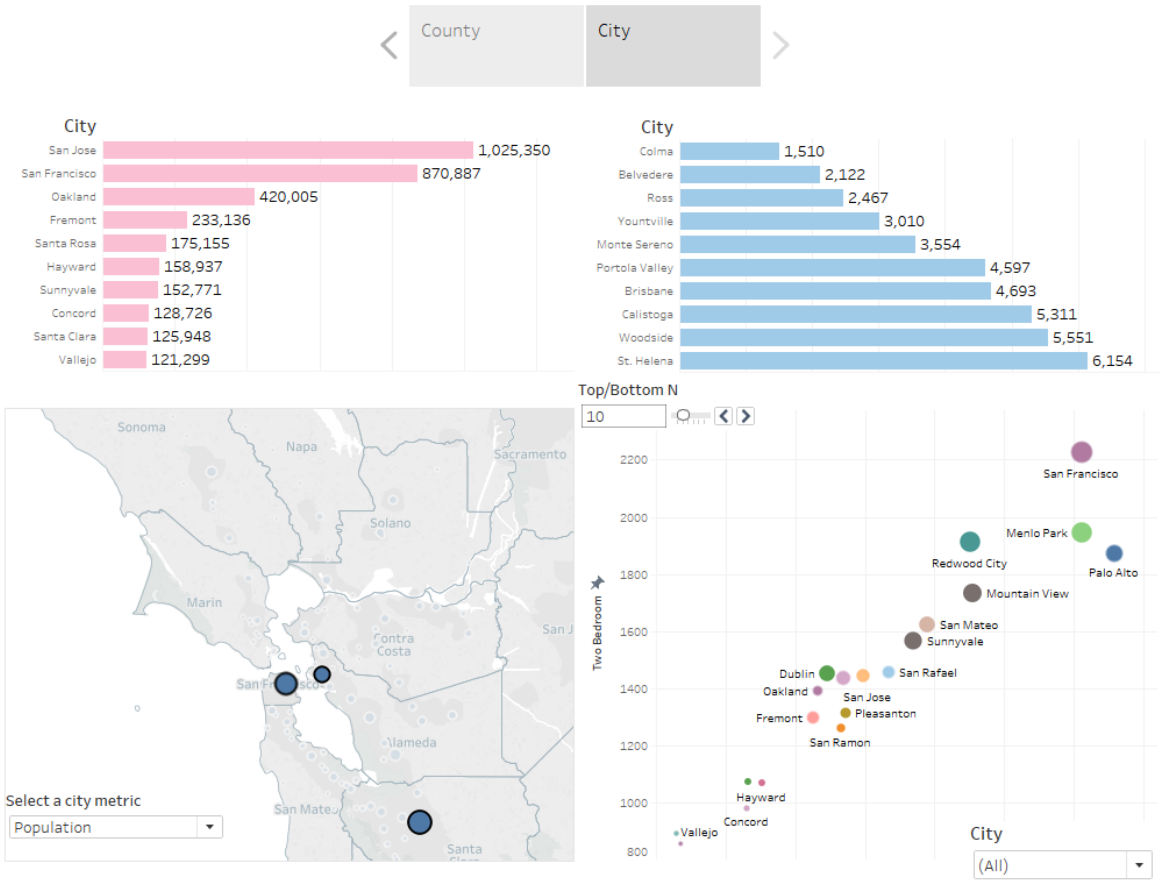


I compared violent crime and property crime in the nine counties. Alameda has the largest number of violent crime and the largest number of property crime. If the reader more concerns about violent crime, he can focus on the y-axis.



The number of housing unit reflects the ease of finding an apartment. Its rankings are the same as population. The rental price in the original data is the median rental price. I averaged the rental price for the last twelve months and then divided it by the number of bedrooms to get the price for per person. For counties, I used the average rental price of one-bedroom, two-bedroom, and three-bedroom in my analysis. I did not include rental prices of studio and four-bedroom because the data is incomplete. I think that is because these two floor plans are not common in the Bay Area and my later house searching confirmed the reason. The rental price is highest in San Francisco. San Mateo has the second highest rental price, but it is much less than that of San

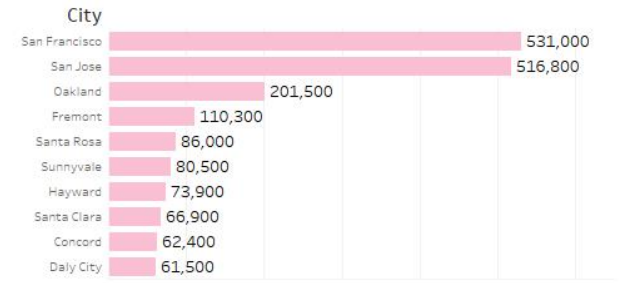
Francisco. All three counties in the North Bay have relatively low rental price, except Marin, which has the third highest rental price. I suppose that's because of its proximity to San Francisco.

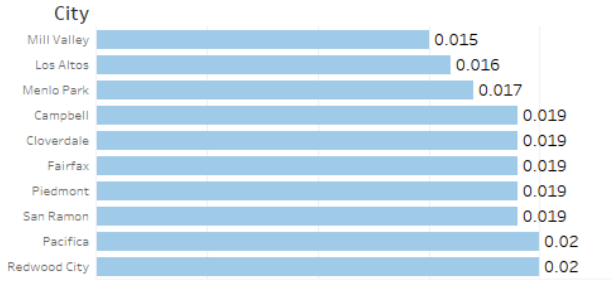


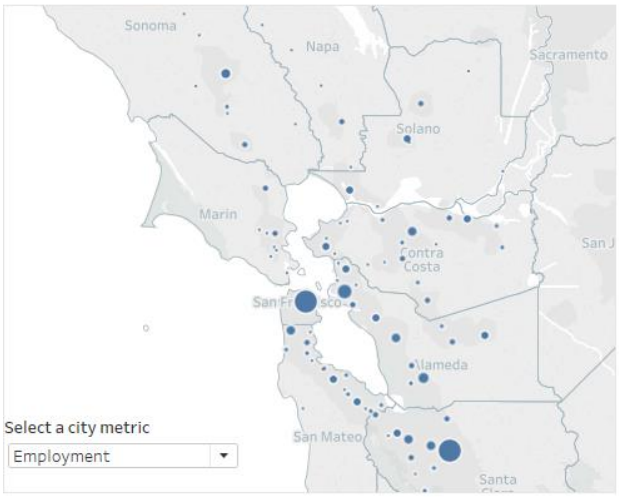
City	Populati..	Employ..	Unemplo..	Crime To..	Crime Ra..	Violent s..	Property..	One Bed..	Two Bed..	Three Be..
Oakland	420,005	201,500	0.048	30,202	0.072	6,059	23,952	2,174	1,390	1,063
San Francisco	870,887	531,000	0.029	53,847	0.062	6,190	47,402	3,338	2,225	1,824
San Jose	1,025,350	516,800	0.036	28,773	0.028	3,887	24,749	2,449	1,434	1,137

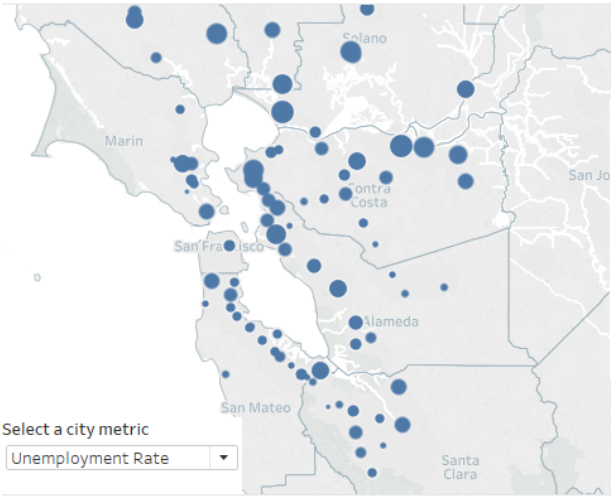
Second, I analyze data by city. I filtered cities in the Bay Area in data manipulation process before visualizing the data. The bar chart shows the ranking, while the map shows the geographical distribution. You can select a metric using the drop-down menu, and if you click a city on the map, the table below the map will show all the data about the city. You can also search a city directly. The left bar chart shows the top N cities ranked by a certain metric, and the right plot shows the bottom N cities. And you can adjust the N using the parameter control Top/Bottom N. The top three cities with the largest population are San Jose, San Francisco, and Oakland, and their

population are much more than other cities'. San Jose is the only city which has over one million populations.

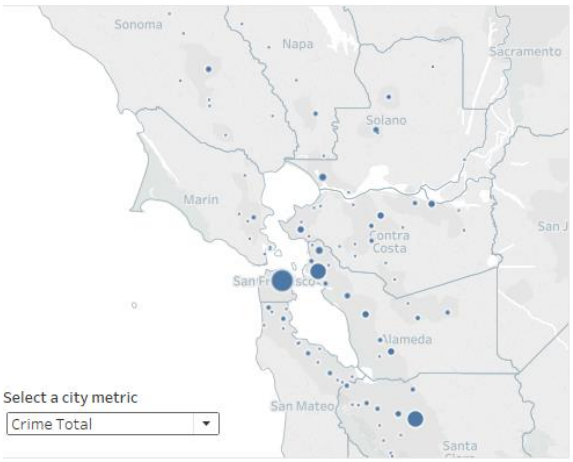


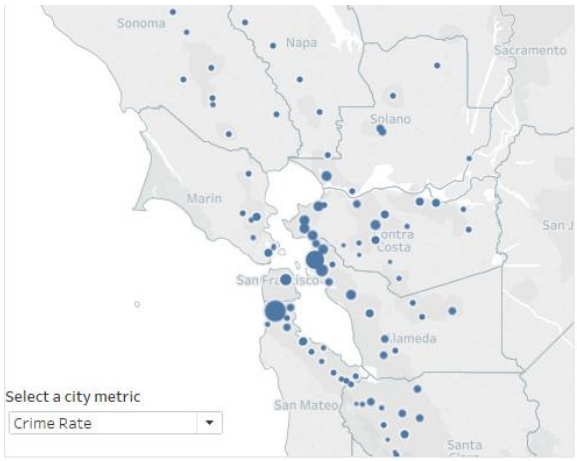






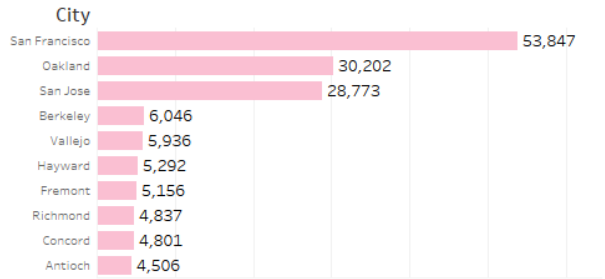
The top ten cities with highest employment number are nearly the same as the top ten cities with largest population, except that the tenth city with the largest population is Vallejo, while the tenth city with the highest employment number is Daly City. However, the rankings are different. If you look at the top 10 cities with lowest unemployment rate, you will find the none of them shows in the top 10 cities with highest employment number. Low Unemployment rate might be caused by few populations and the top ten 10 cities with the highest employment number are famous cities, which we are more familiar with, so employment number is more meaningful than unemployment rate.

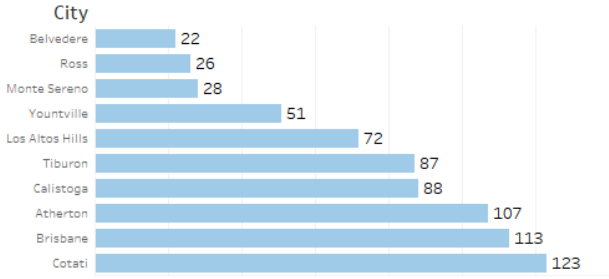




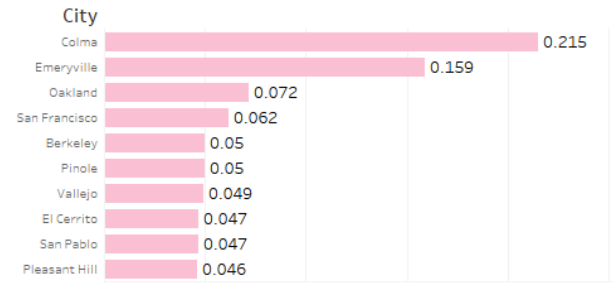
The left plot shows the total number of crime of all cities in Bay Area, while the right plot shows the crime rate of all cities in Bay Area. According to the maps, Cities near Silicon Valley are safe in both aspects. Crime rates of cities near Oakland are high.

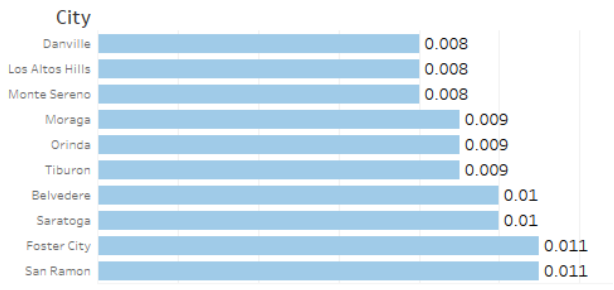
Crime Total:





Crime Rate:

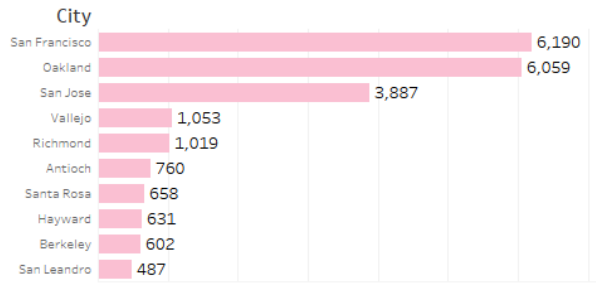




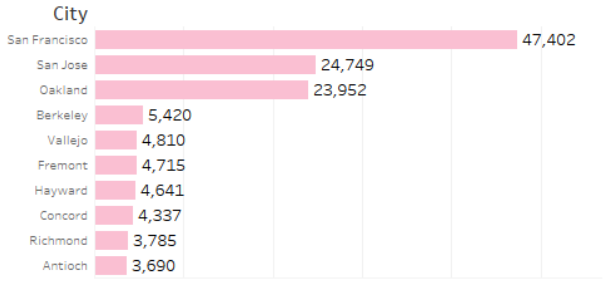
The numbers of total crime of San Francisco, Oakland, and San Jose are extremely large compared to other cities. However, the top ten cities with the highest crime rate are different from the top ten cities with the largest number of total crime. Colma and Emeryvile does not stand out

when looking at the number of total crime, but their crime rates are among the top three. Top ten cities with the smallest number of total crime and the lowest crime rate are all small towns or cities.

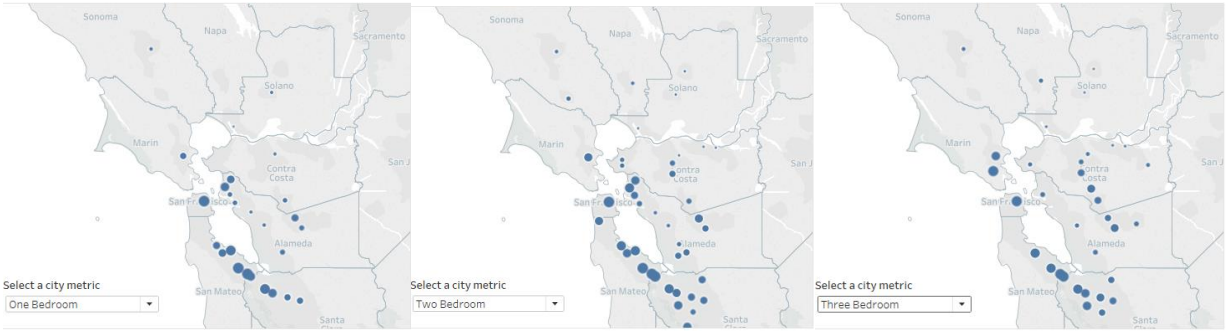
Violent Crime



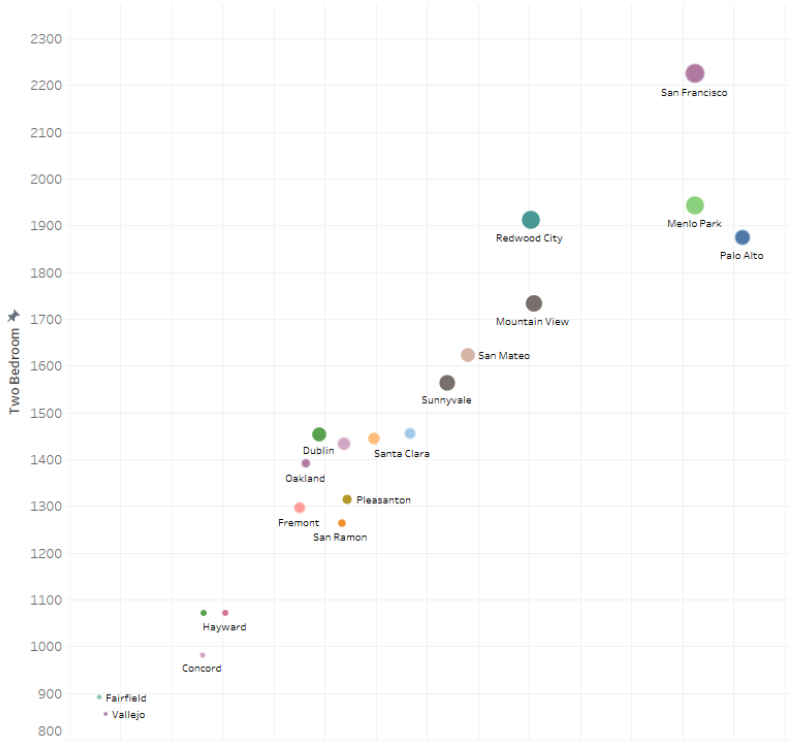
Property Crime



Again, I display violent crime and property crime separately since people are more concerned about violent crime. From the bar charts, we can see that the high ranking of total crime number might be caused by large number of property crime, for example, Berkeley.



The three map plots display the median rental price of one-bedroom, two-bedroom, and three-bedroom, respectively. They don't include all cities in Bay Area because of the incomplete data. The rental prices of cities near San Francisco and Silicon Valley are higher than other cities.



The plot above is a three-dimension plot: The X-axis shows the rental price of three-bedroom, the Y-axis shows the rental price of two-bedroom, and the size of circle shows the rental price of one-bedroom. If you are only concerned about rental price of a certain floor plan, you can focus on one dimension.

Usually, the number of employment and number of crime are positively correlated to population. If you already have a job offer, you don't need to consider the number of employment in choosing a city to live. If you are still looking for a job, I would suggest you relocate to a place full of job opportunities. Also, you can choose to commute between cities. You can live in a city with relatively low rental price and low crime rate, but work in a city with large number of job opportunities. However, you need to sacrifice your time in this way. In addition, you need to consider your own budget in choosing where to live. Solano has high unemployment rate and high crime rate, even its rental price is lowest among all counties, I would advise not to live in this county. There are some restrictions in my project. Work location and transportation condition are

also considerations in choosing where to live. The employment number and unemployment rate are too general, so you should also do research on which city has more opportunities of the job position you are looking for. I searched the data analytics/business intelligence jobs in the Bay Area on Linkedin, and I find most of these positions are in San Francisco, but some of them are in San Mateo, Sunnyvale, and Oakland. The rental price and crime rate in San Francisco are high, so I decided to live in the Peninsula, and San Mateo is an ideal city. Also, if you don't want to drive to work, you can choose to live near a BART or Caltrain station. You can check the Bay Area Regional Rail services using the link in the reference. In addition, I only include the one bedroom, two-bedroom and three-bedroom floor plans in my analysis. However, during my apartment searching process, I find town house is also a popular choice. I did not analyze the rental price of town house because the price is for an entire house and the number of rooms are various, which makes me unable to calculate the rental price for per person. I encountered some challenges when I did my project. First, the same variable is not consistent in several datasets. For example, the same city has different names in different datasets. Therefore, I spent a lot of time to clean and validate data using Microsoft Excel (Power Query). Second, the data was not structured in the way I wanted. Since using Tableau to visualize data requires well structed data, I used R to re-organize data before visualizing it. I analyzed the commuting pattern between counties, but I found the information was useless since most people live and work in the same county, so I discarded this part. To combine all visualizations into a dashboard, I used parameter control. You can select a metric using the drop-down menu.

Reference

https://en.wikipedia.org/wiki/San_Francisco_Bay_Area

https://en.wikipedia.org/wiki/Transportation_in_the_San_Francisco_Bay_Area

Data Sources:

State of California Employment Development Department

http://www.labormarketinfo.edd.ca.gov/data/labor-force-and-unemployment-for-cities-and-census-areas.html

State of California Department of Justice

https://openjustice.doj.ca.gov/crime-statistics/

United States Census Bureau

https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Zillow

https://www.zillow.com/research/data/