Contents

Executive Summary ............................................................................................................. 3

Introduction ......................................................................................................................... 4
  Overview ............................................................................................................................ 4
  History ............................................................................................................................... 5
  Crime ................................................................................................................................ 5
  Toxic Sites ........................................................................................................................ 6
  Income .............................................................................................................................. 6
  API Scores ......................................................................................................................... 6

Problem Statement and Research Question ........................................................................ 7

Methodology ......................................................................................................................... 8
  Deciding on What Data ...................................................................................................... 8
  Where the Data came from ............................................................................................... 9
  Data Overview .................................................................................................................. 9
  Molding the Data ............................................................................................................... 9

Findings ................................................................................................................................. 12

Analysis (1-2 pages) ........................................................................................................... 20
  Median Income and Age, 2000 & 2010 ............................................................................. 20
  Oakland’s Population and Ethnic Distribution .............................................................. 20
  Student’s Ethnic Distribution ......................................................................................... 20
  Crimes of Oakland 2009 .................................................................................................. 21
  Oakland High Schools Ranked by Vicinity to Toxic Sites ................................................ 21
  School’s Ranking vs. Ethnic Distribution ........................................................................ 21
  API Score Prediction ........................................................................................................ 21

Conclusion ........................................................................................................................... 22

References ............................................................................................................................ 24
  Literature ......................................................................................................................... 24
  Data Files .......................................................................................................................... 25

Appendix ............................................................................................................................... 26
  Data Acquisition .............................................................................................................. 26
  The GIS Steps .................................................................................................................. 26
  Molding the Data ............................................................................................................. 26
Executive Summary

Oakland, CA is a city known for its diversity of its people, its culture, and its history. However, Oakland is also known for high crime levels, low income population, toxic sites, and low API scoring schools. These negative factors affect the well-being and performance of the students in the area and their ability to reach their fullest potential.

This study examines whether or not the students in Oakland are given equal opportunities independent of ethnicity. It will observe the ethnicity of the students around the schools that are in areas of lowest income, have the highest crime to student ratio in their areas, are close to toxic sites, and have low API scores. This will see if a certain ethnicity has a worse learning environment than others.

A regression will also be run to see if the environmental factors of income, crime, and toxicity affect a school’s API score, which is solely based on the performance of the students. This basically will see if the students’ performance can be linked to their environments.

With the use of GIS, this project shows that African American Students and Hispanic students go to schools who are worse off in regards to those factors. It also shows that these environmental factors of toxicity, income, and crime can predict the API score most Oakland High Schools. There are three schools that have however scored higher or lower than the regression would have predicted due to these environmental factors.

The data for this project was located from the Census Bureau and Code for Oakland via shape files and excel files that were later linked together and analyzed.
Introduction

Overview

Oakland is located in the State of California, in the County of Alameda as seen in the figure above. The Oakland is also referred to as the East Bay to the locals. The red circle in the City of Oakland map represents the downtown area of Oakland. The peninsula left of that is connected to San Francisco via the Bay Bridge. San Francisco is known to be one of the most expensive cities to live in, beating Manhattan in the running. The Bay Area has become the heart of the tech boom that has been occurring. However, the wealth of Oakland does not match the wealth of San Francisco, even though they are right across from each other.
History

Oakland used to be a prosperous city before and during the Second World War. It had a very profitable shipbuilding and automotive industry that later faltered after the war. Also, the wealthy people who used to live in Oakland moved downtown moved to the suburbs in the 1960s. Therefore Oakland was becoming poorer around that time as well. (Oakland, California, 2014)

Oakland is currently a very ethnically diverse city with the majority of its residents being African American. To be exact, 27% of the population is African American, 26% is White, 25% is Hispanic or Latino, and 17% is Asian. (Oakland, California, 2014)

Oakland is a modern city in the sense that it has a transportation system consistent of underground rail and bus, it has a port and airport. It has a downtown area and suburbs. Oakland boasts large green spaces in its hills and an overall vibrant culture and personality. (About the Port, 2014)

Crime

While Oakland might be home to incredible culture and people, it is also home to a lot of crime. According to Forbes, the city is ranked as the third most dangerous city in the United States, with a crime rate of 1,683 per 100,000 residents. It is also defined as the city with the most robberies in all of the United States. (The 10 Most Dangerous U.S. Cities, 2014). One of the major crime types in Oakland are drug crimes. They are so prevalent that the Oakland police has chosen to focus more on them then the homicides in Oakland. This might come as a surprise since the reason Oakland ranks so high on the list of most dangerous cities, is due to the homicide rate. (Winston, 2013).

There are some studies that have looked at the linkage between the crime rate and schools. One such study specifically looks at the Oakland school system. African American youth made up 73.5% of all juvenile arrests in Oakland between 2005 and 2012. (Black Organizing Project, 2013). Other studies looked at how crime influences peers around them. A study showed that those that were exposed to peers who committed crimes were more likely to commit crimes themselves. In other words, the crime rate around them has an influence on the well-being of a person.
Toxic Sites

The Bay Area might be known as a very environmentally conscious area, there are still many Toxic Sites in Oakland. While many are inactive, it has been seen that those sites can have effects on the lives of people surrounding them hundreds of years later. Such an example exists in the Southside High School in Elmira, NY. It was built in the vicinity of toxic site that was deactivated more than a 100 years ago. Unfortunately, the students have had negative side effects, such as cancer, because they were exposed to dangerous chemicals that can linger for hundreds of years. (Norris, 2014)

There have been a multitude of studies done between schools and the toxic sites in their vicinity using GIS. One such study made toxic sites as one of the many factors that pose environmental threats to student. The Study was conducted by University of North Carolina, Chapel Hill in 2008. The study included more than one type of hazard. It included interstate highways, US routes, hazardous waste facilities, TRI facilities (toxic release inventory facilities), flooding, and railroad tracks. In regards to toxic hazard, waste or TRI facilities, the radius of influence was a quarter mile. They used GIS to see what schools fall within these buffer zones around the hazardous areas. (Salvesen, Zambito, Hamstead, & Wilson, 2008)

Income

Oakland is known as being one of the poorer regions in the Bay Area. The unemployment rate of 7.6% in Oakland is much higher than California’s average of 7.1% and the Nations average of 6.0%. This unemployment high also contributes to the great amount of poverty seen in the region. The national poverty level is at 14.9% and California’s comes in at a bit higher with 15.3%. Oakland however has a staggering poverty level of 20.3% which is much greater than those of the Nation and California. (Oakland, CA Employment & Jobs, 2012)

Income can play a very large part in a child’s life. In a study conducted by the National Bureau of Economic Research, it is seen that income has an effect on a child’s achievement. According to their study, an increase of 2.1% percent of a standard deviation in math test score and an increase of 3.6% of a standard deviation in reading test scores can be seen for every one thousand dollar increase in a family’s income. (Dahl & Lochner, 2005)

API Scores

The API scoring system is a system by which the State of California can measure the academic achievement and progress of its schools. This system of ranking began in 1999 among the public schools after the 1999 Public Schools Accountability Act was
passed. Factors that determine a school’s API score include standardized tests, CST and CAHSEE, the students take and the progress the students are making on those tests over the years. The API scoring system does not take into account the actual school, just the scores that the students receive on the tests. Hence, the spotlight is on all the students. The API score does not measure the quality of the school or does it provide any descriptor of the environment of the school or surrounding the school. However, the schools receive rewards in the form of money for producing a good API score. This incentive is meant to motivate the schools in teaching their students better. The aim for all schools is to reach a score of 800 or more. (Staff, 2014)

There was a study done with Los Angeles charter schools and their API scores compared to regular public schools to see if the charter schools outperformed regular public schools. The study used GIS in order to create a “charter competition” variable through the buffer tool. It was defined as a five mile radius around each of the charter schools found in LA. The study outlined that the mean API score for the charter schools in LA was 788.92 while the mean of the traditional public schools was 719.92. (Trachtman, 2013) These mean API score are much higher than the mean for Oakland’s traditional public schools with a low score of 613.

The studies thus far have looked at certain aspects of this project. However, there are other studies that look at other pressing issues regarding the people of Oakland through the use of spatial analysis and GIS. One study concerned with the crime, neighborhood characteristics, and participation in asthma case management in Oakland used GIS to calculate that neighborhood deprivation does not provide any indication of program participation for asthma. The education level of the neighborhood is a factor of an asthmatic child’s participation probability. (Gale, Magzamen, Radke, & Tager, 2011)

Problem Statement and Research Question

Oakland ranks as one of the most dangerous cities in the United States. In 2010, Oakland was ranked as the 3rd most dangerous city in the United States based on it’s crime rate. The neighboring city of Richmond comes in right behind it. (The 10 Most Dangerous U.S. Cities, 2014). Oakland houses eight toxic sites along the East Bay coast.

However all these figures are just averages and don’t show the great amount of diversity within Oakland. Therefore this study aims to see the more specific environments in which students attend high school in. The schools will be judged on the following factors:

1) Crime per student population ratio of the area
2) Vicinity to toxic sites
3) Wealth of the surrounding area
4) API Scores of each school
This ranking of the schools will then be superimposed with the ethnic distribution of high school students in order to get a better view if all ethnicities are provided for equally. Therefore the research question is as follows:

Are the high schools in Oakland providing the same learning environments equally to all the different ethnicities in Oakland, based on those four indicators?

This part will be studied visually to see any ethnic patterns that come with the school’s traits and characteristics. The second part of the this study will take a look and see if the crime rate per student population, vicinity to toxic sites, and income can predict the API score of each school. Therefore the next important research question that will be asked, discussed, and answered is:

Can crime per student of the area, the school’s vicinity to toxic sites, and the income of the area predict the observed API scores that each school received?

Methodology
Deciding on What Data

For this project, I did not so much find the data, as the data actually found me. My original plan was to study the changes in rent from 1990 to 2014 in the entire Bay Area, capturing the two tech booms that the Bay Area is famous for. However, when trying to find the data, it was not as clean cut as I would have liked, especially since the format of the data changed over the years. Therefore, I decided to focus on only one year and one county, Alameda County – however I could not find crime data for all of Alameda County. Therefore I narrowed it down to focus only on Oakland. Once I found the data for crime, toxic sites, and schools in Oakland, I decided to stick with those and work from there and see what analysis can be done with the data.

Regarding the census data, I searched for things that might directly influence the students. This included the income, median age, ethnic distribution, and actual enrollment of students. These all came in the format of excel files with the tract numbers as their identifier. I therefore had to also download the tracts for both 2000 and 2010, since the tracts have changed between those two censuses. I would later on link the excel data with the shape file via the census tract numbers.
Where the Data came from

The data regarding the income, median age, number of students enrolled in school, and median age for each tract were found on the Census Bureau Website through their search engine. The data came in the form of excel files for Alameda County at the Census tract level.

The data regarding the water outline in the bay was found on DataSF.com. The remaining data was found from Code for Oakland, a website whose purpose was to host a “hackathon” for city of Oakland.

Data Overview

The data that I ended up using and later combining is as follows:

1. Bay Area Water Outline (.shp)
2. Census Tract California 2000 (.shp)
3. Census Tract California 2010 (.shp)
4. School in Oakland (.shp)
5. Toxic Sites in Oakland (.shp)
7. Median Income per tract 2000 (.xls)
8. Median Income per tract 2010 (.xls)
9. Median Age per tract 2000 (.xls)
10. Median Age per tract 2010 (.xls)
11. School Enrollment per tract 2010 (.xls)
12. Ethnic Distribution per tract 2010 (.xls)

Molding the Data

For all Shape Files

- Please note that before importing any shape file I projected it first on:
  - NAD 1983, State Plane III FIPS 0403 Feet
  - This projection was chosen since the guidelines for projections in California were found on the following website:

Census Tracts - California

1. I downloaded the Census tracts for all of California for both 2000 and 2010 from TIGER
2. I downloaded the Bay Area Water general outline shape file from DataSF.com
3. The tracts of California had tracts also for parcels that are water, so I had to erase them with respect to the water shape file. I omitted the water shape file for the final report.
   a. At this point I still thought I was going to work with the entire Bay
4. I then selected the tracts, for both 2000 and 2010, what seemed like to encompass all the counties in the Bay Area.

**Excel File Modifications and Importing**

5. I downloaded from the census bureau income data for 2000 and 2010 at the census tract level to include all the counties of the Bay.
6. Using the income data for all Bay Area Counties, I fixed up the excel file so that it would work with GIS and matched the tract column in the California tract shape file.

**Joining CSV Files and Shape Files – The Process**

1. To give the CSV files a spatial reference, I had to join it with the census tract file.
2. The census tract field in the shape file is in the form of a string, so I would create a new field (that is a double) and use the field calculator to just multiply the previous tract labeled field by 1 to get the same number but as a double.
3. I clipped the shape files to match that of the Oakland crime shape file, and resaved all of them.

**Reworking of platform files**

1. I had to work on the ethnicity file in excel because I wanted the percentage of each ethnicity within the tract rather than the actual number of people.
   a. This created a new csv file which included the percentage of each ethnicity in each tract as well
2. I had to rework the student enrollment file by grouping the students from 1st grade to 8th grade (elementary and middle school) and from 9th grade to 12th grade (high school).
   a. I re-added this csv file into GIS and joined it again to the tracts to create the correct representation I was looking for.
3. I joined the student enrollment file with the ethnicity file. Once joined I exported it so that I could work on it in Excel
   a. In excel I multiplied the percent per ethnicity by the number of students in each group to get the number of each students of a certain ethnicity for each tract
4. I reworked the school location file to only include schools labeled as ES (elementary school), MS (Middle School), and HS (High School).
   a. This was done through editor in GIS
5. I joined the school enrollment file with the crime file.
   a. In the new crime shape file I added a new field to represent the rate of crime per student

**Toxic Site shape file and school shape file**
1. The toxic site shape file had a point location for all the toxic sites found in Oakland.

2. Using Euclidean Distance from the toolbox I created a raster sheet. The values in the raster were for feet rather than miles.

3. Using the Reclassify tool from the toolbox I input the raster I previously created and created it on intervals based on the 0.25 mile, 0.5 mile, 1 mile, 2 mile equivalent in feet. I chose to omit anything beyond two miles since the schools wouldn’t be affected that much.

4. Using the editor tool, I manually input the category of proximity to the nearest toxic site in which the school fell, for the school shape file. If it was more than 2 miles, I labeled it as 3 miles.

**School Shape File**

1. I joined a new school shape file for income of that tract and also the ratio of crime per student for each tract that it fell in. While all joins have been done through table joins, this join was done spatially.

2. Realizing I wanted to do a regression I joined all those separate school files into one.

**Geographically Weight Regression**

1. Using this new school master file that has the API score, crime ratio, tract income, and proximity to toxic sites for each school, I ran a geographically weighted regression (GWR)
   
   a. Dependent variable: API score
   
   b. Independent Variable: income, crime ratio, & proximity to toxic sites

2. This created a new file.
Findings

(The next 7 pages are the findings, which are composed of the maps created with the data)
Oakland's Population and Ethnic Distribution

NAD 1983
State Plane III FIPS 0403 Feet

1 Dot = 10 People
- White
- Hispanic
- African American
- Asian

Each dot represents a group of 10 people. The dot does not give the specific location of each group of ten, but rather the density and distribution for each tract. The maps on the right hand side show the distribution of each ethnicity individually.

Sources: Code for Oakland, Open Street, Census Bureau, TIGER

created by Stasa Zivojinovic, 11/30/14
**Student's Ethnic Distribution**

NAD 1983  
State Plane III FIPS 0403 Feet

- **Elementary/Middle School Students**
  - 1 Dot = 1 Student
  - White
  - Hispanic
  - African American
  - Asian

- **High School Students**
  - 1 Dot = 1 Student
  - White
  - Hispanic
  - African American
  - Asian

Each dot represents one elementary or middle school student of stated ethnicity. The location of the dot does not represent the exact location, but rather the density and distribution of students within each tract.

Sources: Code for Oakland, Open Street, Census Bureau, TIGER

created by Stasa Zivojnovic, 11/30/14
Oakland High Schools Ranked by Vicinity to Toxic Sites

NAD 1983
State Plane III FIPS 0403 Feet

The map on the left hand sides shows the radius around the toxic sites and the location of the high school with respect to that radius. The map on the right then ranks the schools based on where they fall within the radius categories. The larger the circle is that represents the school, the closer the school is to a toxic site.

High Schools in Relation to Toxic Sites

High School’s Vicinity Ranked

Created by Stasa Zivojnovic, 11/30/14

Sources: Code for Oakland, Open Street, Census Bureau, TIGER

Oakland Public High Schools
CRP 386
Each circle represents a high school in Oakland. The larger the circle, the worse the school is with regards to the factor listed for that map. These rankings are then superimposed with the ethnic distribution to see the ethnicity around each of the schools.

1 Dot = 1 Student
- White
- Hispanic
- African American
- Asian

Created by: Stasa Zivojnovic
11/30/14

Sources:
Code for Oakland, Open Street, Census Bureau, TIGER
API Score Prediction

NAD 1983
State Plane III FIPS 0403 Feet

A linear regression was run with the API being the dependent variable and the area's income, crime rate per student, and vicinity to toxic sites as the independent variables. The red dots represent schools who have a higher API score than would have been expected with the characteristics that surround the school. The blue dot represents the school whose API score is lower than expected with the characteristics that surround the school. The yellow dots represent the schools whose API is closely predicted to the actual API score based on those factors.

Piedmont High School
API Score = 904
Standardized Residual > 1.5
The API Score is higher than predicted

Oakland High School
API Score = 633
Standardized Residual < -1.5
The API score is lower than predicted

Wilson (Lionel) College Preparatory Acad
API Score = 792
Standardized Residual > 1.5
The API score is higher than predicted

Standardized Residuals

- Blue: Less than -1.5
- Yellow: -1.5 to 1.5
- Red: Greater than 1.5

Created by: Stasa Zivojnovic, 11/30/14
Sources: Code for Oakland, Open Street, Census Bureau, TIGER

Miles 0 1 2
Analysis (1-2 pages)

Median Income and Age, 2000 & 2010

These maps are meant to illustrate Oakland changes in Oakland before the study year of 2010. Regarding the income, the north east portion of Oakland has increased in income. So has the area around downtown and close to San Francisco. The lower west portions of Oakland have become a bit poorer and less segregated based on income. Overall the wealth of Oakland between the 2000 and 2009 census has seemed to have decreased. It should be noted that there was an economic crisis the year in which the census was taken.

The median age seems to have increased throughout Oakland. The downtown areas have become older while the southern portions have become younger. There seems to be a relationship between the income and age of each tract. Those areas whose income has increased has also seen an increase in median age.

Oakland’s Population and Ethnic Distribution

When looking at the individual ethnic maps, it can be seen that the Asian and Hispanic populations are the most clustered together, while the White and African American population is more dispersed. The Asian population is concentrated in the downtown area and a bit to the east. They are located close to the water front of the East Bay. The Hispanic population is clustered more towards the southern region but extending throughout Oakland.

The White and African American Population are both dispersed, but the White population is dispersed more towards the Hillside (north-west) while the African Population is dispersed more towards northern water front area of the East Bay.

A better overview of the ethnic distribution can be seen by looking at the top four ethnicities superimposed in the large map. There we can see that the majority of residents in the north western portion of Downtown Oakland are of African American ethnicity. The Downtown Oakland Area extending eastwards and southwards along the coast is mostly Asian. The southern west portion near the coast is mostly of Hispanic ethnicity while the Southern east portion towards the hillside is African American. The north east corner is dominantly all white with traces of other ethnicity such as Asian. Overall, it can be seen that Oakland is highly segregated by ethnicity.

Student’s Ethnic Distribution

The student ethnic distribution mimics that of the overall population because those maps are based on the ethnic ratio. However, it is visible that there are more students attending elementary and middle school in Oakland than high school, especially in the north eastern regions. There are also many more elementary and middle schools than there are high schools. There is a general lower number of student in the downtown Oakland area meaning that not many students reside in that area.
Crimes of Oakland 2009

The crimes for 2009 in Oakland are focused along the East Bay coast. With large groupings in the downtown area and the southern region. The crime in the north western area are in majority drug crimes while those along the coast are more sex crime related. The southern portion sees mostly drug crimes with a lot of shootings as well. The area inland has much less crime than the coastal region.

Since there is a lot of crime in the downtown area, but not many students living there, that area is the center cluster for the highest crime per student population rate. There is also some in the region spanning the coast, but much less dense. The northern area of downtown also has a high ratio of crime to students.

Oakland High Schools Ranked by Vicinity to Toxic Sites

The high schools near the downtown area and central Oakland are located the closest to a toxic site since the toxic sites span along the coastal region. The southern schools and inland schools are not as exposed to the toxic sites.

School’s Ranking vs. Ethnic Distribution

Ranking the schools based on the four factors, it can be seen that the African American and Hispanic student population are at the greatest disadvantage. The northern area by downtown has the largest crime to student ratio, is of low income, high vicinity to toxic sites, and low API scoring. The only school to surpass the API scoring goal of 800 is in the north east, in a predominantly white area. The crime to student ratio in the central Oakland area is not as bad as that in the downtown area, it is more evenly spread out throughout Oakland. The Hispanic student population is disadvantaged in regards to API scores, income, and toxicity. However, their crime to student population ratio is not as bad as that which the African American students have to deal with in Downtown Oakland.

This maps answers the first question, “Are the high schools in Oakland providing the same learning environments equally to all the different ethnicities in Oakland, based on those four indicators?”

The answer is that they are not, because students of African American and Hispanic ethnicity have to attend school in worse environments. Their daily life includes more toxins, crime, lower income, and poor test scores than the Asian or White student population does.

API Score Prediction

This map has a geographically weight regression with toxic site vicinity, income, and crime rate as the independent variables and the API score as the dependent variable.

Piedmont High School and Wilson College Preparatory Academy are the two schools that scored much higher on the API scoring system then would have been predicted by the three independent variables. Oakland High School scored lower on the
API scoring system than predicted with respect to toxic proximity, income, and crime rate.

All other schools were predicted properly by the geographically weight regression showing that there might be relationship between the three environmental factors and a school’s API score.

Conclusion

Regarding environmental Justice for the Students, the quality of their potential should not be limited due to their ethnicity. Instead, legislature needs to be passed in order to provide a more equal opportunity in education for all ethnicities. The indicators in this project are linked to factors that affect almost every factor in a child’s well-being as seen below:

- ... ARE NEAR TOXIC SITES ➔ DANGEROUS TO THEIR HEALTH
- ... IN CRIME RIDDEN AREAS ➔ THREATENING TO THEIR SAFETY
- ... IN LOW INCOME AREAS ➔ NOT ENOUGH FINANCIAL SUPPORT
- ... HAVE LOW API SCORES ➔ POOR LEARNING OPPORTUNITIES

Unfortunately students of African American and Hispanic ethnicity face the largest setbacks due to these factors, therefore not being provided an equal opportunity to reach their fullest academic potential. However, it should be noted that there might be another layer of correlation between the environment and the ethnicities for which may not have been accounted for in this study.

The toxic vicinity, area’s median income, and the crime per student of the area have shown to have been able to predict most of the school’s API scores. Since the API scores are based on the performance on the students rather than the school itself, it shows that the environment around the school influences the performance of the students since the regression could decently accurately predict the API score.

There are three high schools for which the regression did not predict the API score correctly, meaning that there are other factors for those schools that cause the observed API scores instead of what can be predicted with the indicators. The schools that performed better than predicted by the indicators is Piedmont High School and Wilson College Preparatory Academy.

One factor is that the Piedmont High School is in a more suburban area, which could have potentially caused the children to perform better. Wilson College Preparatory Academy would be an interesting study to study further since the area is resembles any of the other lower performing schools but has scored significantly higher, and is not necessarily correlated to the indicators.
Oakland High School scored lower than would have been predicted by the regression through the indicators of that area.
References

Literature


Data Files
School Enrollment per tract 2010 [excel file]. Alameda County,
Appendix

Data Acquisition

Most of the data came from the Census Bureau in the form of the 2010 census. The 2000 census was also used. The tracts came from TIGER for all of California. The Oakland specific data – crime, schools, and toxic sites came from the following website “Code for Oakland.”

The GIS Steps

Molding the Data

Census Tracts - California

7. I downloaded the Census tracts for all of California for both 2000 and 2010 from TIGER
8. I downloaded the Bay Area Water general outline shape file from DataSF.com
9. The tracts of California had tracts also for parcels that are water, so I had to erase them with respect to the water shape file. I omitted the water shape file for the final report.
10. At this point I have two “clean” shape files for the tracts in all of California. (At this point I still thought I was going to work with the entire Bay
11. I then selected the tracts, for both 2000 and 2010, what seemed like to encompass all the counties in the Bay Area.

Excel File Modifications and Importing

12. I downloaded from the census bureau income data for 2000 and 2010 at the census tract level to include all the counties of the Bay.
   a. I couldn’t not download for all of California since the files were too large
13. Using the income data for all Bay Area Counties, I fixed up the excel file so that it would work with GIS
   a. This part, before I got the hang of it, was extremely time consuming and complicated. The file requirements and layout was extremely important. Therefore it took multiple trial and errors to figure out the format that worked well with GIS
14. To create the proper file, I would delete all columns that were not the data I needed.
15. I then would use the “text to column” tool to convert the long string of words and numbers and separate the phrase “Census Tract 400, Alameda County” into a word per column.
16. I would then delete all the columns that were not the actual census tract number
17. Since the columns in the shape file were 40010 instead of 400.10 like in the excel file, I then multiplied the entire column in the excel file and then copied it in again as values instead of equations.
18. I would convert all the values in the columns to numbers so that GIS will pick it up as a double rather than a string
19. The headers in the columns were saved as general, but they were not allowed to have spaces. Therefore I had to use an underscore instead of spaces.
20. I would then save the file as a CSV because the import into GIS proved to be easier via a CSV than a 97-2003 Excel File.

**Joining CSV Files and Shape Files – The Process**

3. To give the CSV files a spatial reference, I had to join it with the census tract file.
4. The census tract field in the shape file is in the form of a string, so I would create a new field (that is a double) and use the field calculator to just multiply the previous tract labeled field by 1 to get the same number but as a double.
5. Through this process I created 6 shape files as follows:
   a. Income 2000
   b. Income 2010
   c. Median Age 2000
   d. Median Age 2010
   e. Ethnicity 2010
   f. School Enrollment 2010

**Overview of Platform Files**

2. File a-d were for all of the bay area, and the files e and f were only for Alameda County
   a. Files a through d was before I decided to narrow down my project
   b. Files e and f could not be downloaded on the city level, but were downloaded at the county level. (Alameda County).
3. The files that I was able to download as shape files include:
   a. School location
   b. Toxic Site Location
   c. Crime in Oakland 2010
4. I clipped the shape files to match that of the Oakland crime shape file, and resaved all of them.

**Reworking of platform files**

6. I had to work on the ethnicity file in excel because I wanted the percentage of each ethnicity within the tract rather than the actual number of people.
   a. I did this part in excel because it was easier to drag across multiple columns
   b. This created a new csv file which included the percentage of each ethnicity in each tract as well
7. I had to rework the student enrollment file by grouping the students from 1st grade to 8th grade (elementary and middle school) and from 9th grade to 12th grade (high school).
a. I re-added this csv file into GIS and joined it again to the tracts to create the correct representation I was looking for.
b. I then exported it again.
8. I joined the student enrollment file with the ethnicity file. Once joined I exported it so that I could work on it in Excel
   a. In excel I multiplied the percent per ethnicity by the number of students in each group to get the number of each students of a certain ethnicity for each tract
      i. This was done in excel because of the dragging across columns capabilities
9. I reworked the school location file to only include schools labeled as ES (elementary school), MS (Middle School), and HS (High School).
   a. This was done through editor in GIS
10. I joined the school enrollment file with the crime file.
    a. In the new crime shape file I added a new field to represent the rate of crime per student
    b. I exported this file as the new crime file.

**Toxic Site shape file and school shape file**

5. The toxic site shape file had a point location for all the toxic sites found in Oakland.
6. I had to turn on the spatial analyst in the extensions
7. In the environment settings I had to make the processing extent that of the Oakland tract file
8. Using Euclidean Distance from the toolbox I created a raster sheet. The values in the raster were for feet rather than miles.
9. Using the Reclassify tool from the toolbox I input the raster I previously created and created it on intervals based on the 0.25 mile, 0.5 mile, 1 mile, 2 mile equivalent in feet. I chose to omit anything beyond two miles since the schools wouldn’t be affected that much.
10. I then changed the labels in the legend to match how I reclassified and placed the location of the schools on top.
11. I added a field for the high school file to represent proximity to toxic site and were it fell.
12. Using the editor tool, I manually input the category of proximity to the nearest toxic site in which the school fell, for the school shape file. If it was more than 2 miles, I labeled it as 3 miles.

**School Shape File**

3. I joined a new school shape file for income of that tract and also the ratio of crime per student for each tract that it fell in. While all joins have been done through table joins, this join was done spatially
4. Realizing I wanted to do a regression I joined all those separate school files into one.

**Geographically Weight Regression**

3. Using this new school master file that has the API score, crime ratio, tract income, and proximity to toxic sites for each school, I ran a geographically weighted regression (GWR)
   
   a. Dependent variable: API score
   b. Independent Variable: income, crime ratio, & proximity to toxic sites

4. This created a new file.

**Summary of Metadata**

**For all Shape Files**

- Please note that before importing any shape file I projected it first on:
  - NAD 1983, State Plane III FIPS 0403 Feet
  - This projection was chosen since the guidelines for projections in California were found on the following website:

**Potential Sources of Inaccuracy**

One thing that needs to be taken into consideration is that the data for this project is based on 2010 data rather than 2014. Many changes could have happened between then and now.

There could be other factors that influence the student’s API performance than the three I have listed. Such factors could include natural disasters, tax rates, public transportation accessibility, walkability, and health care system of the area. I touched upon the three which I felt were the most obvious and blaring factors that could affect a child academic performance.

This study would have been more accurate if I had the time to incorporate elementary and middles schools as well, in order to increase the sample size and get more accurate results.

The vicinity to toxic sites was done with categories rather than actual distance. Using the actual distance from the school to site would increase the accuracy of this study. Also to consider is that even though a school could be farther away from the toxic sites, if it is surrounded by more than one, it could potentially be more hazardous than just being close to one.