I worked with Jacob Jaillet on creating a database of containing the birthplace location of every major league baseball player since ~1880. To do this, we used <https://www.baseball-reference.com/bio/> as our main source for every player’s hometown. We gathered information off the site on a state by state basis as csv files. Using a batch geocoder from <https://geocode.localfocus.nl/> we were able to generate a list of each city a baseball player was born in and that citie’s latitude and longitude.

From there Jake worked on putting this information into QGIS and generating a heatmap and going from there.

I created a database in postgres in PSQL and then used the command line to copy data from the csv files into the database to populate the tables using commands such as:
baseball-> \copy Cities (CID, City, Latitude, Longitude) FROM

'C:\filepath\citiesfile.csv’ WITH (FORMAT CSV, DELIMITER ',',

HEADER TRUE);

I then used queries within postgres to join player data gotten from baseball-reference with city location gathered from the batch geocoder to create a csv that contained each player from the state of Texas with the latitude and longitude of their birthplace city. We chose Texas to start with because it has two major league teams and also contains 3 of the top 10 largest cities in the country that do not have major league teams.

This csv can then be added to QGIS as a delimited text layer with a geometry definition of point coordinates. This will put a point on the map at each location a player was born. A heatmap of major league baseball player’s birthplaces can then be generated:



We were interested in looking at how many players came from cities that had a major league baseball team vs how many players came from cities without a team.

I went back to postgres and got a list of the number of players from the two areas with teams and the 3 largest cities that do not have teams. I looked at the amount of players from each city and then took a percentage of the amount of players from each city against the city’s total population to account for more players coming from a particular city simply because the population is higher.

You can see that the two areas with teams (Houston and Dallas/Fort Worth/Arlington) have higher percentages than the three cities without a team. In the most extreme case, Houston produces nearly 3 times more major league baseball players per population count than El Paso.