



# Vital Statistics of Professional Athletes

Eric Kuo

General Assembly

Capstone Project



# Problem Statement

What impact does a career in professional sports have on the life expectancy of an athlete?

- ▶ According to Wikipedia article about Football Players:
  - ▶ The average life expectancy or lifespan of an American football NFL player has been reported to be extremely low, only 53 to 59 years depending on playing position.
  - ▶ However, a 2012 study reported that retired NFL players have a lower death rate than men in the general population.

# Some Ways To Answer That Question...

- ▶ Can we build a model to predict how many athletes will die in a given year?
- ▶ How does the distribution of ages at death of athletes compare to that of the general population?
- ▶ Of athletes who were born in a certain time period, can we predict whether they are still alive, or how long they lived based on the length of their playing careers or their physical dimensions?

# Collecting Player Data

- ▶ Player data was collected from the following sites:



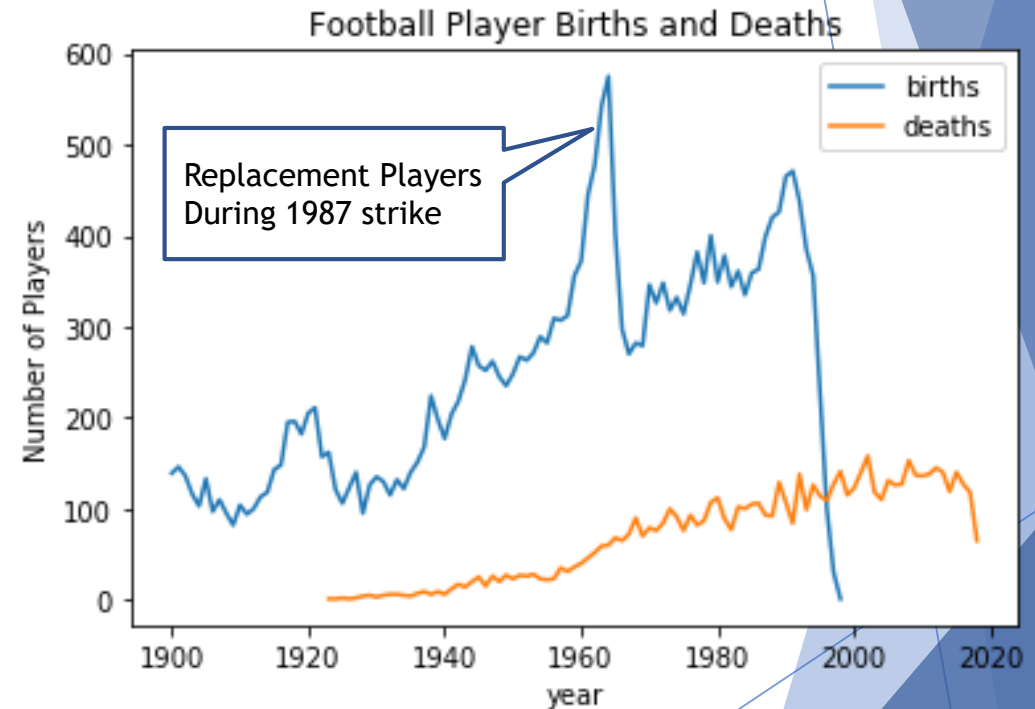
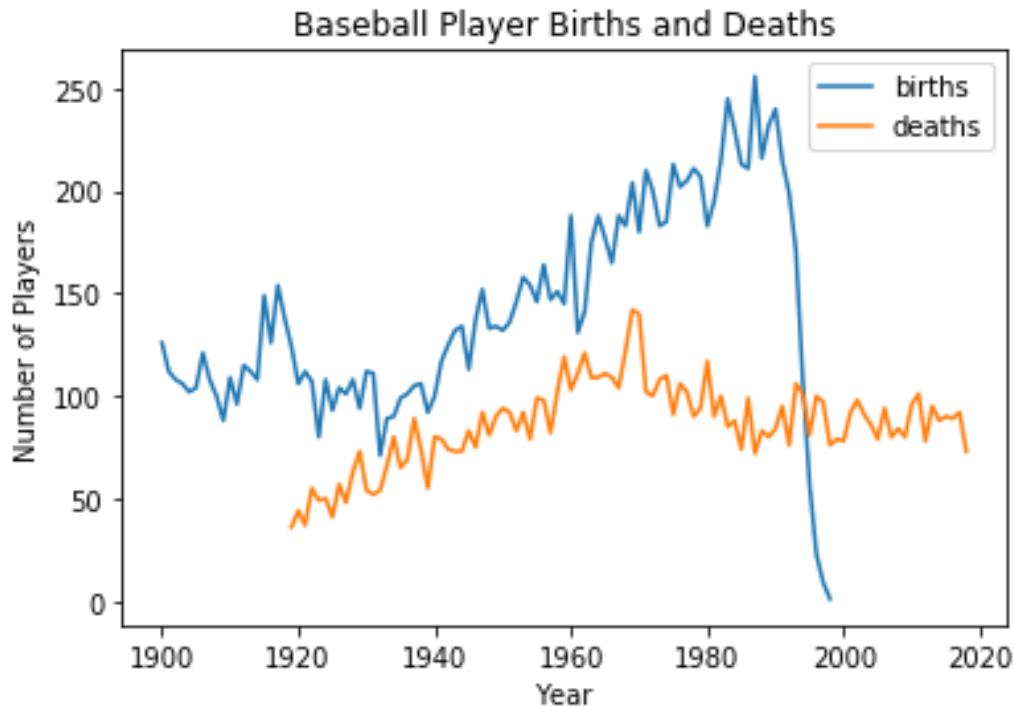
- ▶ From each site, data was collected (or inferred) about each player:
  - ▶ Date of birth and death
  - ▶ Age at death
  - ▶ Length of career: number of seasons and games played
  - ▶ Height and weight (for selected baseball and football players)

# Vital Statistics of General Population

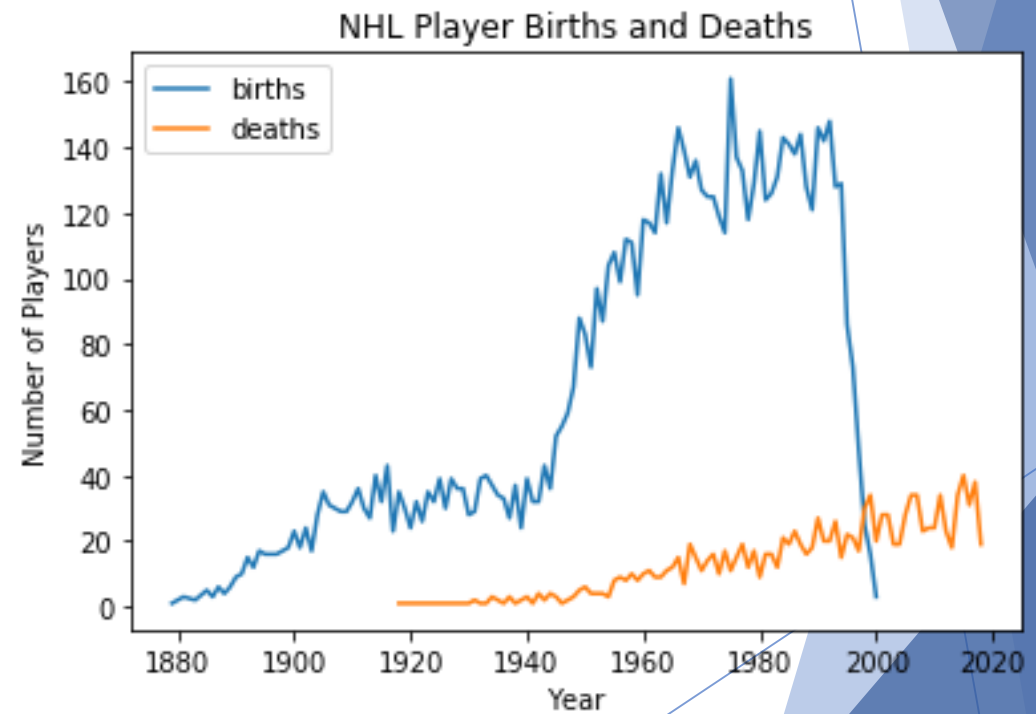
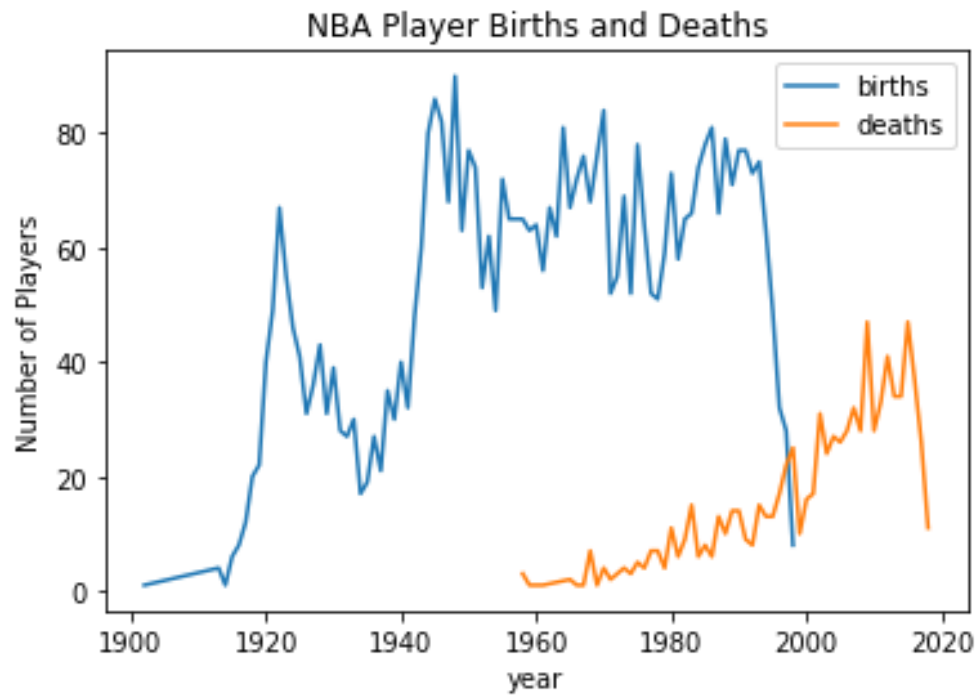
- ▶ Life expectancy tables from 2004 to 2015 were collected from the Social Security Administration
- ▶ Number of annual live births in the United States were taken from the Center of Disease Control and Prevention



# Annual Number of Births and Deaths of Professional Athletes



# Annual Number of Births and Deaths of Professional Athletes

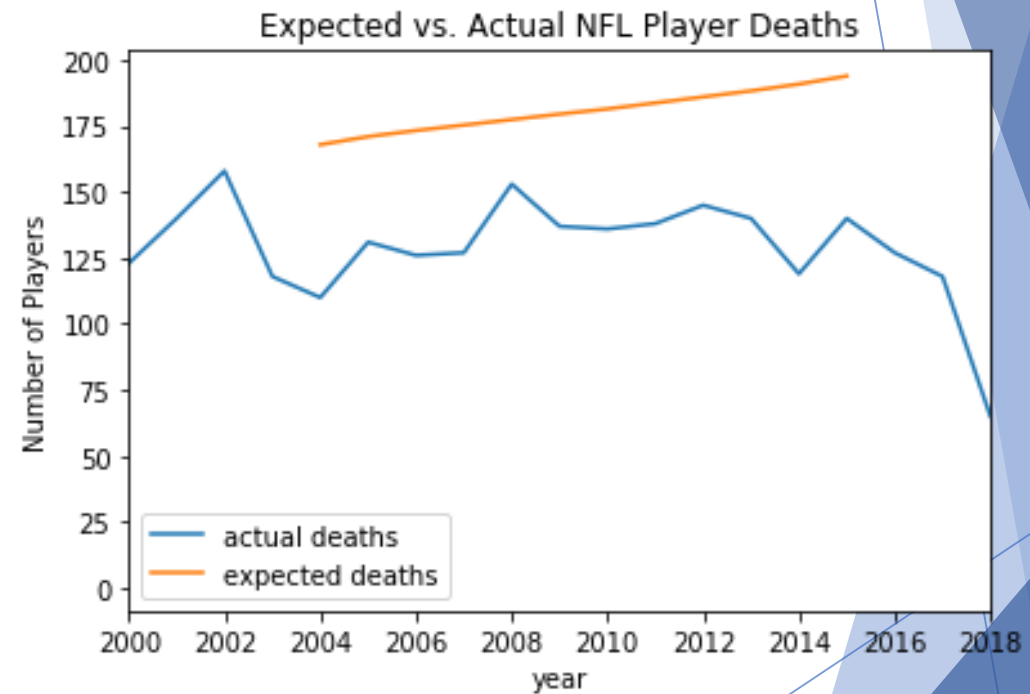
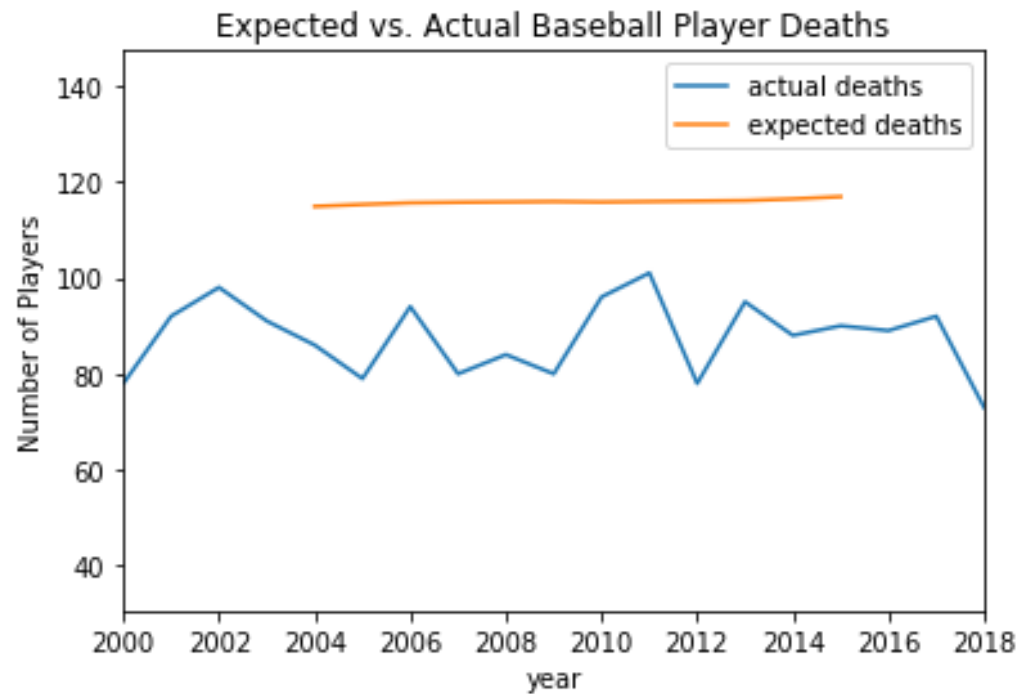


# Modeling Number of Athlete Deaths in a Particular Year

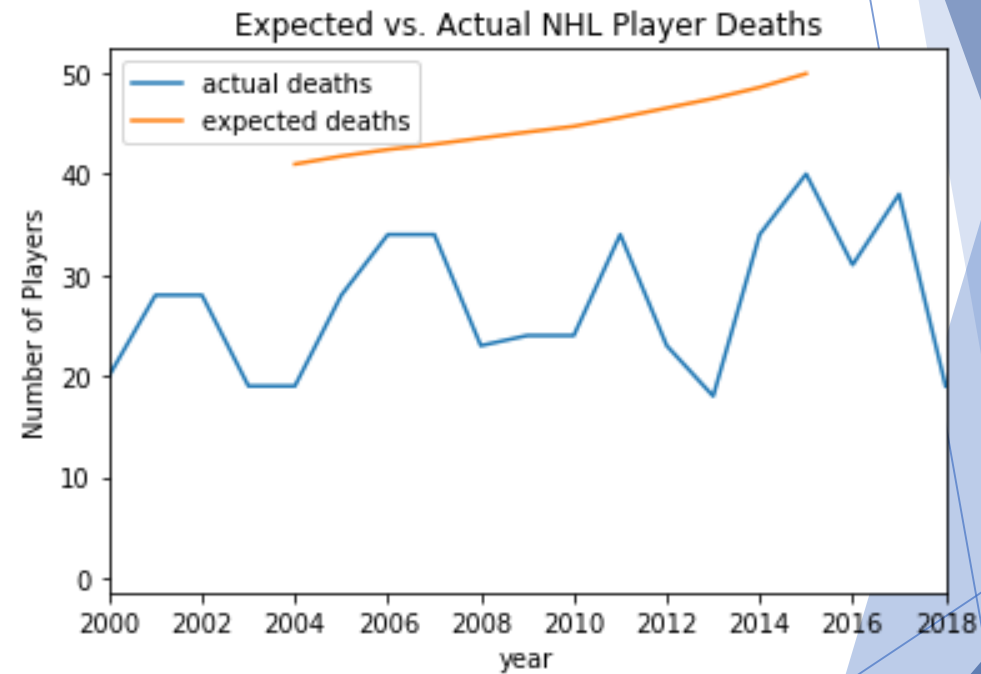
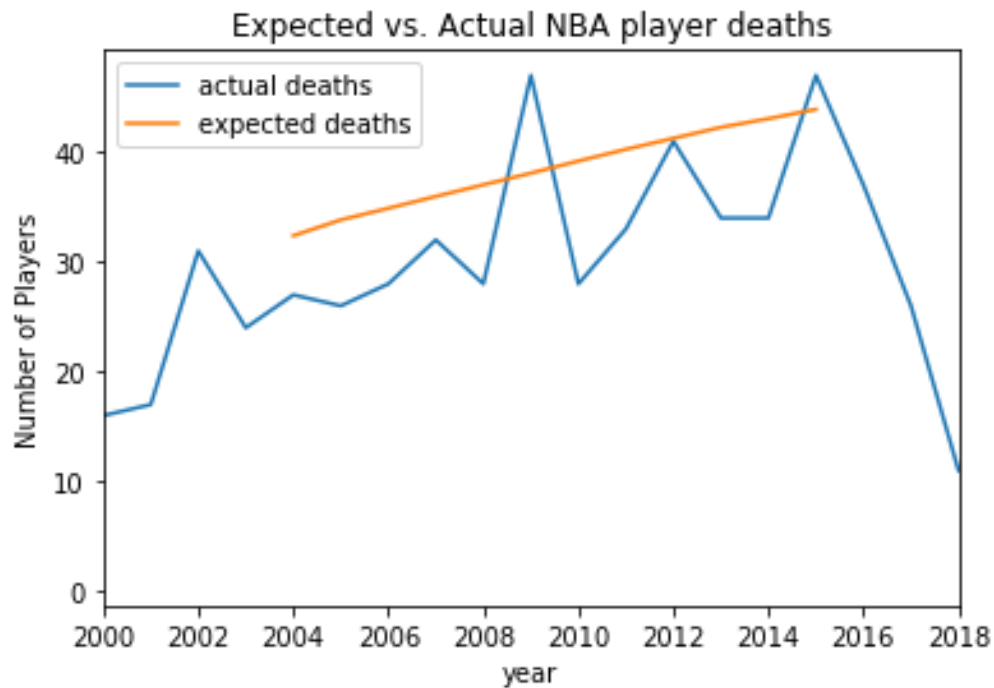
- ▶ Example: how many NBA players will die in 2015?
- ▶ For each year in the 2015 SSA Life Table, estimate percentage of males who were born that year that will die in 2015
  - ▶ Example: roughly 3.5% of males born in 1930 will die in 2015
- ▶ Multiply that estimated percentage by the number of athletes who were born in that year to get expected number of athletes who will die at that age
  - ▶ Of the 39 players born in 1930, about  $0.035 * 39 = 1.365$  will die in 2015
- ▶ Sum up the expected number of deaths for each birth year to get total expected number of athlete deaths in 2015
  - ▶ For NBA players 2015, expected number is 43.9 deaths. In actuality, 47 players died.



# Predicted vs. Actual Number of Deaths



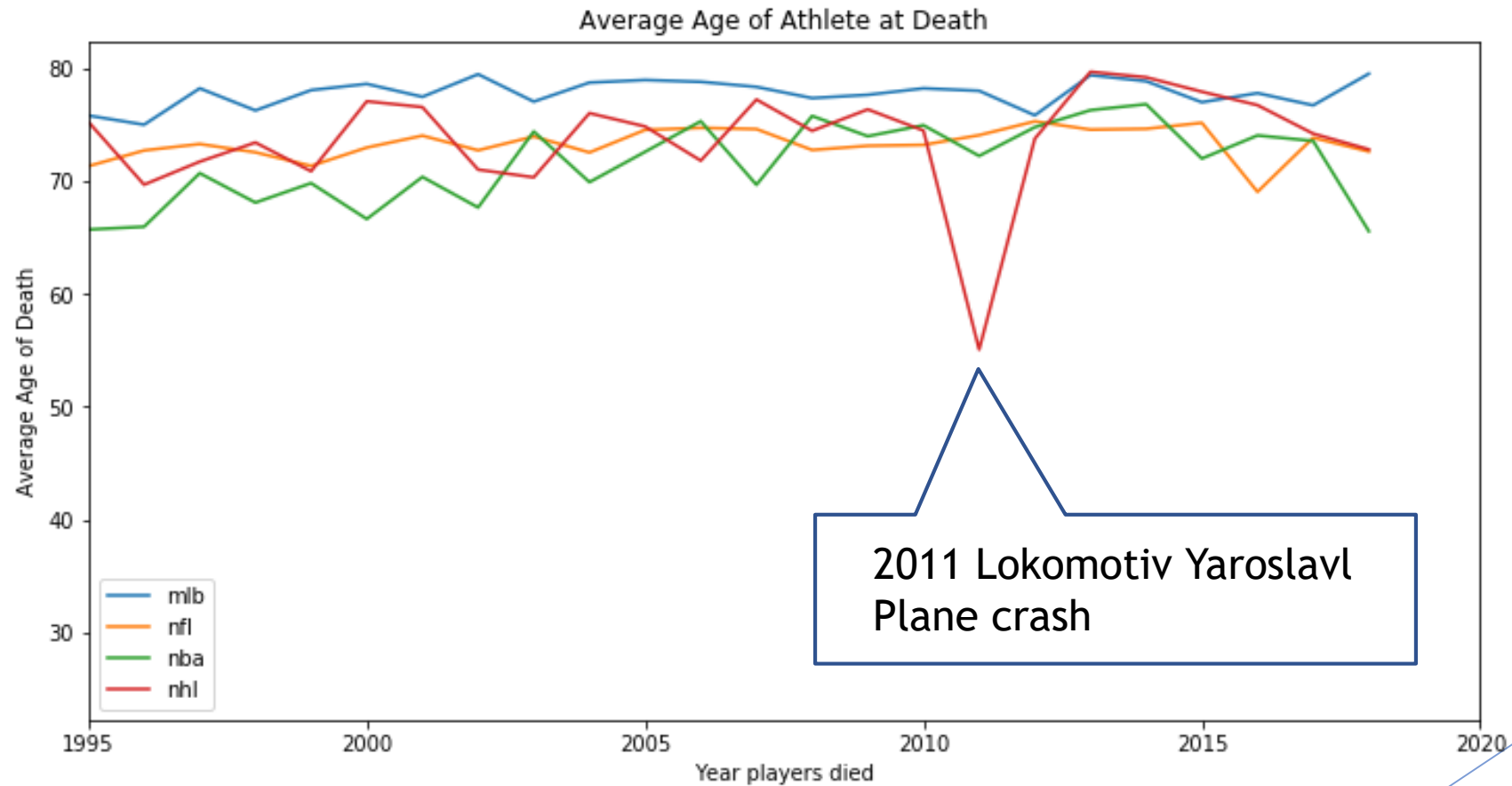
# Predicted vs. Actual Number of Deaths



# Comparison of Distributions of Ages at Death

- ▶ Conduct an A/B test between two population samples to determine how likely their ages at death are from same distribution
  - ▶ First group: all athletes of a particular sport who died in a particular year
  - ▶ Second group: general population of American males who died in a particular year
- ▶ For NBA, NHL, and NFL athletes, A/B test did not reveal any conclusive differences between the athletes' ages at death with those of the general population
  - ▶ One exception: NHL player deaths in 2011
- ▶ However, for Major League Baseball players, the A/B test revealed a difference between their age distribution and that of the general population

# Average Age at Death of Pro Athletes



# Attempt at Classification

- ▶ Of all Major League Baseball players who were born between 1930 and 1945, 42% are deceased
- ▶ Can we build a classification model to predict whether a player is still alive?
- ▶ Based on length of career, numbers of games played, listed height and weight, no classification model achieved over 60% accuracy
- ▶ As it turns out, no correlation between any of those characteristics with whether a player is still alive, or how long a player lived
  - ▶ At best, a slight (and obvious) correlation exists with the year of birth and how long a player lived

# Conclusions

- ▶ Prediction of number of athlete deaths works reasonably well for NBA players, but consistently overestimates for the other sports (MLB, NFL, NHL).
- ▶ Distribution of ages at death is not significantly different from the general American male population for NFL, NBA, or NHL players
- ▶ However, Major League Baseball players seem to have higher life expectancies than players in other leagues
- ▶ No correlations exist between life expectancy and length of playing career, or even physical dimensions of professional athletes

Life expectancy of NFL players is 59 years?

