Coding Book Camp project 1

Presented by Jennifer Mayfield, Grecia White, & Kevin Thai Huy

**What school ratings were available for the city of “Austin”?**

We found two websites with relevant data called School Digger: <https://developer.schooldigger.com/> and Great Schools: <https://www.greatschools.org/api/request-api-key>. We choose School Digger since it offered a free option to do API calls. Great Schools didn’t offer free API calls and there was a 3-5 day wait for an API key. In the end we paid for the API calls with School Digger since we quickly exhausted our calls. School Digger only allows you to call one page at a time with up to 50 schools per page and 20 calls per day. Once we started working with School Digger we quickly noticed there were some zip codes that did have any schools or did not have all three stages of primary schools in its zip code. We decided early on to use zip codes as the primary key between all our data sets. We didn’t average school ratings for all schools in a zip code since some school boundaries would cover several zip codes. Finally, we decided to focus on elementary school ratings. Elementary schools located in zip code were more likely to cover a smaller area and fewer zip codes.

The ratings scale was on a scale of 0 to 5. School Digger ranks schools based on test scores supplied by the state’s Department of Education. An average standard score is calculated and is assumed to be the norm. The average is calculated across all tests and grades. Schools ranked in the top 10% are given five stars. School Digger calculates an Average Standard Score by normalizing and averaging each school's test scores across all tests and grades.

We initially found 41 zip codes in the Austin area. Not all zip codes had schools and all three stages of primary education. We decided to focus on elementary schools since their ratings were more likely connected to only one zip code. Our initial data set consisted 324 schools on seven combined pages and dictionaries within dictionaries. Once we sorted and narrowed our data set we had 132 elementary schools. We experimented with different bar charts. It showed there were great disparities between schools in different zip codes throughout the city. There were more zip codes better than average. The overall average star rating was 2.7. 22 zip codes were above this average.

**What census data was available for the zip code and how to get the data for multiple zip codes?**

Yes, but not recent data, they only have population data from 2010 Census which is too far from now. We used for loops to pull in data for multiple zip codes from 2018. The 2018 numbers were estimated, but they are more closely related to the 2017 data we used from the City of Austin and School Digger.

**What crime data was available for the city of “Austin”?**

Through the city's open data portal we accessed the Annual Crime Dataset for 2017 using their API endpoint. We had access to the following: clearance date, clearance status, district, highest offense description, location, zip code, coordinates and highest NIBRS/UCR offense description. From these we used zipcode, highest offense description, highest NIBRS/UCR offense description. We used the sub categories; Theft, Burglary, Auto theft, Robbery and Aggravated Assault from Highest NIBRS/UCR offense description. We merged population data against zipcodes from a different website, zipcode.com. From this we calculated the percentage of crime vs population against zip codes. For each graph we had adjust the number of zipcodes since not every sub category of crime occurred in every zipcode. The adjustment was needed so the length of lists would match and we could create visuals

Future research could be done with why so many rape cases had incomplete information in the City database. I would be helpful to do a heatmap with an overlay of the zipcodes to get a clearer picture of what areas of Austin have low/high school rankings as well as low/high crime rates. It would also be useful to create a heatmap showing which specific crimes have the highest numbers across Austin zipcodes. I wish I had had more time to dig into the 3 different categories for clearance statuses. It would be nice to see which offenses tend to get cleared the most and cross reference that with zipcodes.

**How are all these data sets related?**

We decided to use the data from schools that had the most data and ratings that correlated more closely to zip codes. This left us with 35 zip codes that was verified to my excel results. As we started combining data sets we starting dropping records because we were missing related information.

**What trends and analysis could be observed when we combined the data sets?**

There is a strong correlation between crime rates and school ratings in Austin. We focused on elementary schools which would be more closely tied to one zip code. There are distinct zip codes where you see highly rated schools and low or no crime rate. All schools rated 5 were west of MoPac. The best zip codes with low crime rates and 5 star elementary schools are 78732 (Steiner Ranch), 78733 (Barton Creek), 78738 (Bee Caves) , & 78759 (Great Hills/Arboretum). We found high crime rates and low school ratings in 78722(French Place), 78723(Mueller/Springdale), and 78752(Windsor Park). Some of our outliers included low crime rates and low school ratings in 78728 (Wells Branch) and 78758 (Doman). There are many crimes not reported or don’t result in conviction and are not recorded. Test scores are only one factor that can define a school and how we rate a school. There are limitations on data and what they can show.