07DEC2019

Project 2: NHL Stats Project Proposal

Scope:

The scope of our project was inspired by previous research on the birthdate effect in higher level and professional sports. The birthdate effect is a trend observed by social workers and psychologists as a relationship between birthdate and academic or athletic success. The trend shows that based on the marker for achievement, being born at a select time of year can be a benefit or hinderance. For sports, mental maturity as well as physical size are advantageous toward athletic success. The NCAA reported on their website that a correlation can be found in the proportion of student athletes with birth dates at specific times of the year and success in their chosen sports (<http://www.ncaa.org/about/resources/research/birthday-effect-college-athletics>). Male athletes born in the beginning of the calendar year represent a large proportion of NCAA athletes in ice hockey and tennis. However, a large proportion of male athletes in basketball, baseball, football, and soccer have birthdays around the beginning of the school calendar. This observation was further studied in Malcom Gladwell’s novel, Outliers: The Story of Success. In the novel, Gladwell discusses the observation that a large proportion of Canadian hockey players in the National Hockey League have birth dates in the beginning of the calendar year.

In this study, we wanted to further these studies by determining if this trend was true for NHL players of other nationalities. We also wanted to determine if birthdates at the beginning of the calendar year correlated with a higher success rate when the athlete makes it into the NHL. Based on Gladwell’s study, there are many successful outliers to the trend. We defined success as number of games played, career total for goals, and career total for assists.

Our data came from a Kaggle source (<https://www.kaggle.com/martinellis/nhl-game-data#table_relationships.JPG>), which ultilized several csv files to develop an SQL database. Of these csv files, we utilized four to create our own SQL database: “player\_info.csv”, “game.csv”, “game\_skater\_stats.csv”, and “game\_goalie\_stats.csv”. The web source included a useful ERD displaying how the data sets were connected. The file does not include an explanation of the metadata; however, the terms are common in hockey data. A few that may need clarification are as follows:

For the “game.csv” file:

Column “game\_id” is a numeric value assigned to each game played.

Column header “type” refers to regular season verses playoff game.

In columns “home\_team\_id” and “away\_team\_id”, a numeric value was assigned to each team.

Column “venue\_link” refers back to the api website the Kaggle source used.

For “player\_info.csv” file:

Column “player\_id” is a unique numeric value assigned to each player.

The column “link” is a web address to the original api.

For “game\_skater\_stats.csv” file:

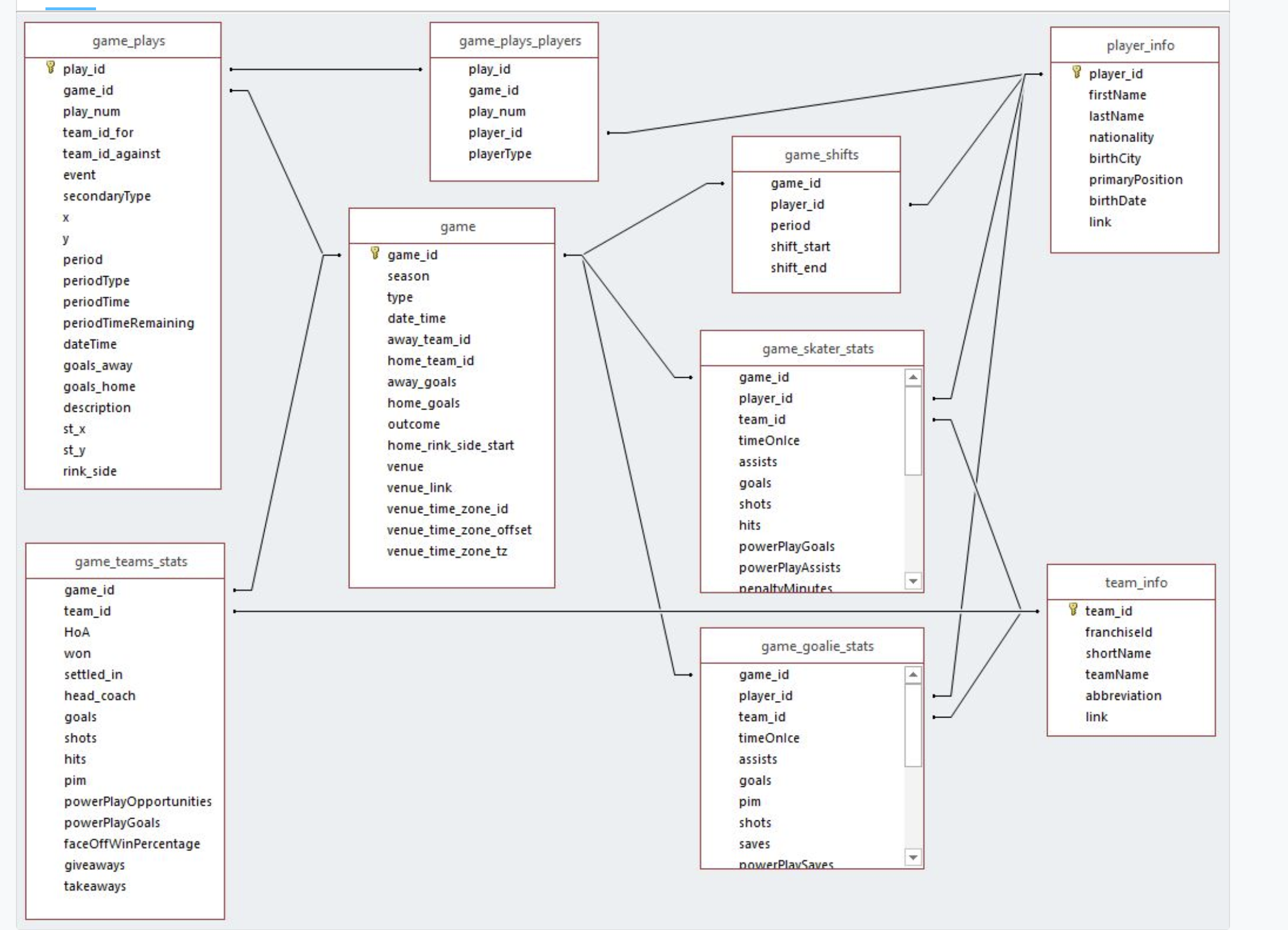
The columns “game\_id” and “player\_id” are foreign keys to the two tables above.

For “game\_goalie\_stats.csv” file:

The term “pim” stands for penalties in minutes.

The column “decision” refers to a win or loss.

Image 1: ERD included with the data files on Kaggel.



Visualization Inspiration:

Image 2: Inspiration for scatter plot of birthdate vs. count selectable by nation of origin.

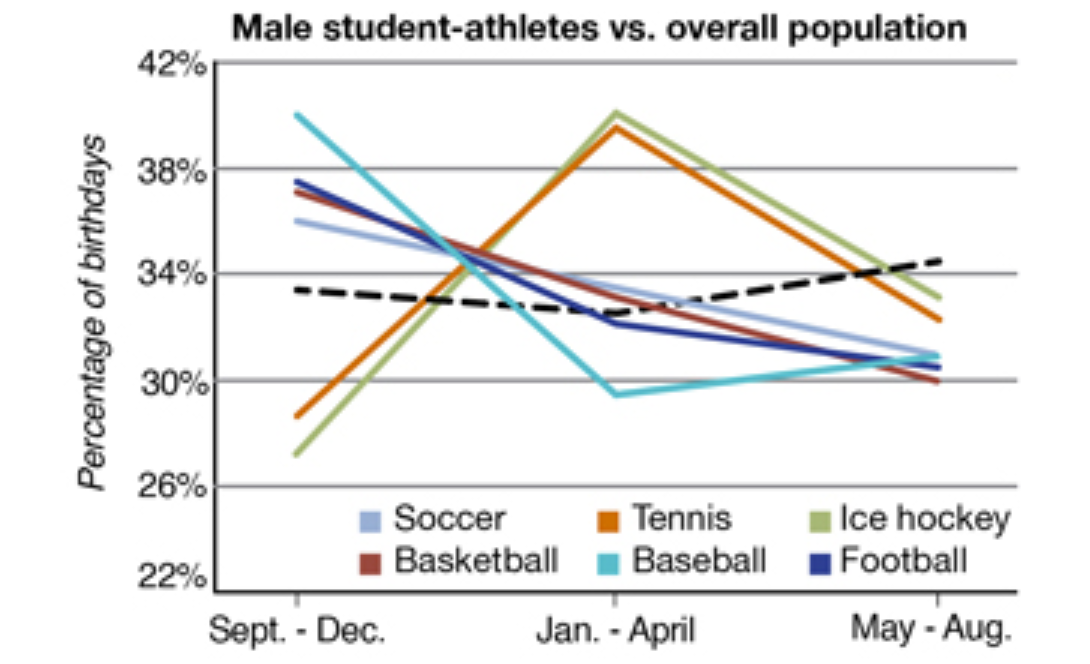


Image 3: Example of simple scatter plot.

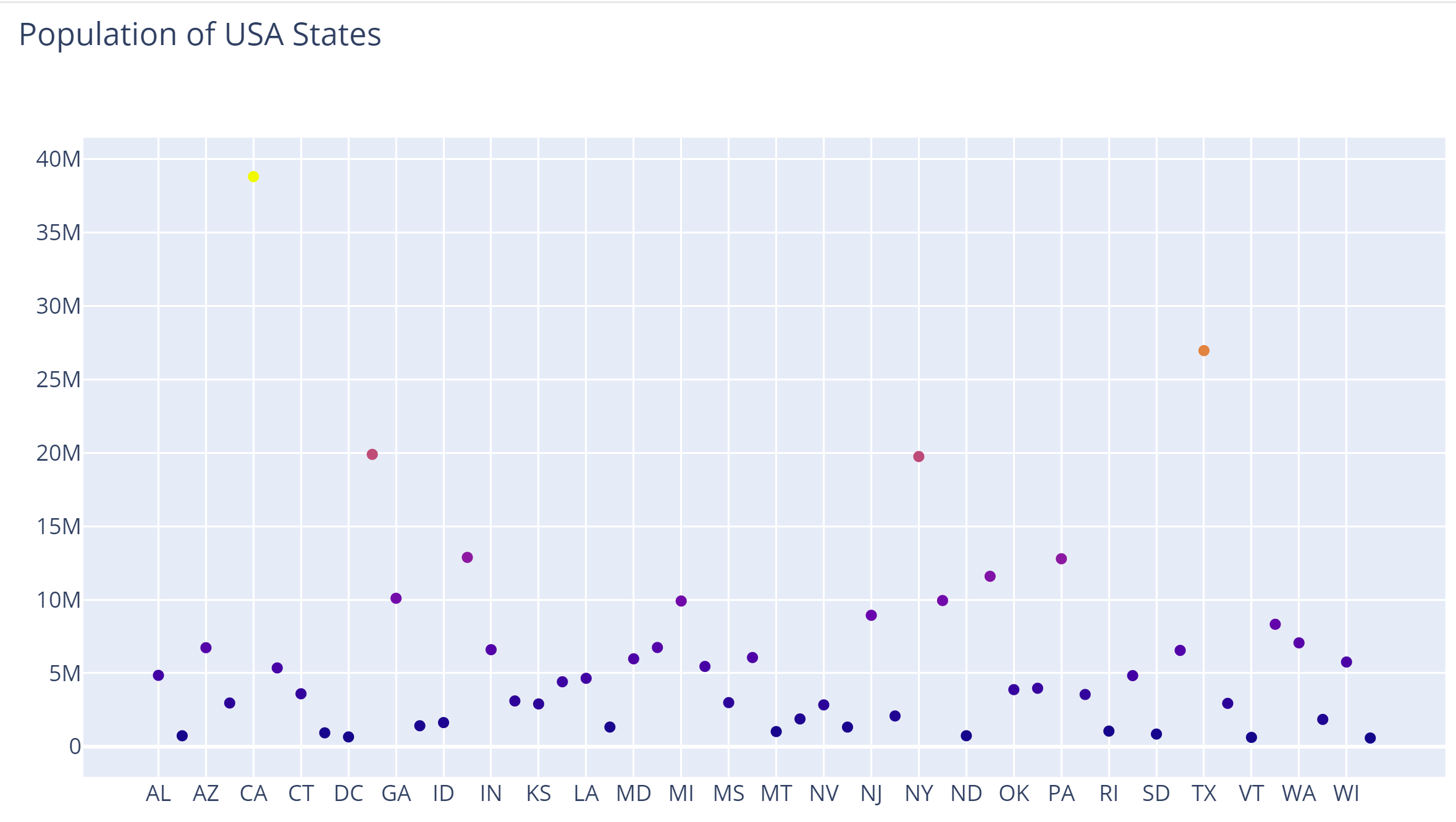
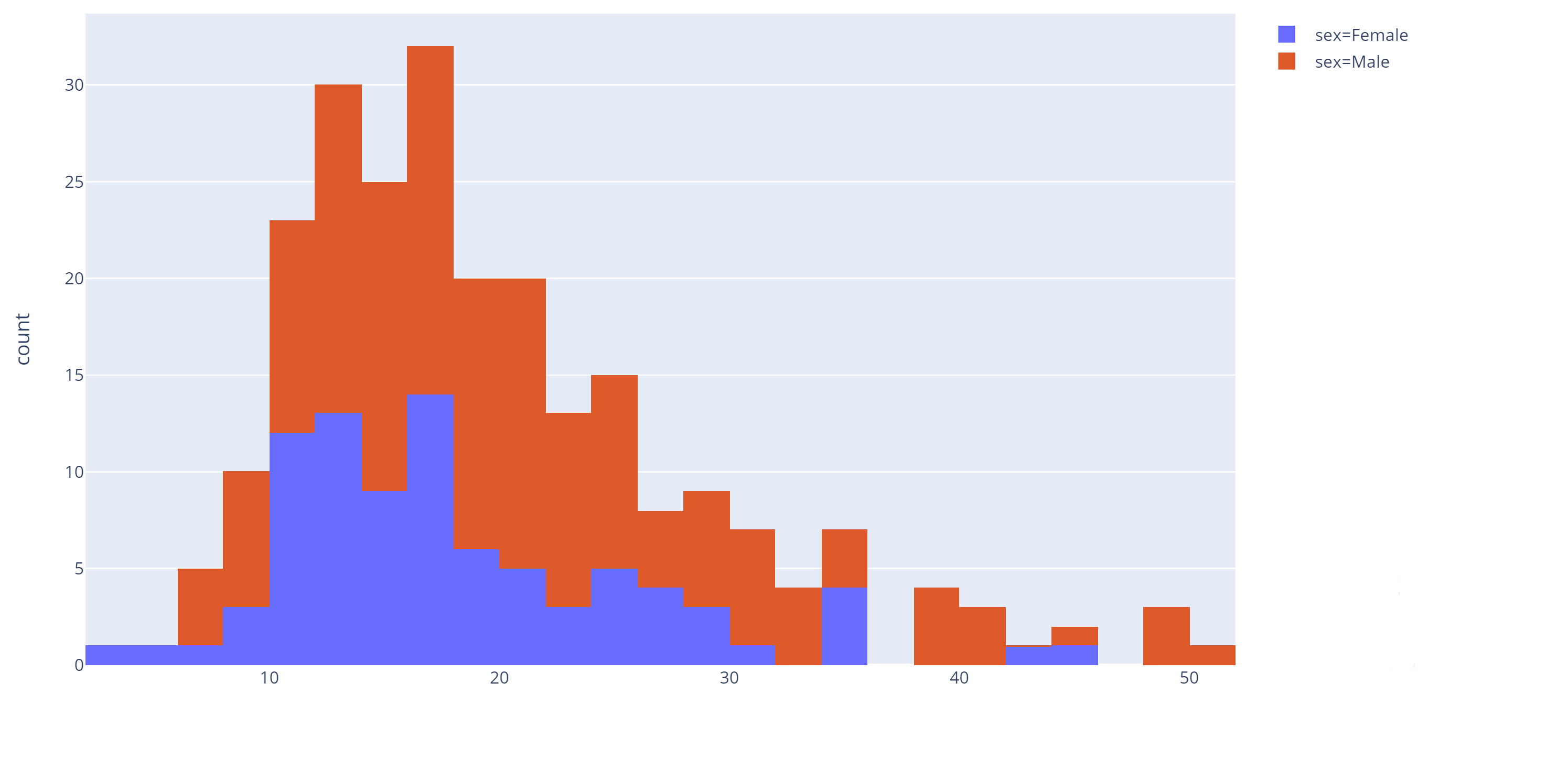
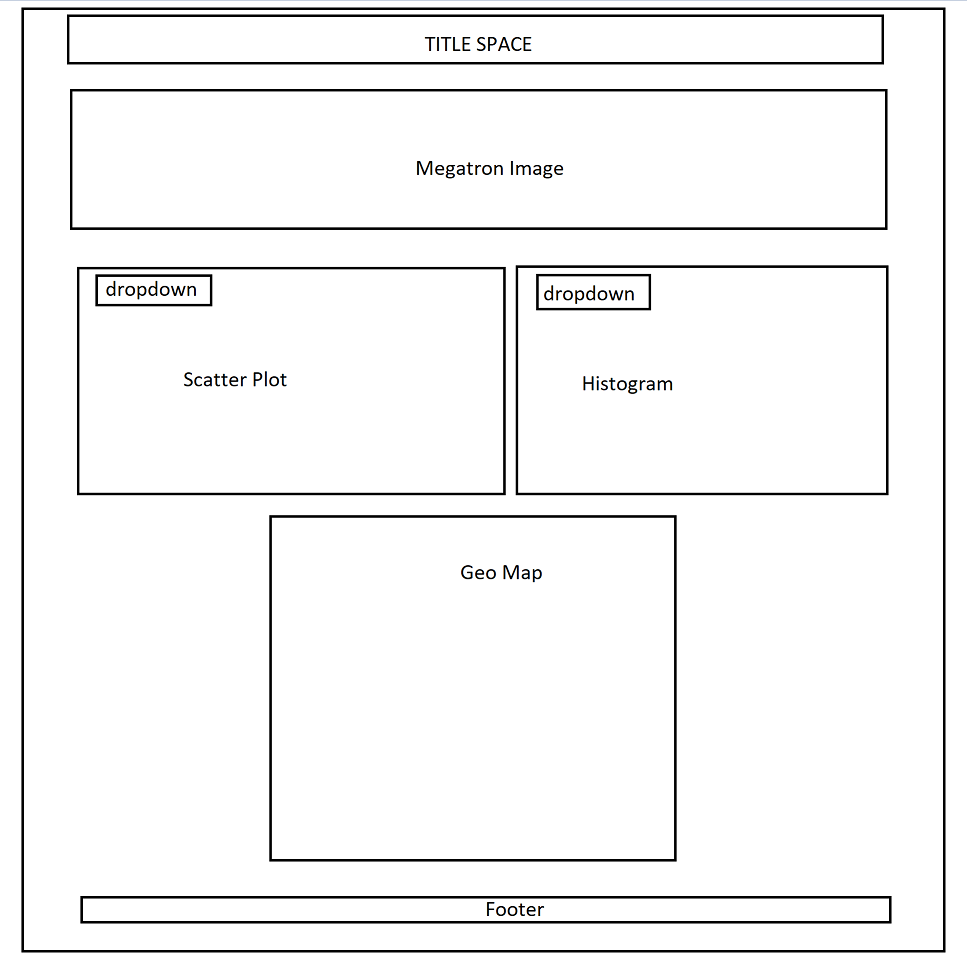


Image 4: Histogram example to be used for markers of success, selectable by nation of origin.

Proposed Layout:

Image 5: Proposed Web Page Layout.



The final project can be found on the github repository <https://github.com/gwathen2019/Project2_NHL_stats>).

Project Requirements:

* RESTful Flask App (provides the visualization to the host) **Gabbie**
* HTML/CSS **Greg**
* JavaScript **Clay**
* JS Library not used in class **Clay**
* At least one DB (SQL w/ Postgress) **Clay**
* A dashboard page w/ multiple charts that update from same data source ALL
* Plotly (scatter plot and histogram) **Megan**
* Data contains at least 100 entries ALL
* User Interaction: menus, dropdowns, textbox (included in our chart design) ALL
* At least three views ALL
* Proposal write up **Gabbie**
* Presentation Format/Set up ALL

Additional Opportunities:

If time permits, other points of interest:

* GeoMap of player origins **Greg**
* Add a chart of the top 10 outliers (late birth, high success) **Gabbie?**
* Most successful hall of famers vs birth date
* Web scrap other source for data **Clay?**
* Use to predict the success of current players
* Additional point of success: salary, playoff games played, Stanley cup rings **Gabbie?**
* Alternative chart types **Megan?**
* Look to adding plus/minus as a measure of success **Gabbie?**