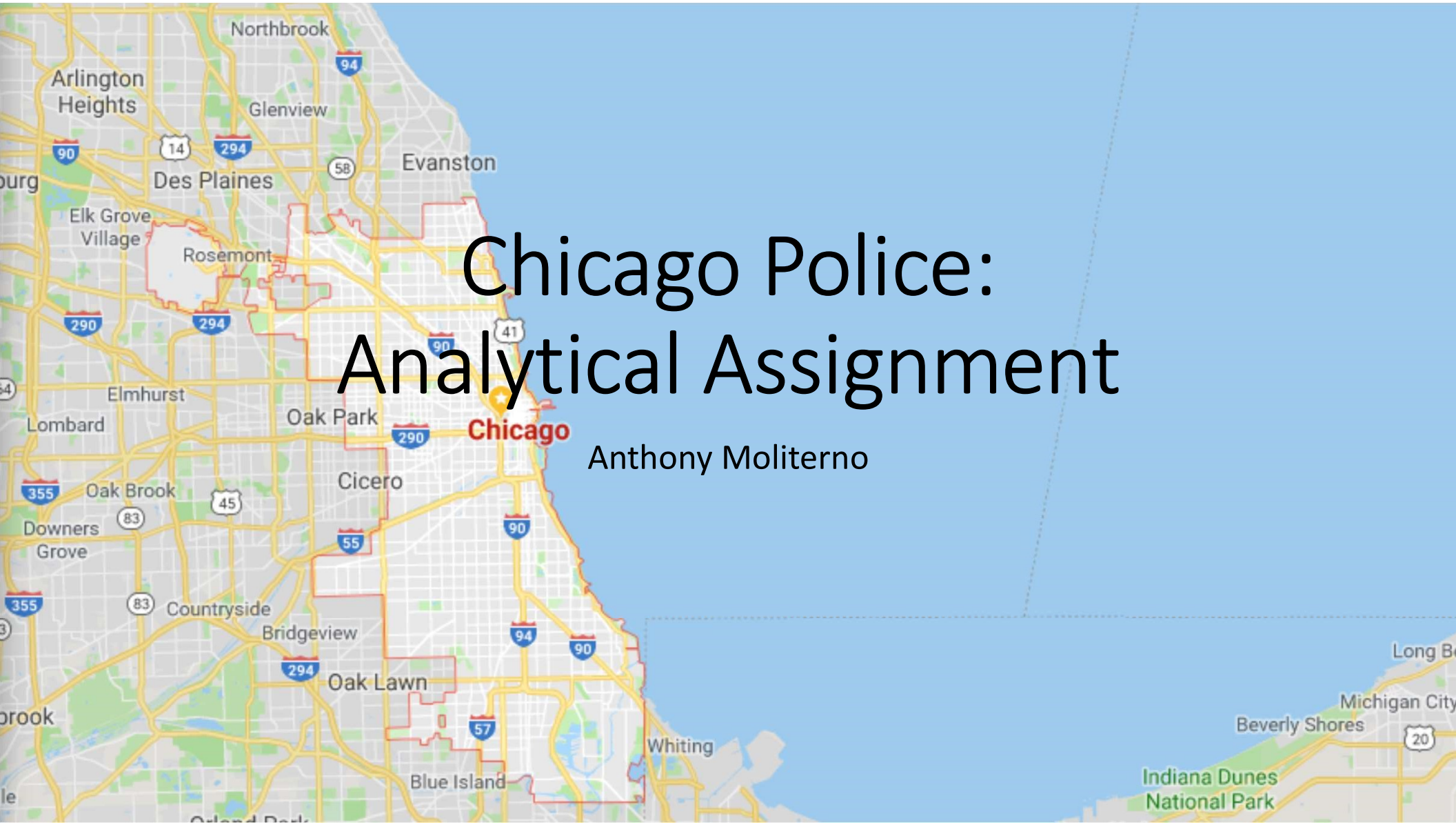


Chicago Police: Analytical Assignment

Anthony Moliterno

Objective

"For this exercise, attached please find a dataset that includes minimally-cleaned information about recreation activity participation in the Girls Play Los Angeles (GPLA) program. ***The GPLA program offers subsidized sports and fitness activities for girls ages 8-15*** *in order to increase young female participation in the City's recreation offerings and promote healthy lifestyles, especially among historically underserved communities. Yet, in practice, how well are the City's most vulnerable communities currently served by GPLA? Using the attached dataset (which has been modified for confidentiality), please* ***illustrate how GPLA addresses equity concerns, as well highlight potential gaps and areas for improvement****."*

Executive Summary

- Good news – the program is very equitable, in terms of…
  - District income vs. Enrollment
  - Age Range vs. Clubs offered
  - Hours of Clubs offered
  - Supply and Demand for sport types

- Areas for improvement:
  - District population vs. Clubs Offered
  - Seasonal offerings

- Caveats:
  - The data only provides a **small glimpse** into Clubs offered and Enrollment patterns, therefore all conclusions **should be taken with a grain of salt**
  - Important information was absent which could be significant, such as **weekday schedules** and **student participation**.

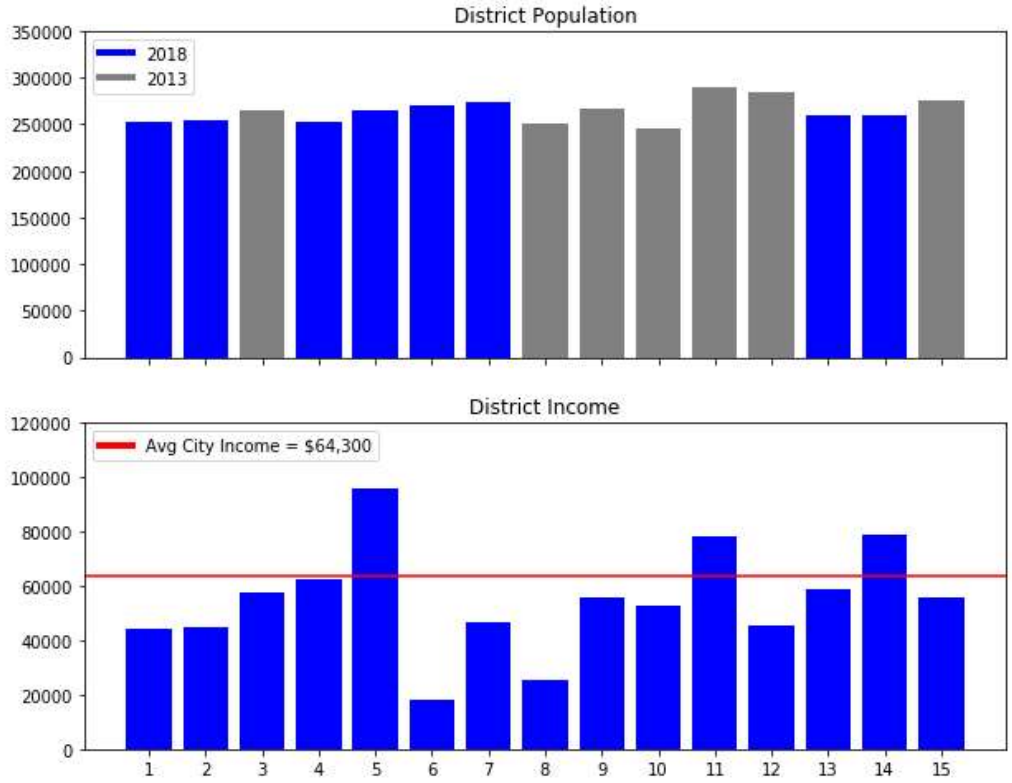
Data Review

- Only 2017 data available – cannot make meaningful time trend comparisons due to seasonality effects
- Data Cleaning
  - Possibly input errors:
    - Some sports have 12-hour time spans, what does this mean?
    - One datapoint had zero enrollment but “1000 females” listed, error?
  - Definitely errors:
    - Typos were fixed (ex: “Basketbll”)
    - Seasons had some NA’s
    - Daytimes had inconsistent upper/lower casing
- Factor Engineering:
  - Added income + population
  - Club times, and number of hours

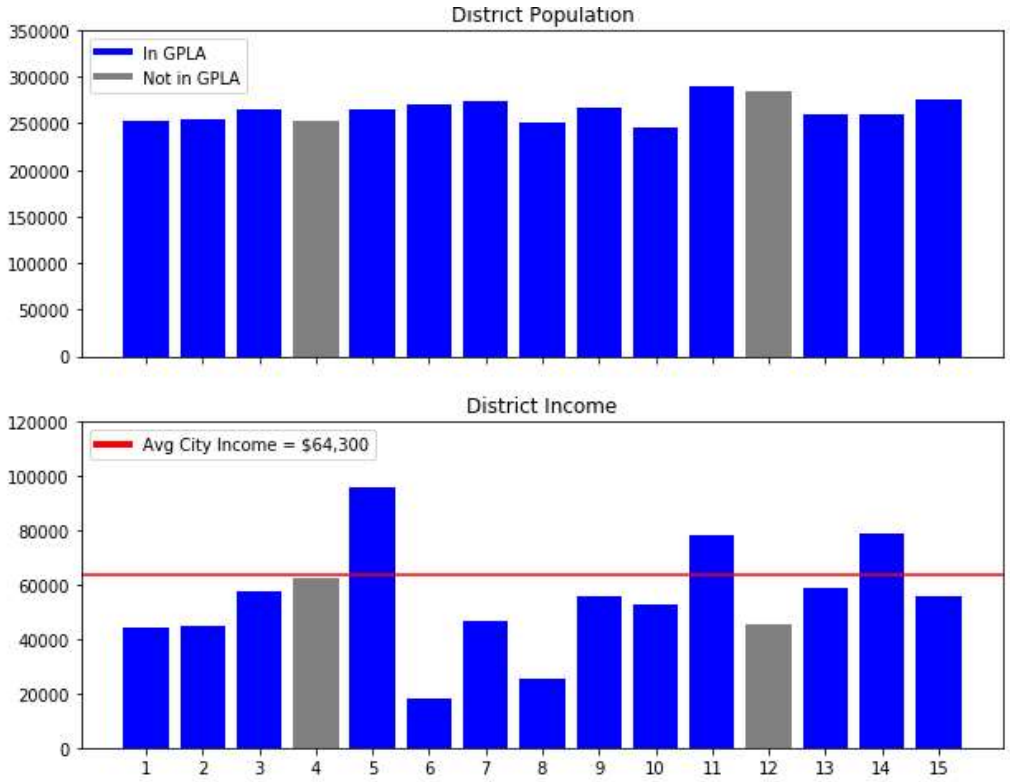
External Data Source: District Population + Income

- Population & Income report from LA Chamber of Commerce
  - Reports **2018** data (one year ahead of GPLA's 2017 data)
- Populations for Districts 3,8,9,10,11,12, and 15 were absent, and **so used 2013** data from LA Controller site

Populations Are Similar Between District, But Income Less So…

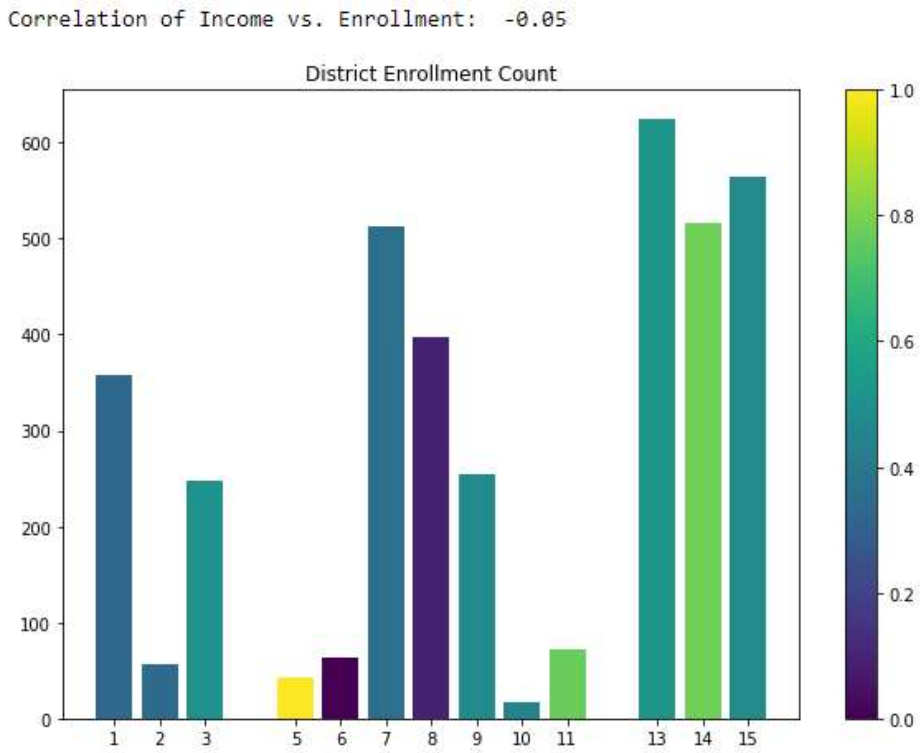


GPLA Represents Most Districts, Which Is Positive



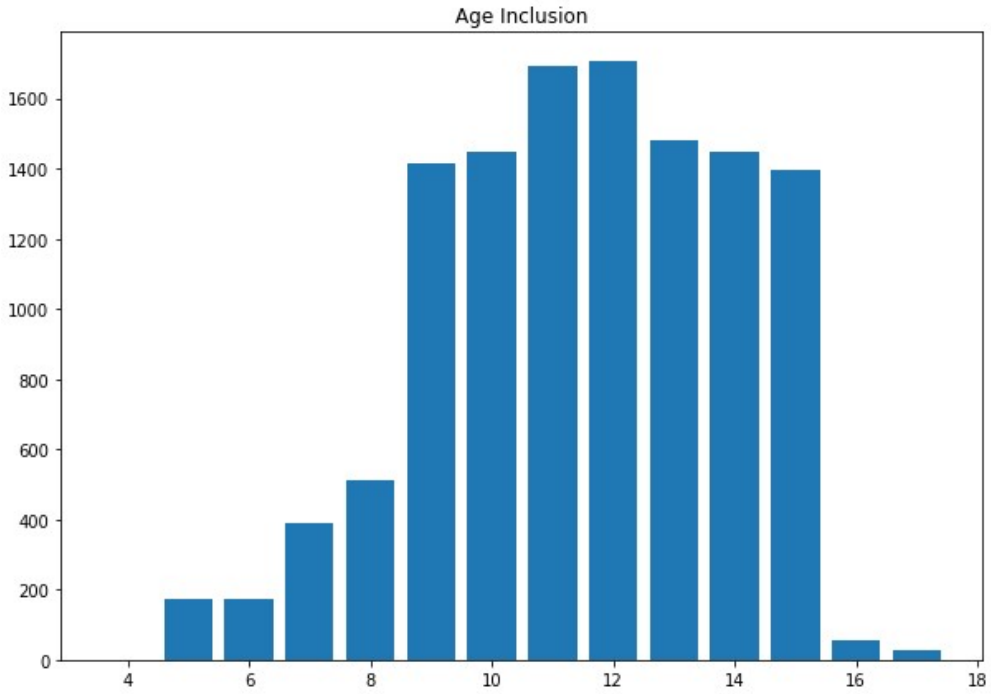
No Apparent Bias to High-Income Districts

- High variance in enrollment by district
- However, enrollment is NOT correlated to income
- What else could explain the variance?



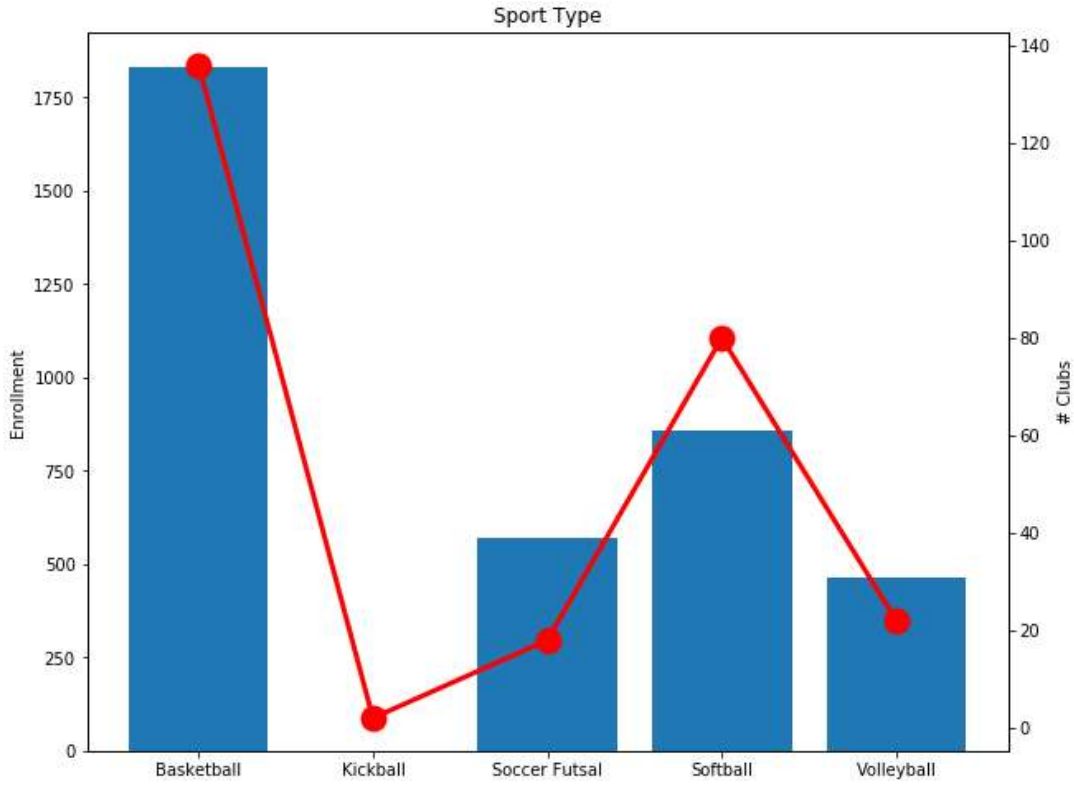
Opportunities by Age

- Relatively fewer opportunities for the very young to participate
- Older enrollment sharply drops off – graduated to high school?
- This *may or may not* be an issue, depending on the demand for opportunities
  - Would need data of individual player ages to answer this



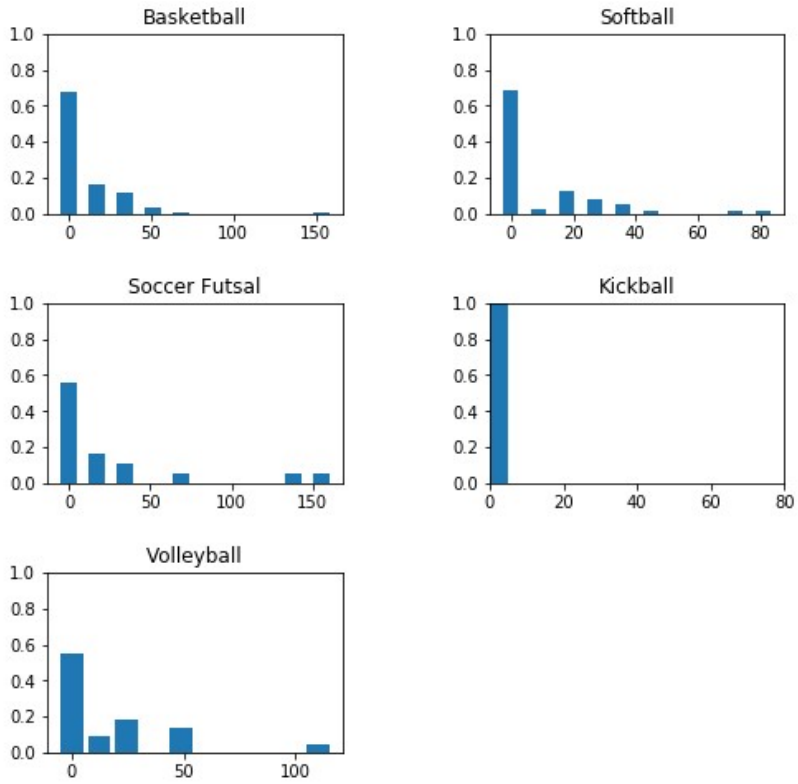
Supply & Demand for Sport Types

- Sport Enrollment vs. Club offerings
- Some sports less costly to start than others.
- Some sports require different numbers of players than others
- Hard to tell which sports are under/over utilized…



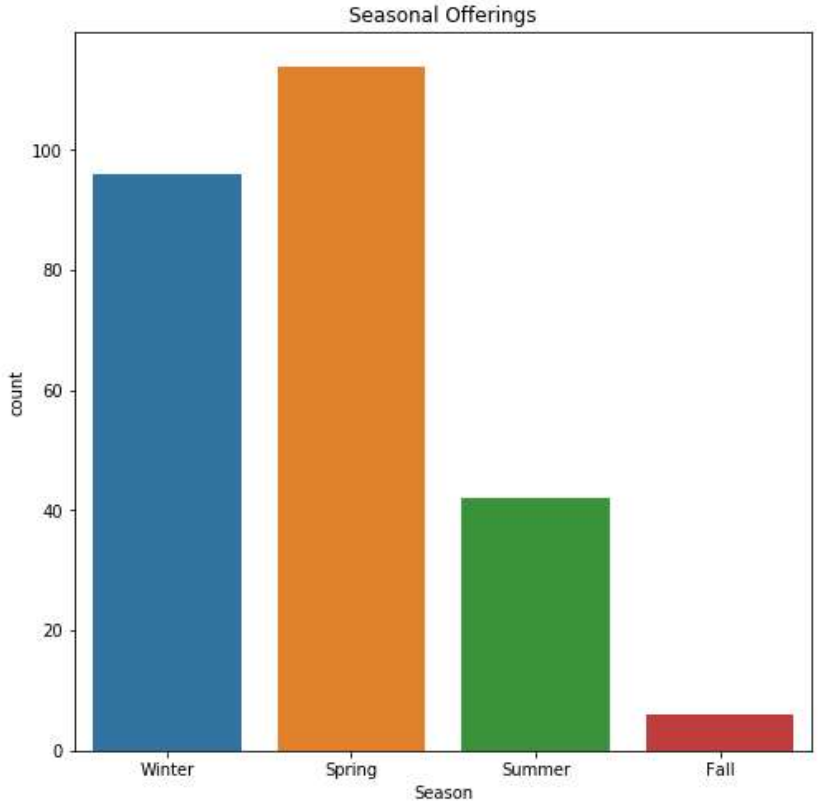
Supply & Demand for Sport Types

- Rearrange visual to show distribution per sport
  - Y-axis show %
- More Sports are available than the children choose to join
  - Ex: No kickball team formed
- Relative under-capacity in volleyball, over-capacity in Basketball



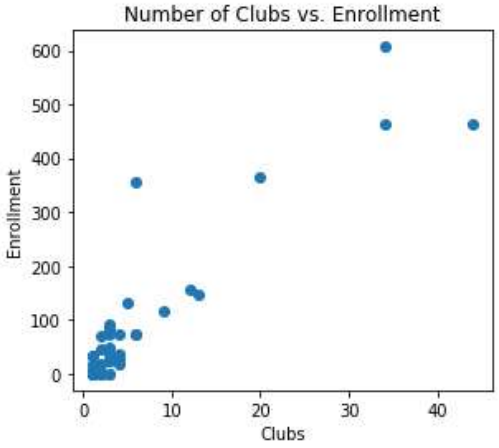
Supply & Demand for Seasons

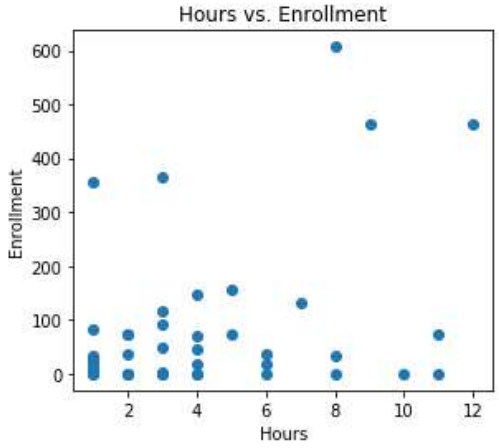
- Later half of the year has a relative lack of opportunities
- Possibly an area for improvement
- Especially useful for when kids are off on summer vacation or just beginning a new school year



Do Hours matter?

- While it makes sense the more Clubs correlates to Enrollment, the relationship for Club Hours is much weaker
- Club Timespans concentrated around 9-5pm, but **vary significantly**
- Do weekdays vs. weekends matter? Can this be improved?





	Clubs	Enrolled	Hours
9:00 am-5:00 pm	34.0	607.0	8.0
9:00 am-9:00 pm	34.0	464.0	12.0
9:00 am-6:00 pm	44.0	462.0	9.0
5:00 pm-8:00 pm	20.0	364.0	3.0
4:30 pm-5:30 pm	6.0	356.0	1.0
9:00 am-2:00 pm	12.0	156.0	5.0
5:00 pm-9:00 pm	13.0	148.0	4.0
9:00 am-4:00 pm	5.0	131.0	7.0
6:00 pm-9:00 pm	9.0	116.0	3.0

Further Work & Areas for Improvement

- Area
  - Districts 4 and 12 did not have any enrollment, are they underserved?
  - High variance in enrollment by district, while populations are similar
- Times
  - Summer and Fall have relatively few opportunities
  - Better assess best hours for students by using most popular times (broken down by weekday vs. weekend times)
- Further questions
  - 80-20 rule? Could there be a concentration in relatively few kids playing most of the sports?
  - Could park proximity to district correlate to enrollment?
  - Do age demographics vary much between districts?