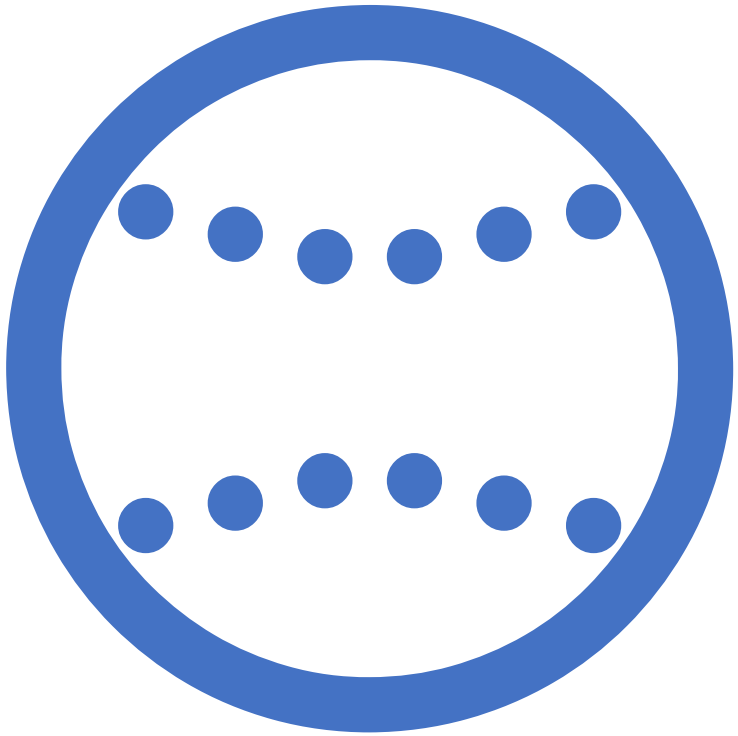
Pro Baseball Focus Consultancy

~ The Best Pitch, Every time

1. Tyler Gigot
2. Cindy Kim
3. Gavin Master

IST 687

AGENDA



01

Introduction

- Briefing the audience about the problem that is faced and introducing the dataset that is used to counter the problem.

02

Business Questions

- Framing the questions that will help to understand the problem in a better way and eventually solve it.

03

Data Preparation

- Provide some background on the underlying dataset and explain how it was cleaned.

04

Exploration

- Take a preliminary look at some of variables and their relationships leading into the modeling

05

Modeling Techniques

- Explore the data and implement various Models to interpret the parameters on which our business questions rely upon.

06

Interpretations

- Deriving actionable insights from the models used.

Introduction

Introduction

Objective

We as a consultancy firm focus on solving a business problem for Major League Baseball (MLB) Team Managers. Utilizing pitching data, we can create comparison sets to check the parameters where pitchers begin to decline during their careers. The main goal is to identify parameters that will aid in predetermining when a pitcher will begin to decline, therefore signaling your team to begin your search for a new pitcher.

Definitions

- W → Wins
- L → Losses
- IP → Innings Pitched
- ER → Earned Runs
- BFP → Batters Facing Pitcher
- SO → Strikeouts
- BB → Base on Balls (Walk)
- HBP → Hit-By-Pitch
- HR → Home Runs
- H → Base Hit
- R → Run
- ERA → Earned Run Average

Business Questions

Business Questions

Does pitching performance tend to decline with age?

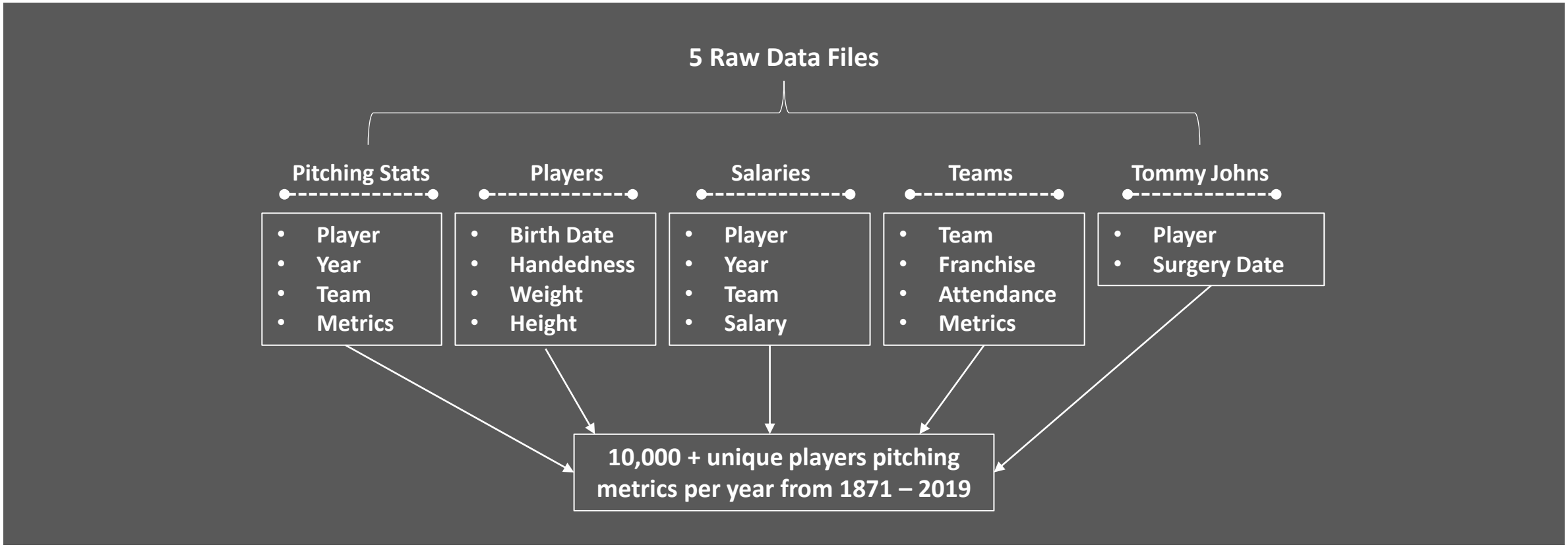
Can pitching metrics be accurately predicted statistically?

Does Tommy John surgery have an influence on pitching?

What does strikeouts show you in relation to ERA?

Data Preparation

Overview of Our MLB Pitching Data



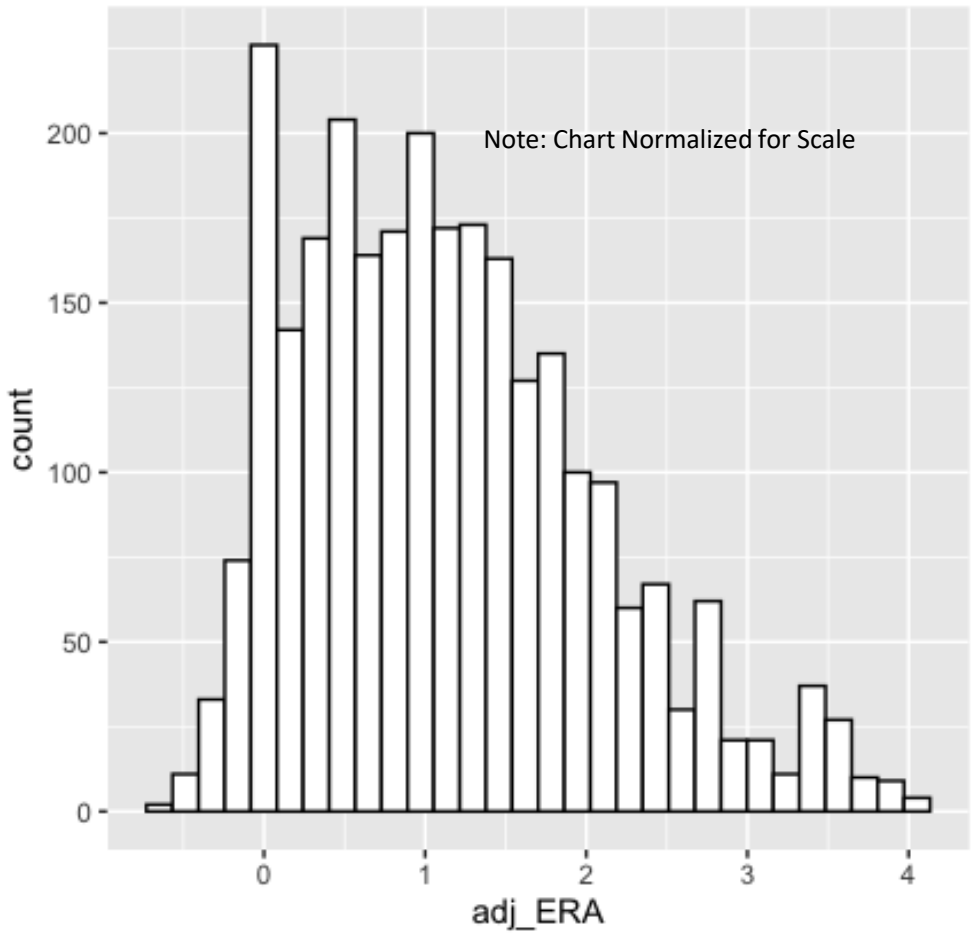
Credit to Sean Lahman for the publically available database used in the makings of this project
www.seanlahman.com/baseball-archive/statistics/

Cleaning The Data

Issue	Resolution
The number of teams changed frequently frequently in 20th century	Use only the years of 2005 – 2019 where it stayed at 30 teams
Florida Marlins changed to Miami Marlins in 2012.	Changed the teams of Florida to Miami to be consistent
A lot of records had a trivial amount of innings pitched and good ERA's	Deleted all records with < 100 innings pitched in a season
The salary data was not complete	For NA's, used the average of that specific players salaries, or if they had none then the average of all players

Exploration

Distribution of ERA for All Players



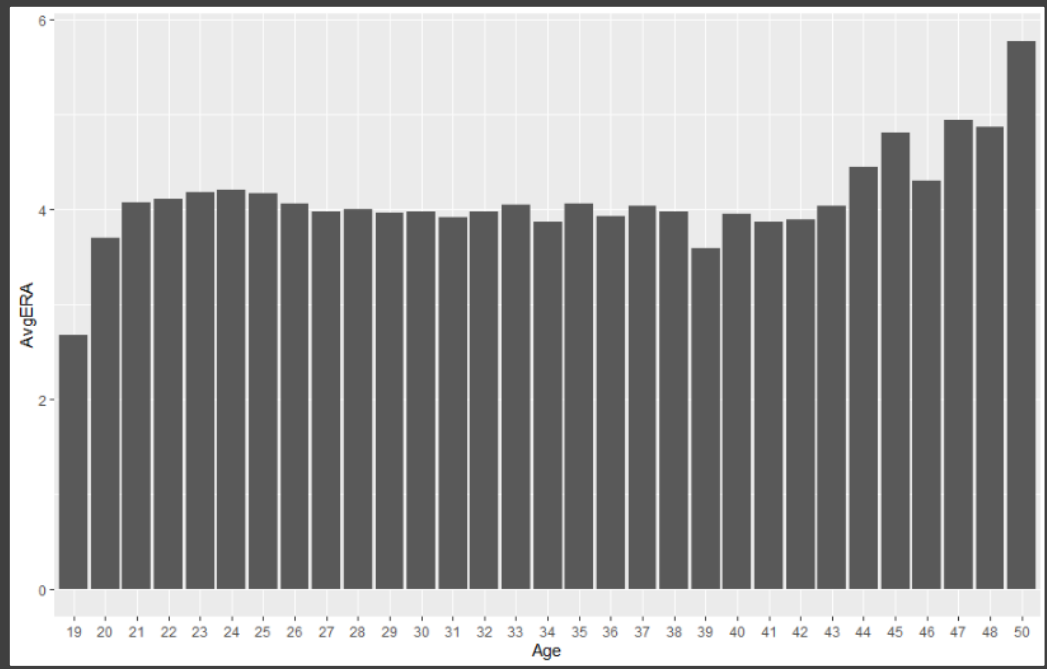
What is ERA?

- The lower a pitcher's ERA, the better
- It becomes more difficult to maintain a low ERA as more innings are pitched
- The pitcher with the best ERA is Clayton Kershaw at 2.445
- The pitcher with the worst ERA is Jose Reyes at 54
- Pitchers with an ERA of 5.549 are considered average

Age and Average ERA

- Our data started to show that there was some sort of relationship between age and ERA
- Higher level review of the data suggested that your age has a direct correlation to your ERA
- However, why do we see ERA declining around the late 30's and early 40's?

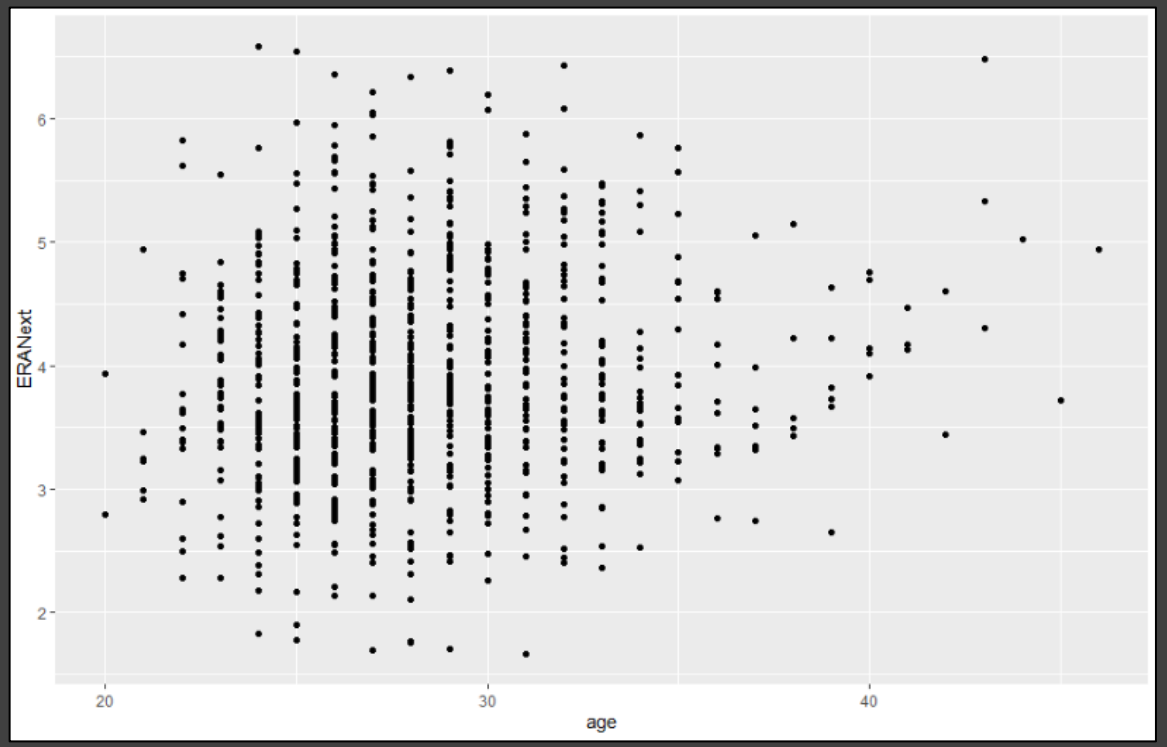
Average ERA vs Age



Correlation Between Age and Following Year ERA

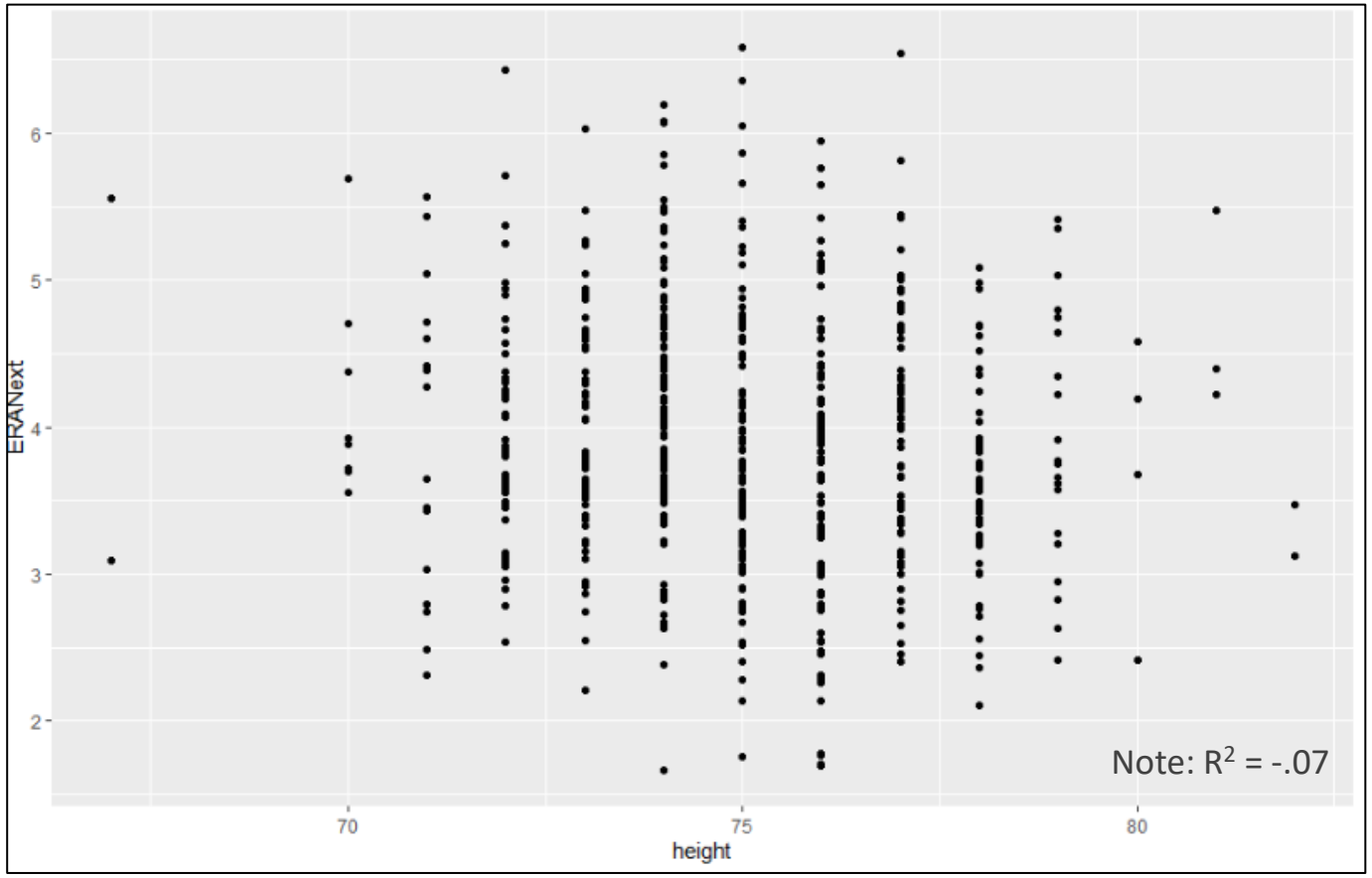
- Our data suggested a relationship between Age and ERA
- The relationship between Age and ERA result in $R^2 = .10$
- So, if Age is a factor in a player's ERA, one would think height and weight would also play an important role

Age vs Following Year ERA



Height

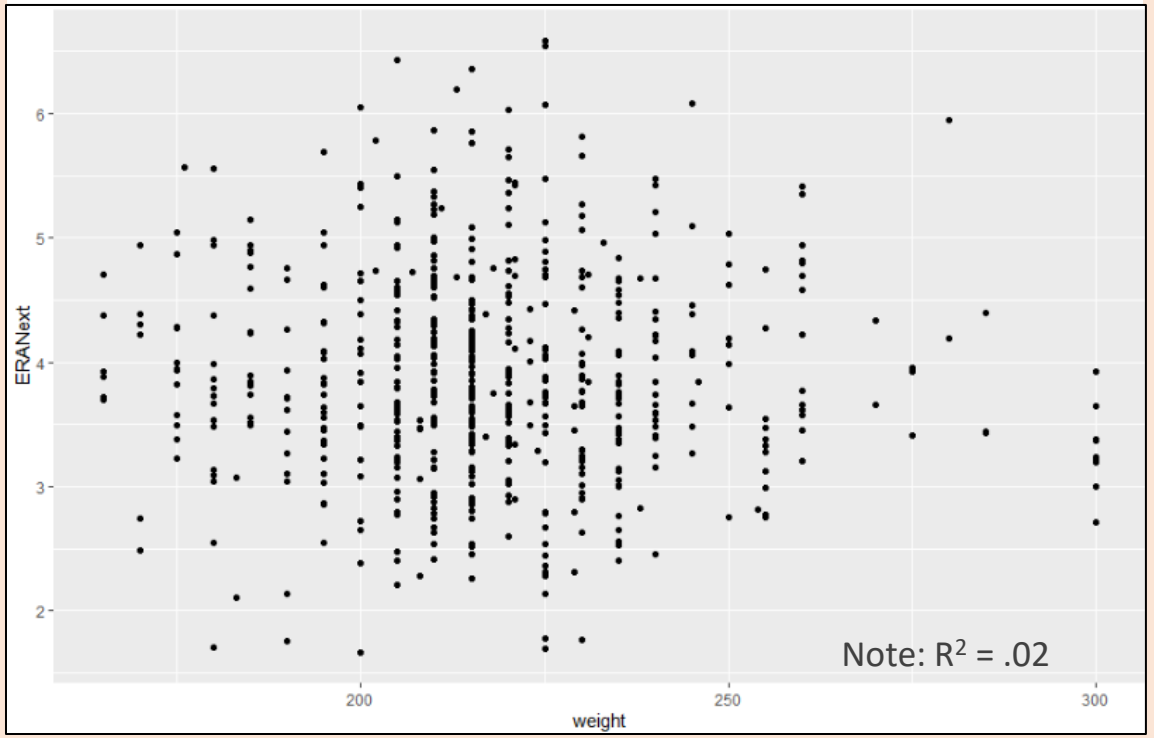
Height vs Following Year ERA

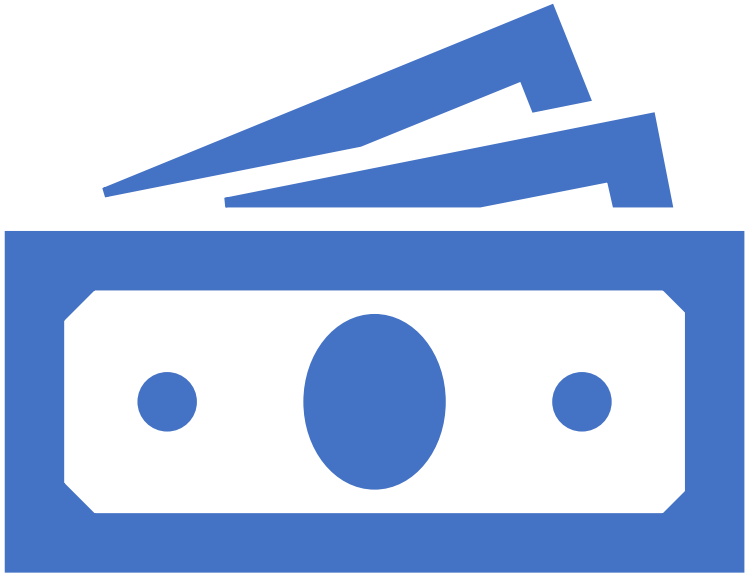


- Our data shows a correlation between age, but not height and ERA
- What about weight?

Correlation Between Weight and Following Year ERA

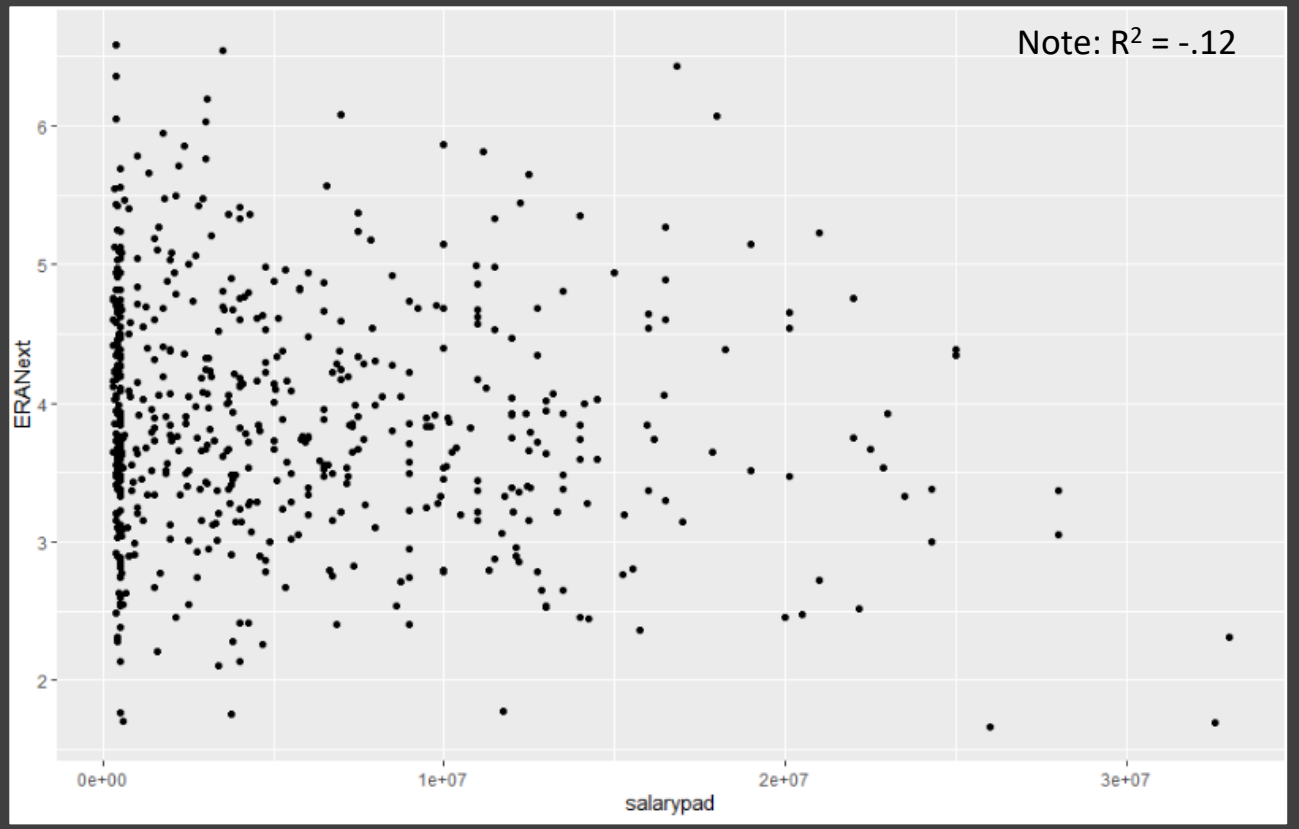
- No correlation between Weight and Next Years ERA
- Age but not Height and Weight
- Would salary drive a player's following year ERA?





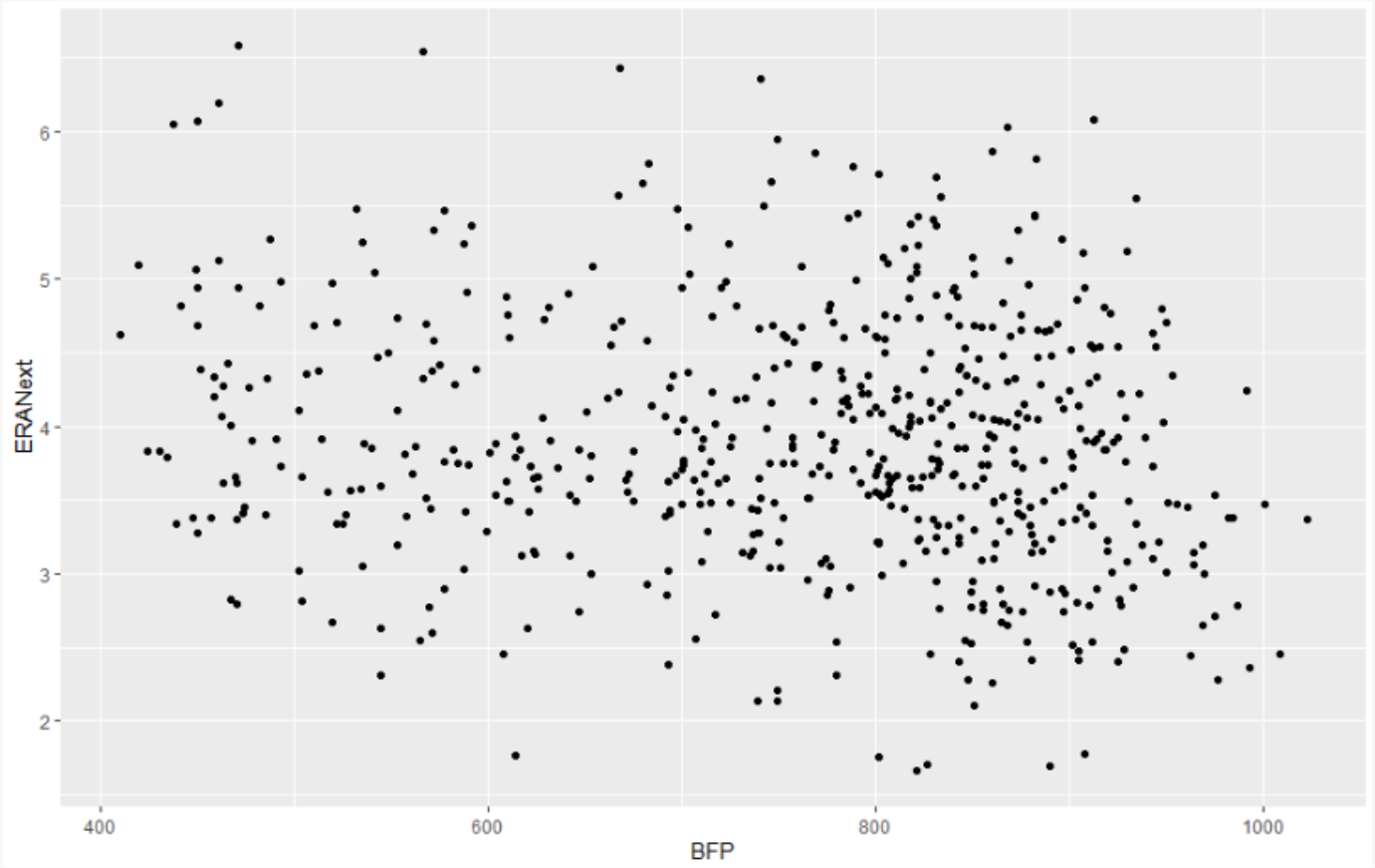
Correlation Between Salary and Following Year ERA

- Slight correlation between Salary and the Following Year's ERA
- Insufficient data to suggest that if you pay your pitcher more, their ERA will be better the following year
- Sufficient data that points to teams paying more for pitchers who have a better ERA



Correlation Between Batters Faced by Pitcher and the Following Year ERA

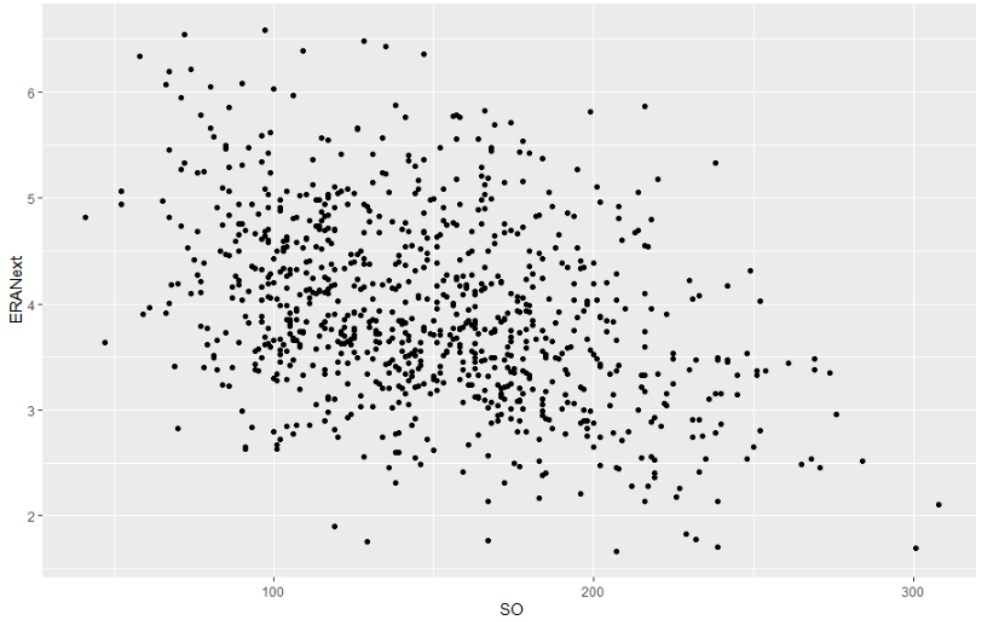
- Slight correlation between BFP and the following year ERA
- No personal attributes minus age factor into a player's ERA
- BFP and Salary
- What about Strikeouts?



Note: R^2 = -.14

Correlation Between Strikeouts and the Following Year ERA

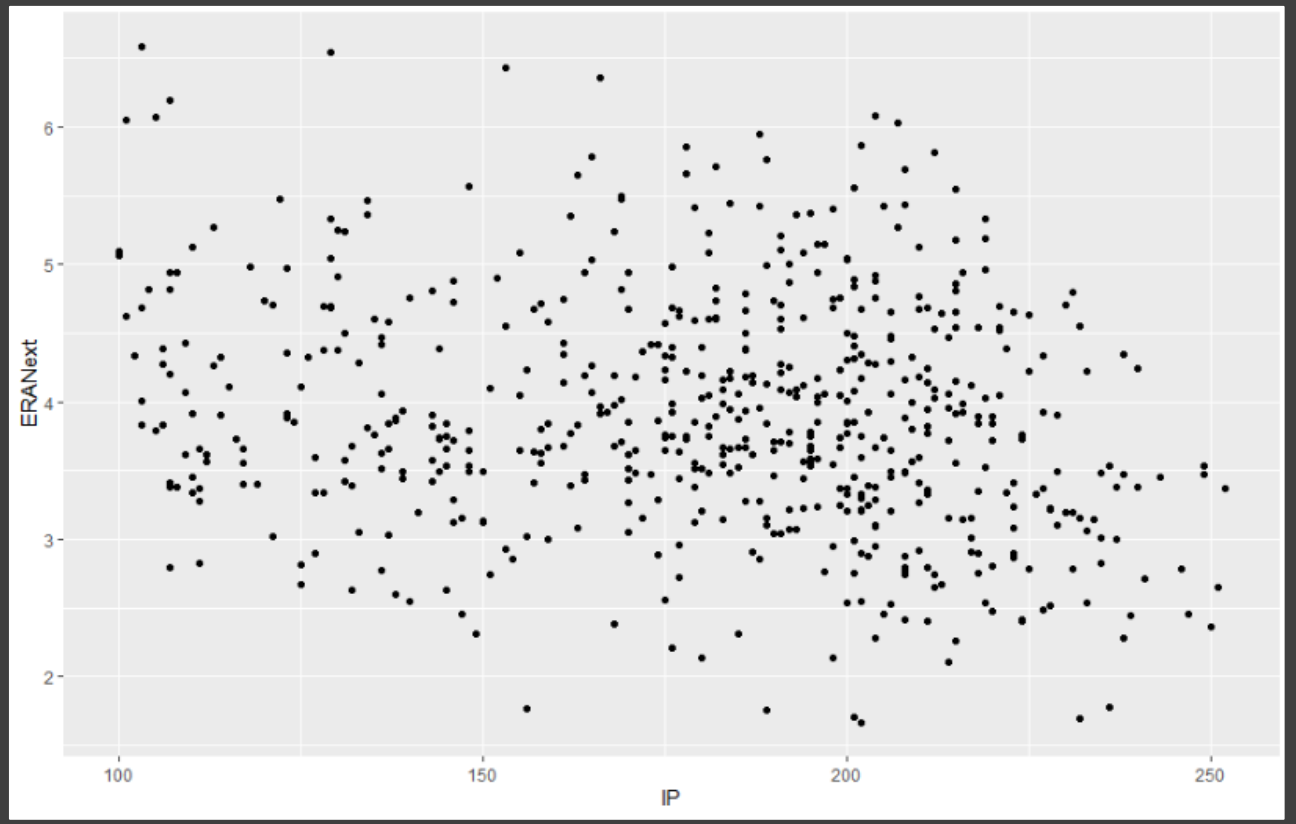
- Very strong correlation between strikeouts and a pitcher's ERA
- **Nolan Ryan** has the most career strikeouts in Major League Baseball. During a record 27-year career, he struck out 5,714 batters.
- A Pitcher gets their strikeout percentage by using the following calculation:
  K% =*Strikeouts / Batters Faced*
  27% is considered Excellent



Note: R^2 = -.38

Correlation Between Innings Pitched and the Following Year ERA

- There is a slight relationship
- More innings = Better ERA



Note: R^2 = -.19

Tommy John Surgery

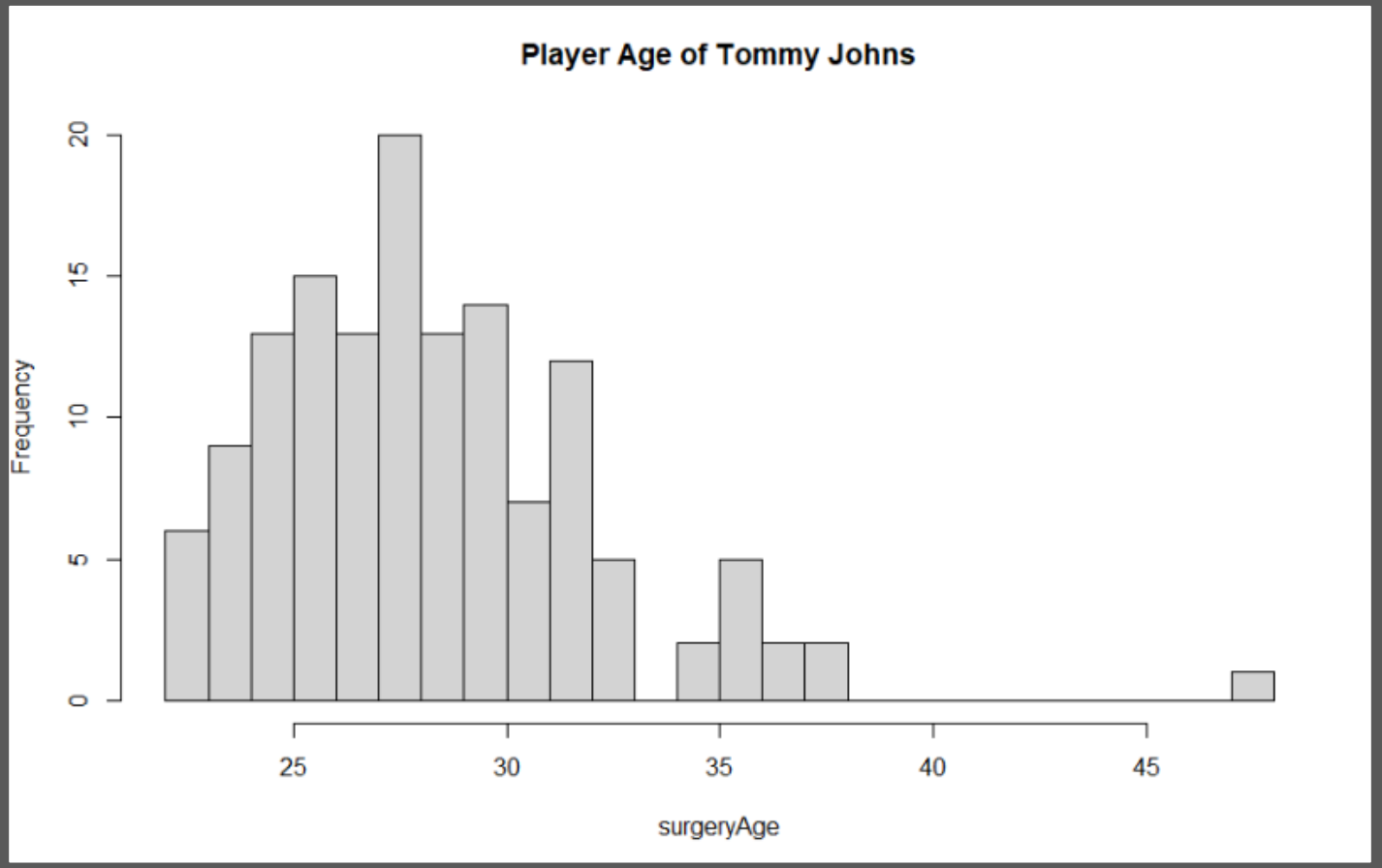
What is Tommy John Surgery?

a surgery commonly used to repair a torn ulnar collateral ligament inside the elbow by replacing it with a tendon from elsewhere in the body. The goal of the surgery is to stabilize the elbow, reduce or eliminate pain and restore stability and range of motion.

Analyzed the effect of Tommy Johns surgery on a pitcher's performance

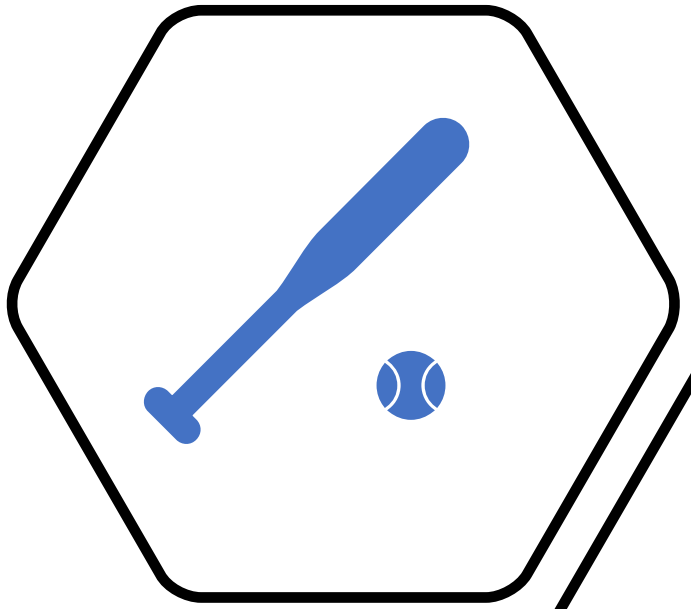
139 pitchers who pitched at least 1 season both before and after the surgery

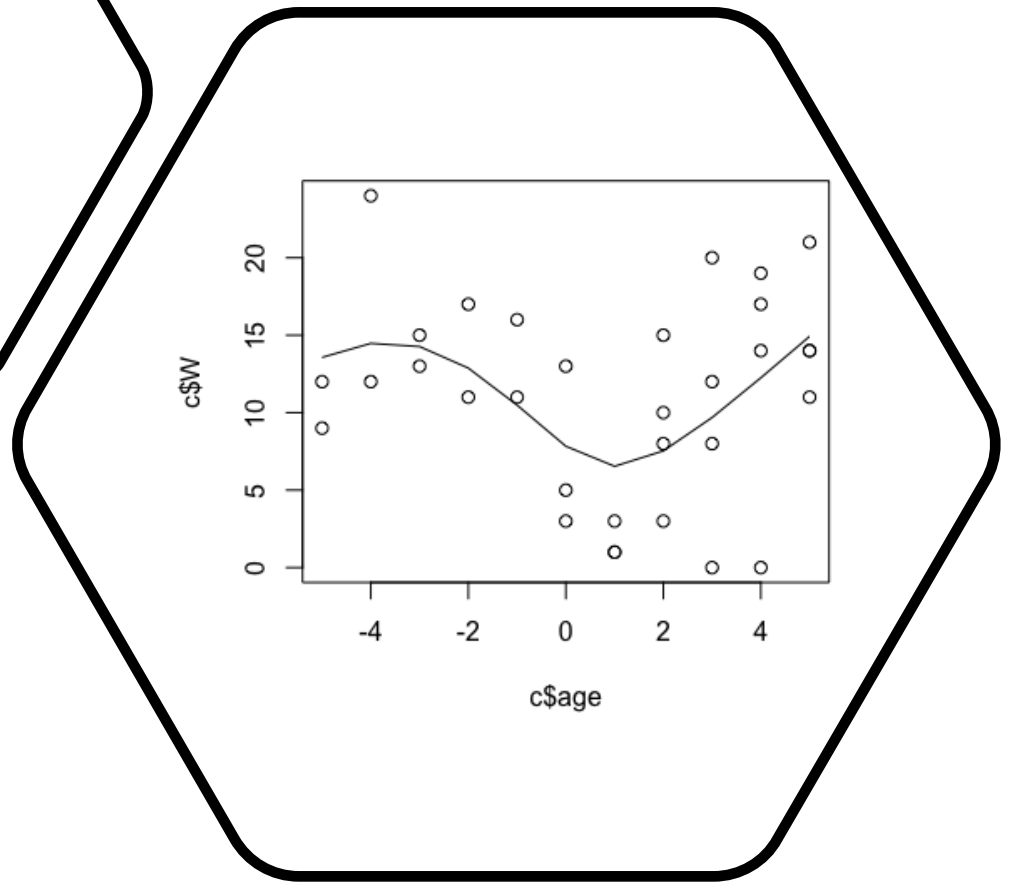
Average Age of Tommy Johns



Average Age of Tommy Johns	28
Average ERA's before Tommy Johns	4.1
Average ERA's after Tommy Johns	4.4

Players React to Tommy John Surgery





- Zero is when the player had surgery
- This graph shows the plot and trend line of wins for a pitcher before and after Tommy John Surgery
- This will give you an idea if a pitcher is a right fit for an organization if they have received Tommy john surgery

Correlation Between Handedness and ERA

Left-Handed Average ERA

3.79

Left-Handed Average Salary

$ 3.32 Million

Right-Handed Average ERA

3.93

Right-Handed Average Salary

$ 3.08 Million

Who has the better ERA?

Left-Handed Pitchers

Modeling Techniques

Rules

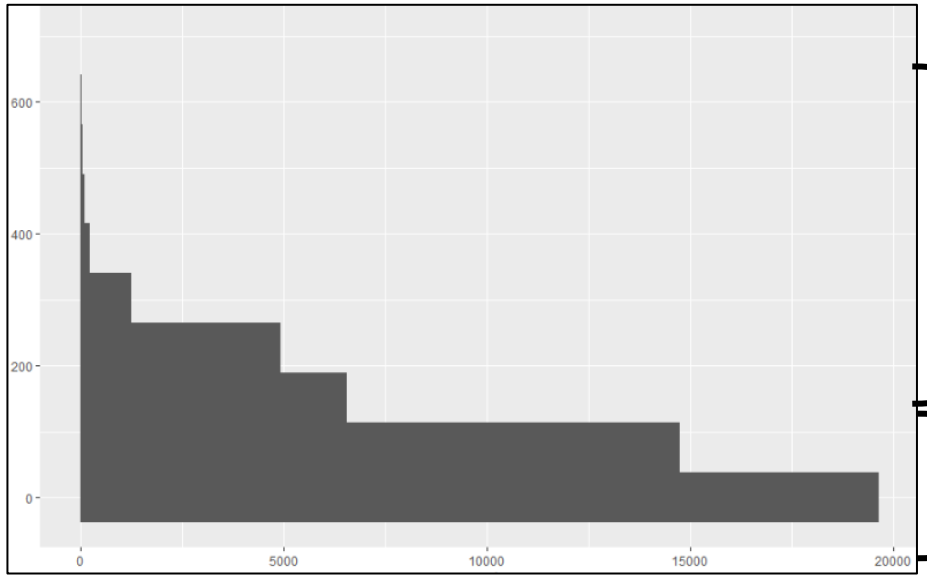
The data used for these models contains all data for starting pitchers (those who have pitched over 100 innings and pitched for at least two consecutive seasons) between 2005 to 2019

All teams in the dataset are as of the last MLB name change which occurred in 2005

Various attributes are added in to test and validate our modeling techniques and to see their effects on a pitcher’s ERA

Starter vs Relief Threshold Innings Pitched > 100

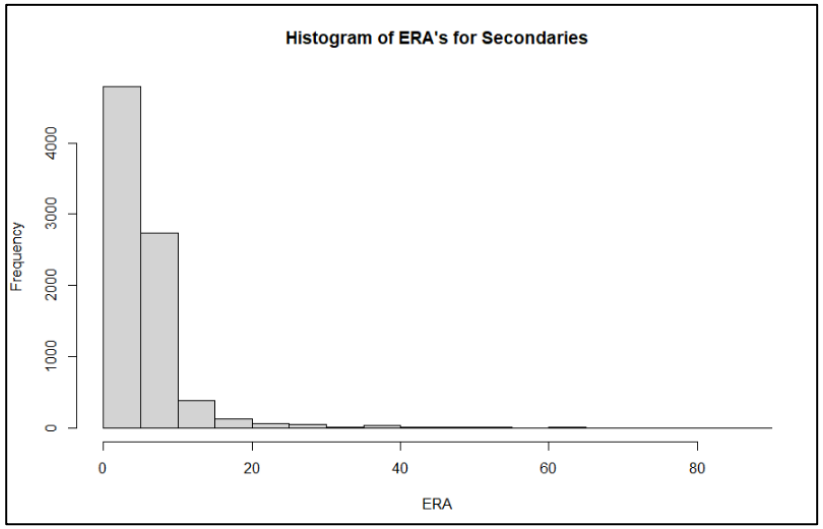
Histogram of Innings Pitched for all Pitchers



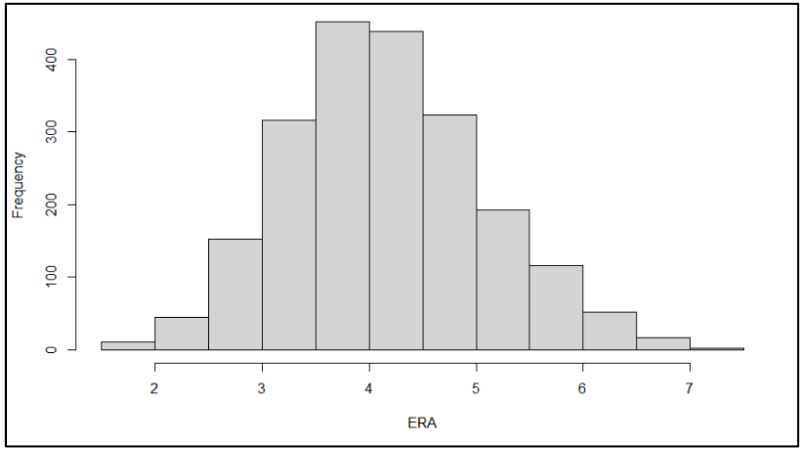
Innings Pitched

Frequency

Histogram of ERA's for Secondaries



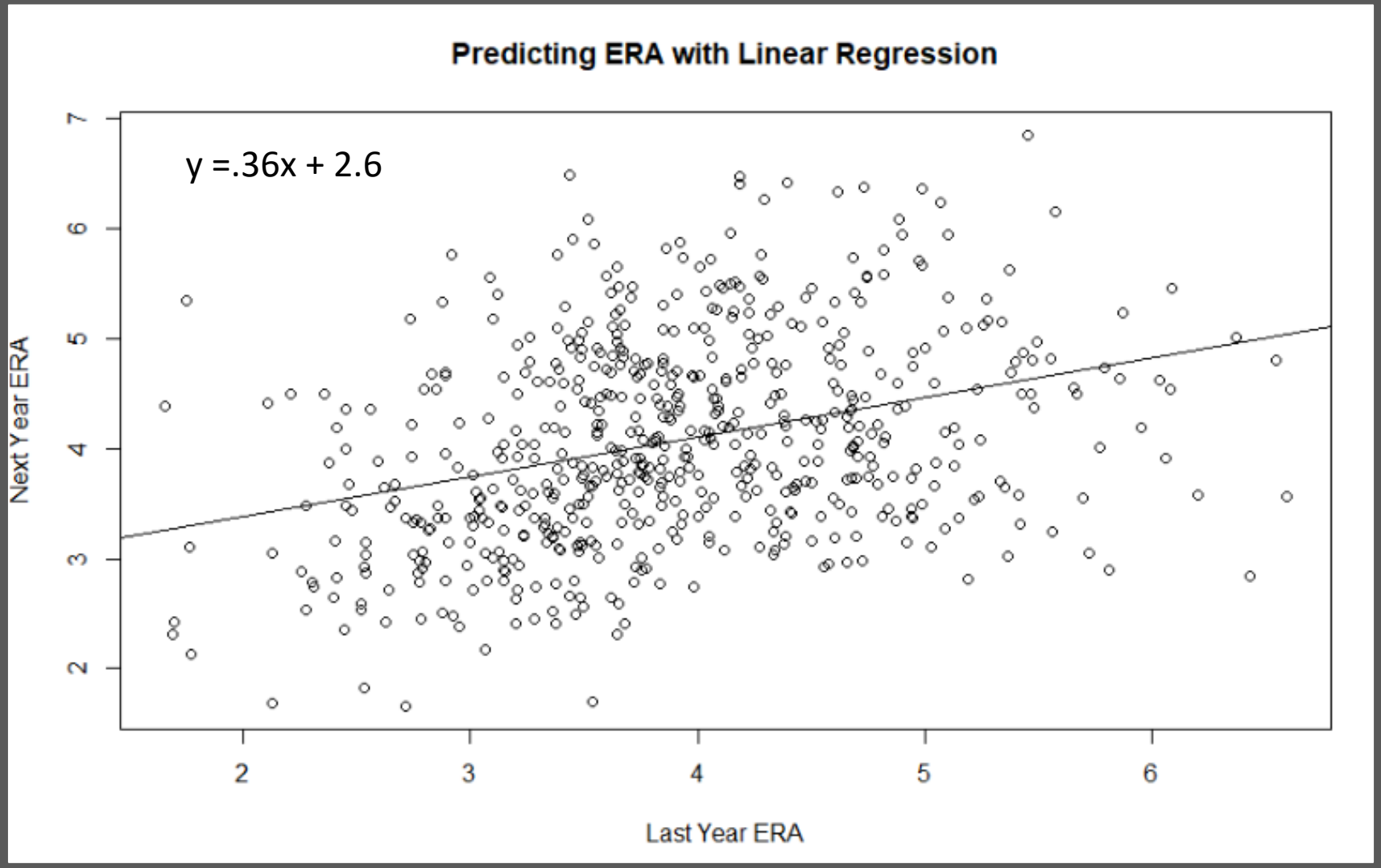
Histogram of ERA's for Starting pitchers



Model #1 : Linear Regression

Predicting Future ERA Utilizing Linear Regression

83% Accurate



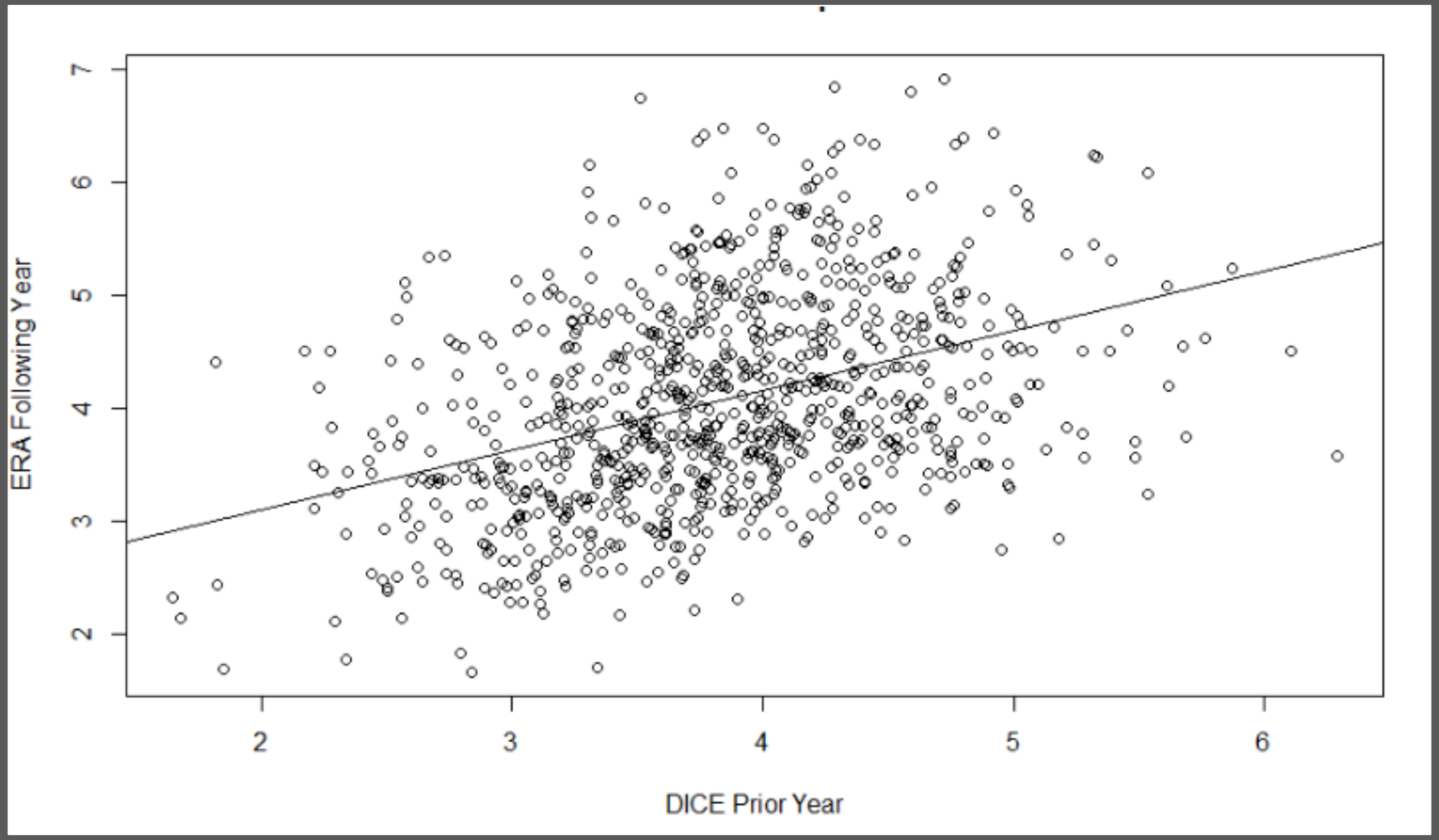
Model #2 : DICE Method

DICE = Defense Independent Component ERA

$$\text{DICE} = 3.00 + \frac{13HR + 3(BB + HBP) - 2K}{IP}$$

Predicting Future ERA Utilizing Last Year's DICE

81% Accurate

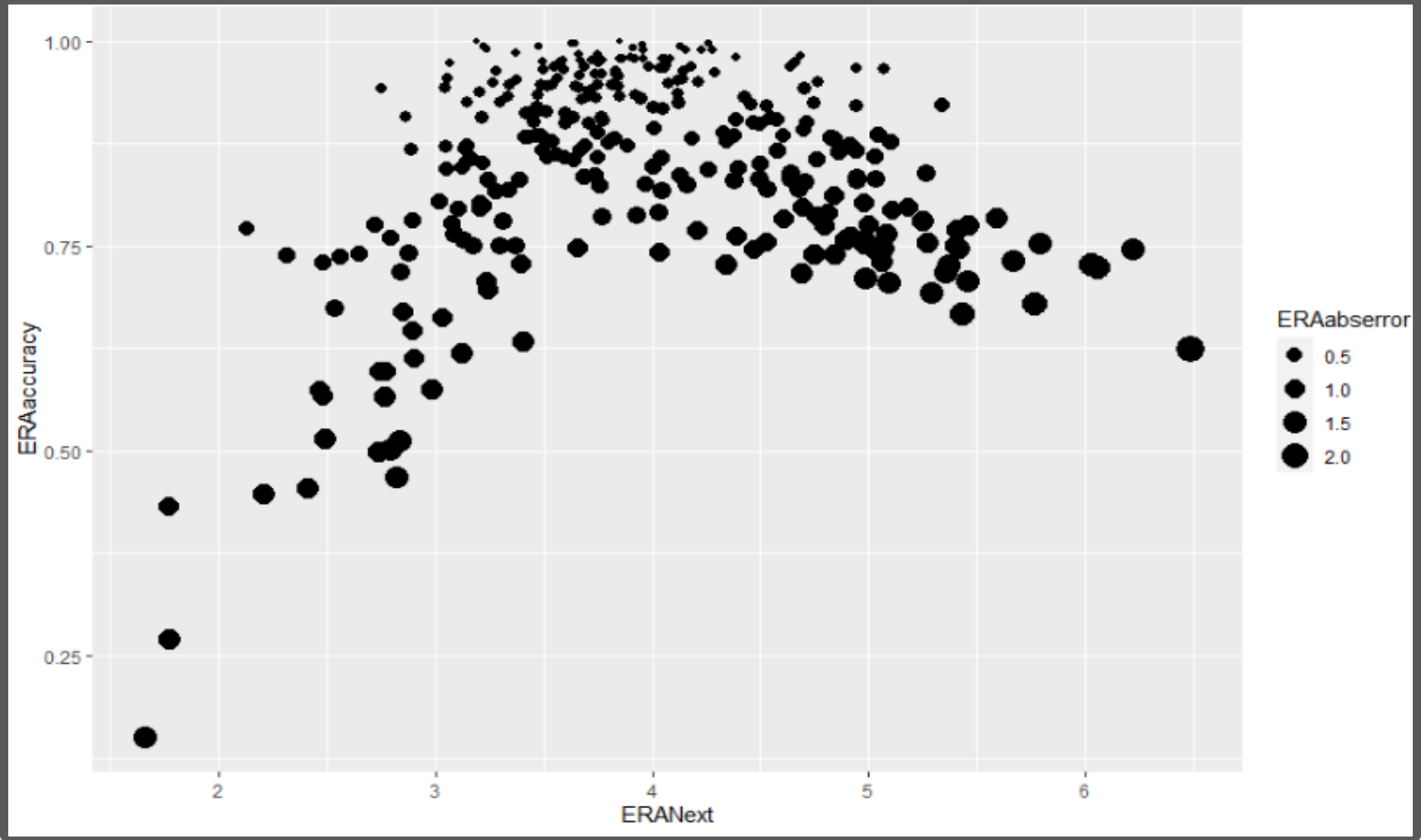


Model #3 : Support Vector Machine

Additional variables include walks, hit by pitch, strike outs, and home runs

Predicting Future ERA Utilizing SVM

85% Accurate

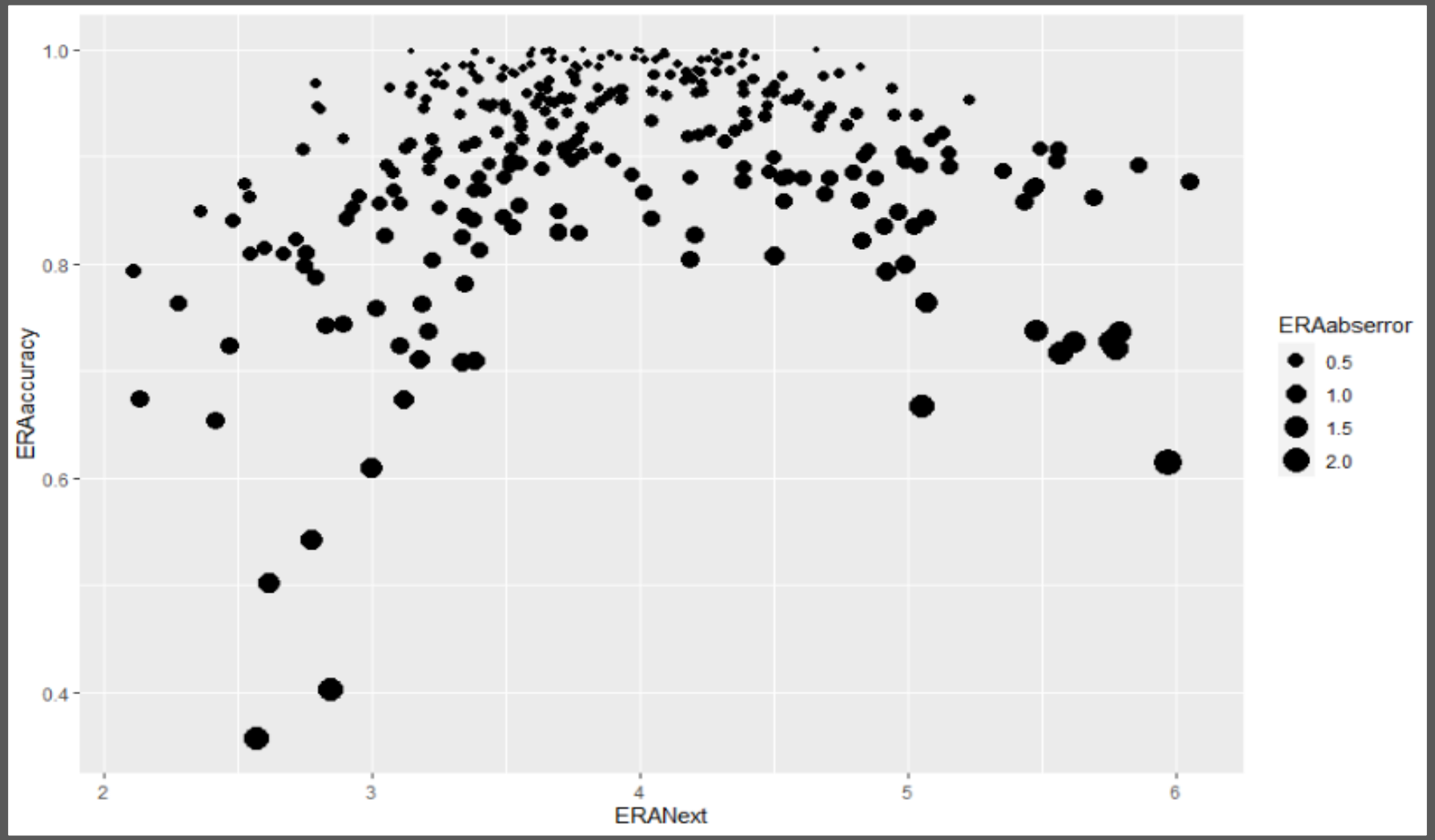


Model #4 : Random Forest

Additional variables include height, weight, handedness, salary, and age

Predicting Future ERA Utilizing randomForest

90% Accurate



Interpretations

Interpretations

Age has an influence on a pitcher's ERA. We recommend searching for a new pitcher when their age reaches the mid-thirties.

Pitchers who have elected to receive Tommy John Surgery can continue to perform strongly. You should not rule out a pitcher who have received this surgery, although it can be a sign that their career may begin to slow down and does present some risk so proceed with caution.

Utilizing our firm's extensive modeling techniques, we accurately predict your pitcher's ERA for next year by almost 85%.

Lefties are slightly better than righties in terms of their stats. They also tend to get paid more by about 200K-300K on average.

Key take away: Focus on a player's age, current ERA, surgery history, and strikeout percentage. This data will provide you the picture your team needs heading into the next MLB Draft or Free Agency.