

Will my future son have a chance to play MLB?



We plan to have children in near future. I am curious if we have a healthy son, if it is a realistic goal sending him to all sorts of baseball skill learning / training activities in hopes of him becoming a professional baseball player with MLB.

My height is 71 inches and my wife's height 67 in. My weight is 220 lb and, my wife's weight is 130lb. We are both of East Indian descent, right handed, and born in NYC. According to Dr. Chao-Qiang Lai, height primarily comes from genes and can be adjusted slightly based on nutrition, exercise and lifestyle changes. Height is directly related to heritability. Chinese have heritability of 65% and will use this as a marker as Chinese are closest to East Indians' heritability % in terms of geographical factor since exact East Indian heritability is unknown. 35% will be contributed by environment and nutrition. East Indian males have average height of 65 in., and females have average of 60 in. Weight can be adjusted by nutrition, exercise, and lifestyle. According to A. Chris Gajilan (Senior Producer, Medical News), there is a 90% chance our future son will be right handed.



Analysis:

The expected height for my future son would be:

- Expected height difference from the East Indian male population mean = $(0.65 \times ((71 - 65) + (67 - 60))) / 2 = 4.2$ in
- Expected height is $65 + 4.2 = 69.2$ in
- Additional height can be gained from environmental and nutrition = $(0.35 \times ((71 - 65) + (67 - 60))) / 2 = 2.3$ in
- Therefore, with additional optimal environmental and nutrition factors of USA, my son may grow to $69.2 + 2.3 = 71.5$ in

According to the height vs weight plots of MLB players from 2011-2016, this height is acceptable. My future son would need to have a weight of approximately 200 lb to be ideal. In position vs height plots, it seems there are more lefty pitchers. Therefore, as my future son has 90% chance of being right handed, he should concentrate on another position. Also, it seems there are more ambidextrous catchers, and this would also not be a great fielding position. According to histogram of height vs salary, out of top 100 MLB salaries, my future son's projected height made \$17,500,000 for 2016, but that does not consider how many years a person has been playing.

I considered most people born in a particular state grow up in that state within the USA. Choropleth map shows that I need to move to California or Pennsylvania to raise my future MLB son. I am praying my future computer scientist job will take me to California :-) to root for my favorite current athlete Albert Pujols of LA Angels and groom an MLB player of my own! I did not consider population or boundary area per state while looking at birth state density of MLB players. I considered total of baseball players born from each state from 1871-2016 and did not consider by year to see if there is a trend.

Citations:

1. Lahman, S. (2017, February 26). Download Lahman's Baseball Database. Retrieved May 1, 2017, from <http://www.seanlahman.com/baseball-archive/statistics/>
2. MLB Rankings. (2017). Retrieved May 1, 2017, from <http://www.spotrac.com/mlb/rankings/>
3. Gajilan, A. (2007, August 13). Left-handedness and your health. Retrieved May 1, 2017, from <http://www.cnn.com/HEALTH/blogs/paging.dr.gupta/2007/08/left-handedness-and-your-health.html>
4. Lai, C., MD. (2006, December 08). How much of human height is genetic and how much is due to nutrition? Retrieved May 1, 2017, from <http://www.scientificamerican.com/article/how-much-of-human-height/>

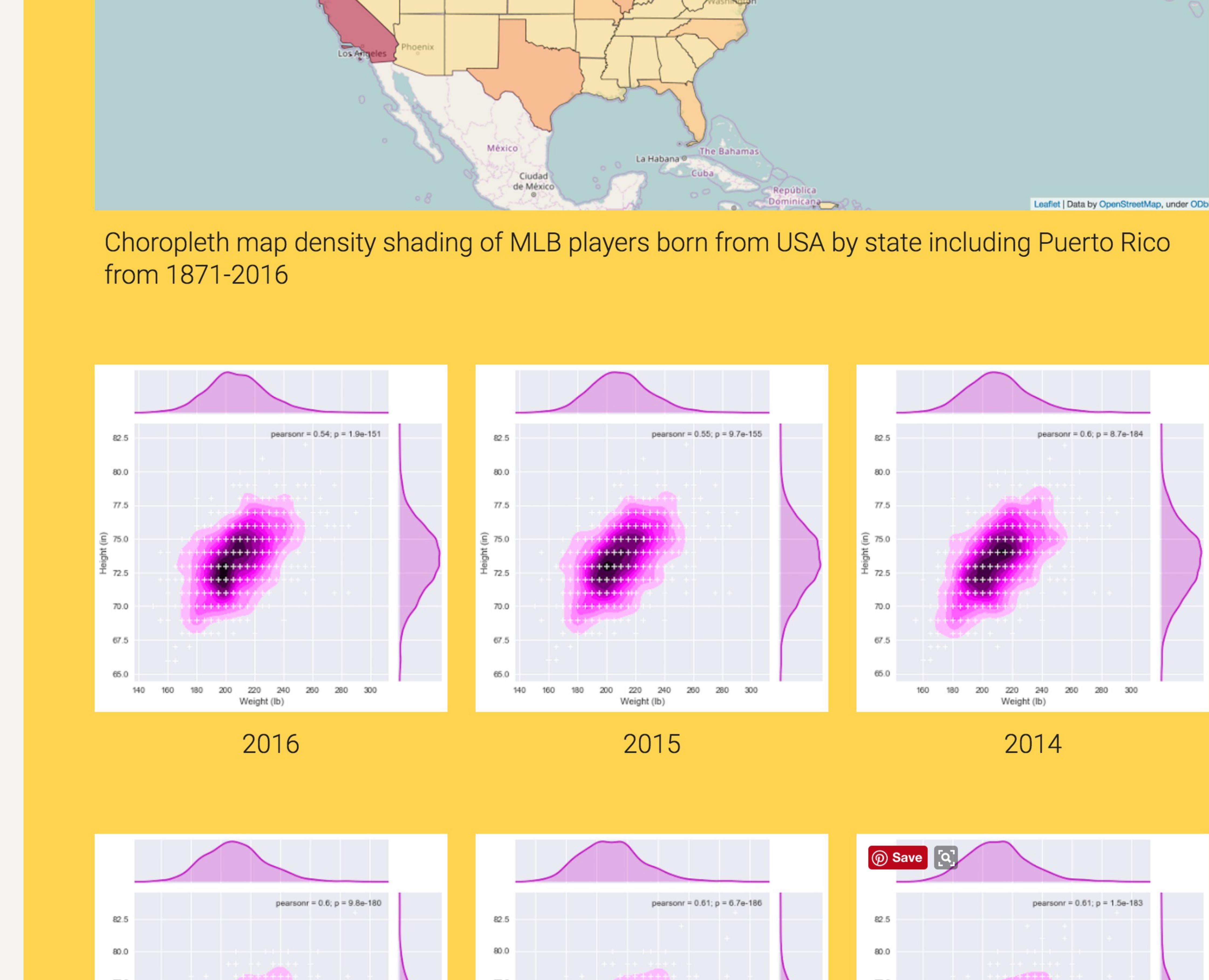
5. Grus, J. (2015). *Data science from scratch: first principles with Python*. Sebastopol, CA: O'Reilly.

GitHub:

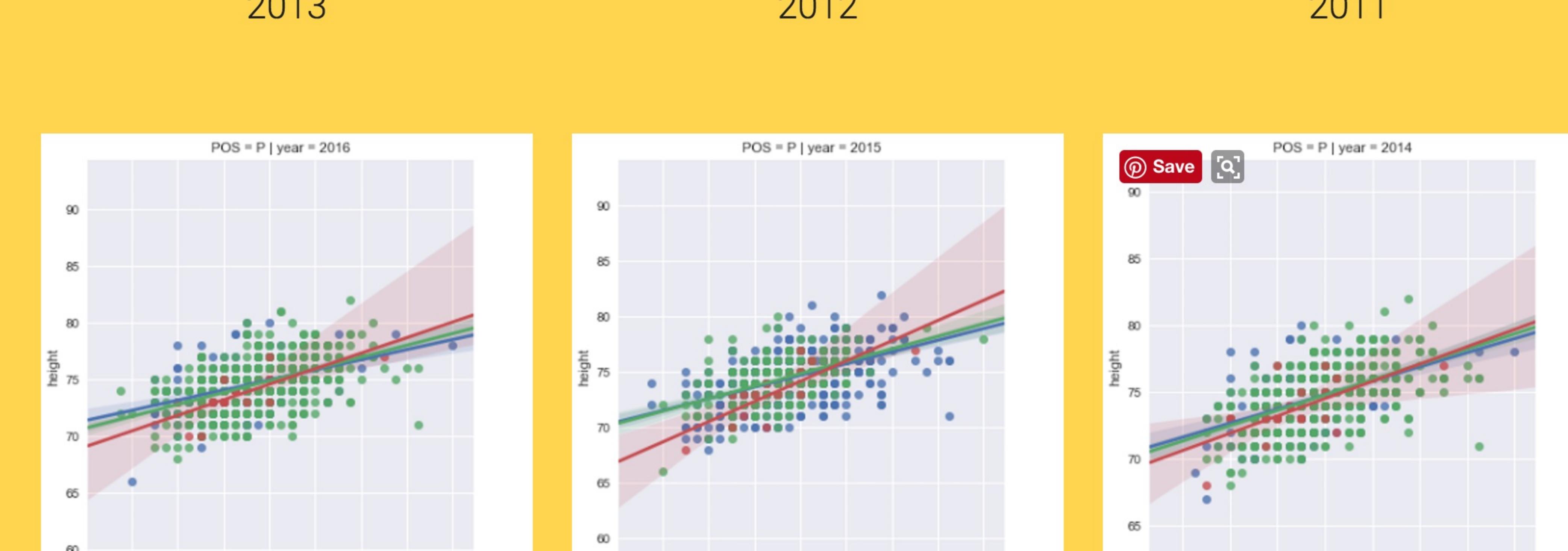
Data-Science-Project - Analysis of MLB data sets

github.com/syamas02/Data-Science-Project

Plots



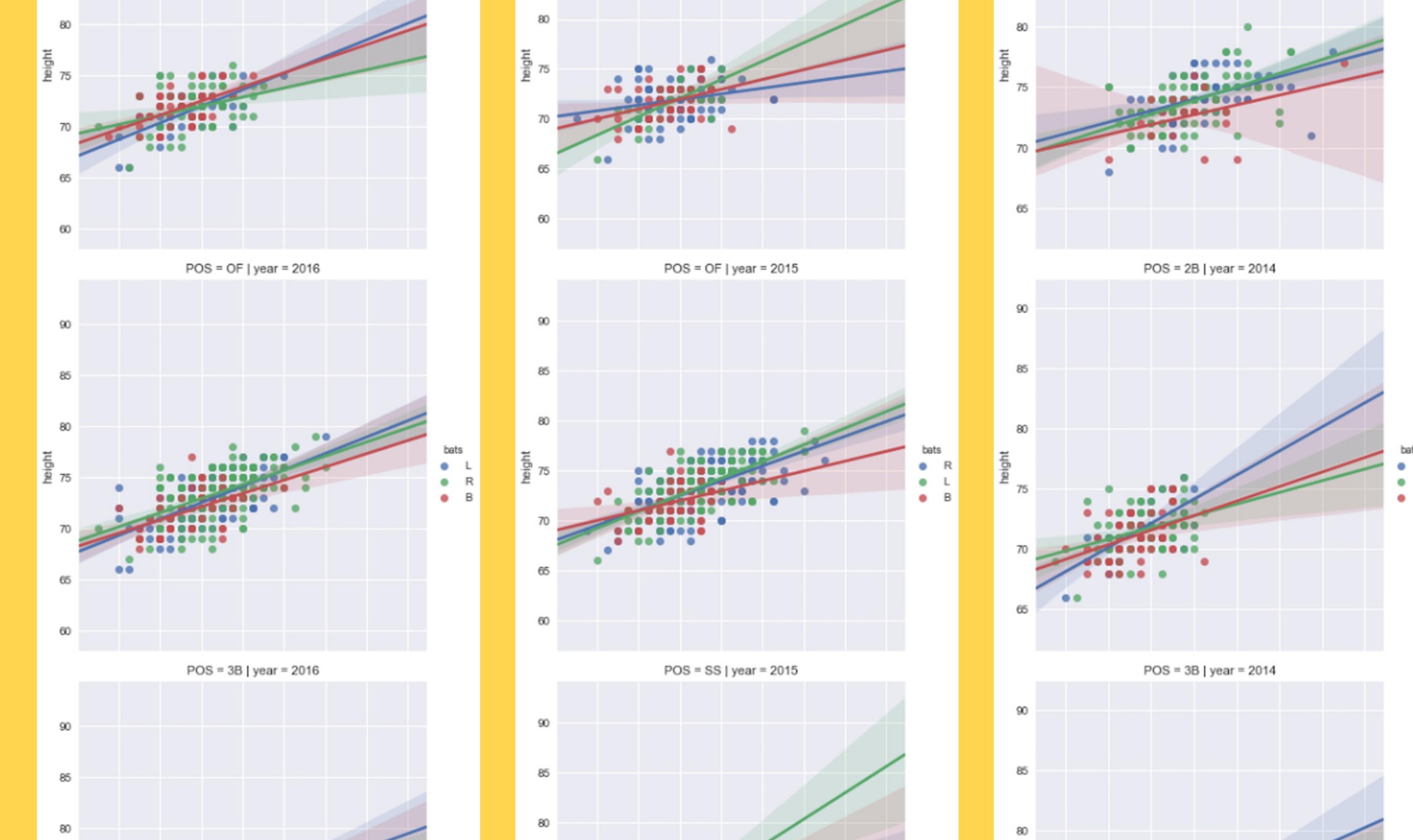
Choropleth map density shading of MLB players born from USA by state including Puerto Rico from 1871-2016



2016

2015

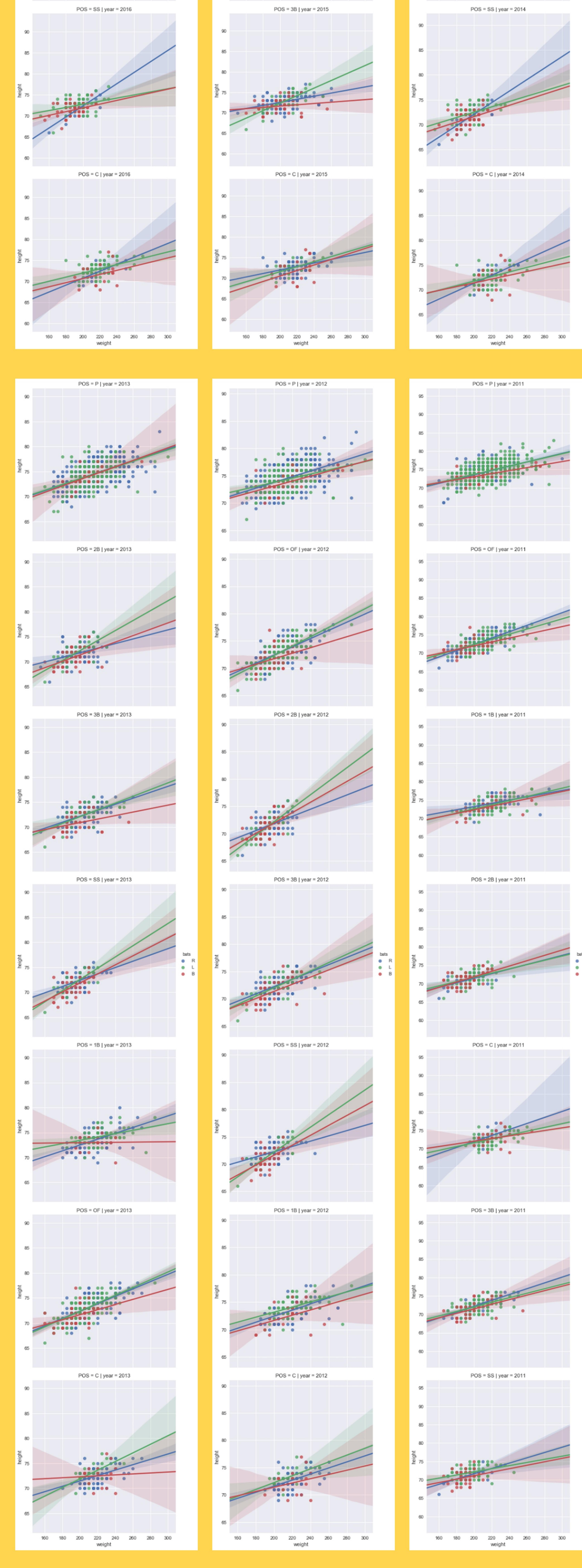
2014



2013

2012

2011



year = 2016

year = 2015

year = 2014



year = 2013

year = 2012

year = 2011

Techniques

I worked on this project by myself using Python 3.6 in Spyder IDE and testing via Jupyter Notebook. I learned a lot about data frames, pandas, and beautiful soup as I had to web scrape, manipulate, retrieve, and combine data from html and CSV files. While trying to web scrape, I noticed, not everything is consistent in html, and more thinking is needed. I used regular expressions to eliminate commas and dollar signs from salaries. I also learned, not every container or collection is ideal with holding and retrieving data for plots. At first, I was trying to put everything in a dictionary as dictionary inserts and retrievals are O(1), but creation and retrieval was too complex for plotting. Then I moved to data frames which is ideal in using seaborn library for plotting. I had first started plotting with matplotlib but seaborn has a lot better visuals for understanding stats. It's also easier to manipulate. I even started taking a course (<https://www.udemy.com/python-for-data-science-and-machine-learning-bootcampin data science>) which I am a 1/3 through to strengthen my knowledge of python data science tools. I also learned google sites is an easy way of publishing a web site. I learned, it's not easy to find data on some of the real problems you would like to solve. I had to improvise my prediction as there is need for more experimentation and readily available data. I also finally placed items in my Github account!