Final Report for MA678 Project

Daniel(Chen) Xu BUID:U49903384

12/10/2021

Abstract:

The National Basketball Association (NBA) is a professional basketball league in North America. The league is composed of 30 teams (29 in the United States and 1 in Canada) and is one of the four major professional sports leagues in the United States and Canada. It is the premier men's professional basketball league in the world. The league was founded in New York City on June 6, 1946, as the Basketball Association of America (BAA). It changed its name to the National Basketball Association on August 3, 1949, after merging with the competing National Basketball League (NBL). The NBA regular season runs from October to April, with each team playing 82 games. As the premier professional basketball league in the world, the change in style of basketball in the NBA led to the change in style of modern basketball. Here rises problems: Which technical statistics best captures the changing style of modern basketball and how these statistics determine the injury situation for players of each season? To address those problems. I use Exploratory Data Analysis to find the factor(Average number of 3 points attempt per game) related to the change of play style and use some statistics to build a multilevel model. However, the model shows that the variables all have a slight impact on the injury situation for each player and is slightly different between Seasons as the style of basketball changes as time passes. This report consisted of 5 main parts: Introduction, Method, Result and Discussion.

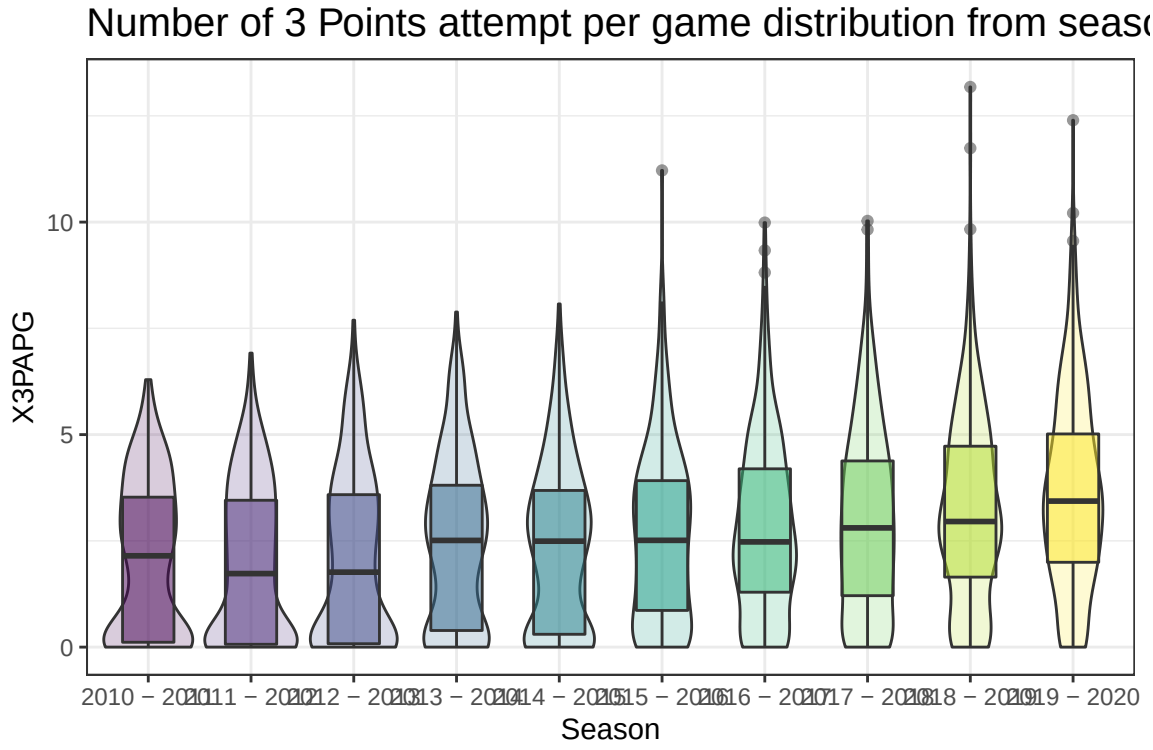
Introduction:

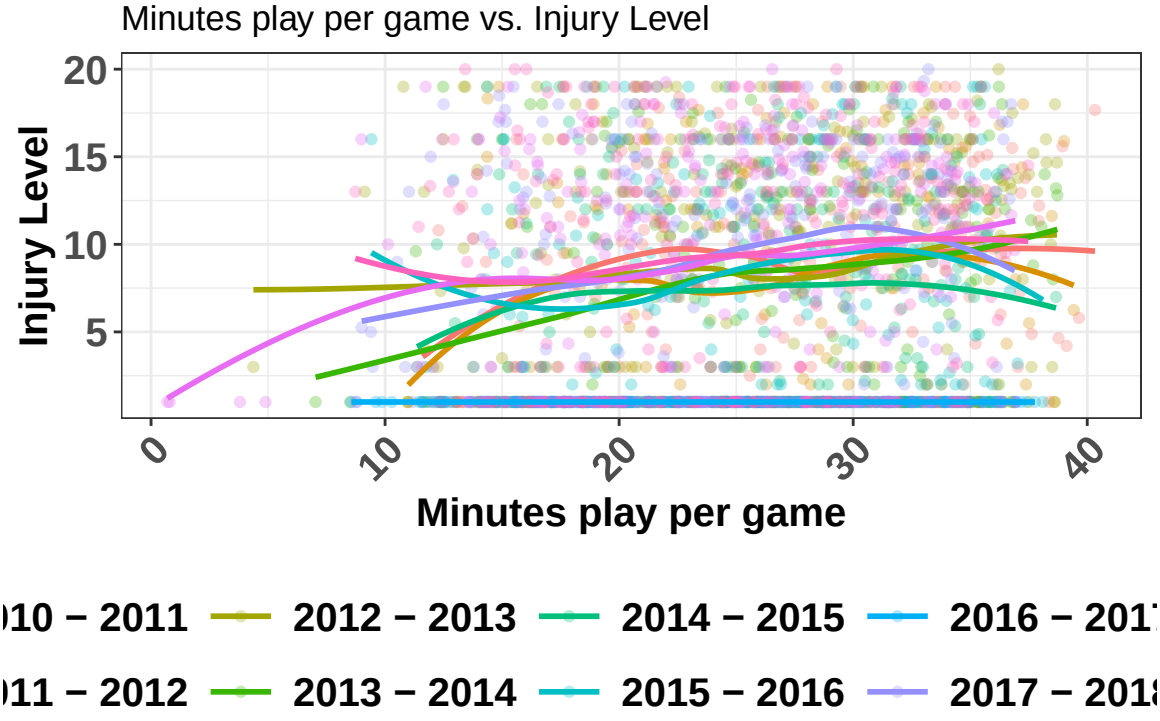
Method

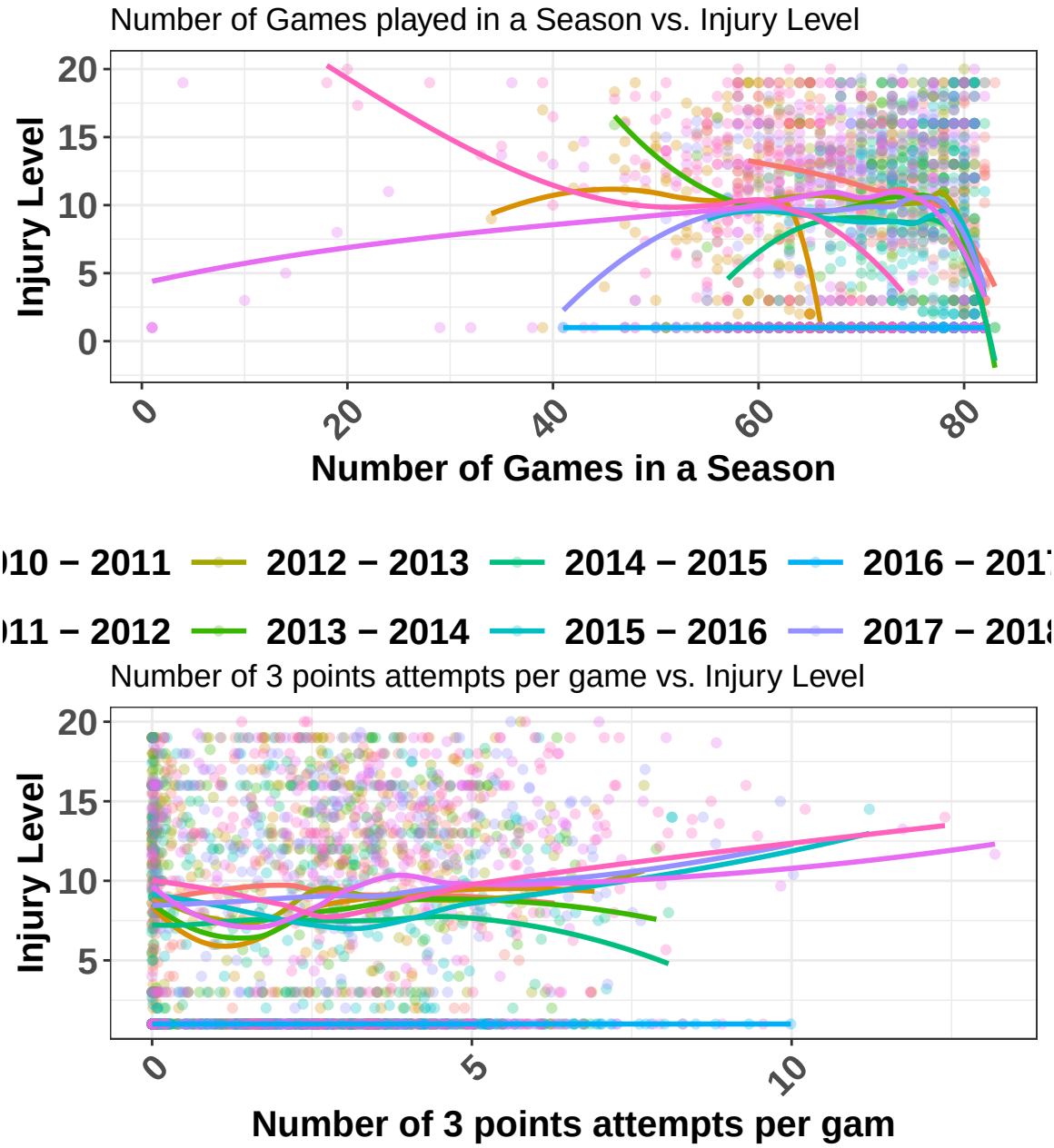
Data Cleaning and Processing:

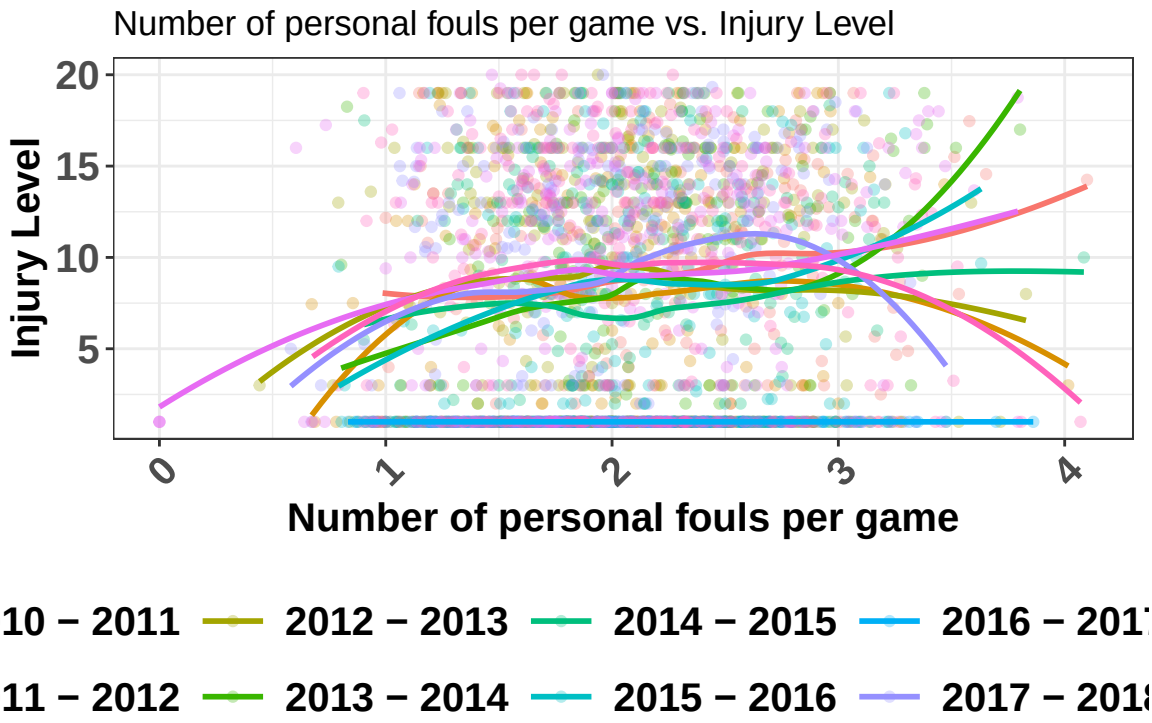
Column Names	Explanation
Season	NBA regular season year range
Age	Player's Age
PER	Player Efficiency Rating
GP	The number of games a player play in a season
MPG	Average minutes a player play per game
PFPG	Average number of personal fouls per game
X3PAPG	Average number of 3 points attempt per game
height_cm	Player's height in centimeter
weight_kg	Player's weight in kilogram
injury_level	A level summarizes the injury situation for each player of each season

Exploratory Data Analysis:






Number of Games played in a Season vs. Injury Level
Injury Level
20
15
10
5
0
0
20
40
60
80
Number of Games in a Season
)10 − 2011
2012 − 2013
2014 − 2015
2016 − 201
)11 − 2012
2013 − 2014
2015 − 2016
2017 − 201
Number of 3 points attempts per game vs. Injury Level
Injury Level
20
15
10
5
0
5
10
Number of 3 points attempts per gam
)10 − 2011
2012 − 2013
2014 − 2015
2016 − 201
)11 − 2012
2013 − 2014
2015 − 2016
2017 − 201



Model Fitting:

Result:

Model build:

We can conclude that the formula:

$$InjuryLevel = 9.93 + 0.014 \times height_{cm} + 0.013 \times weight_{kg} - 0.0016 \times Age + 0.14 \times MPG + 0.041 \times X3PAPG - 0.16 \times GP + 0.38 \times PF$$

Indicator select: Season, MPG, GP, PFPG, X3PAPG, height_cm, weight_kg, Age, PER, PFPG

Model Validation:

Discussion:

Citation:

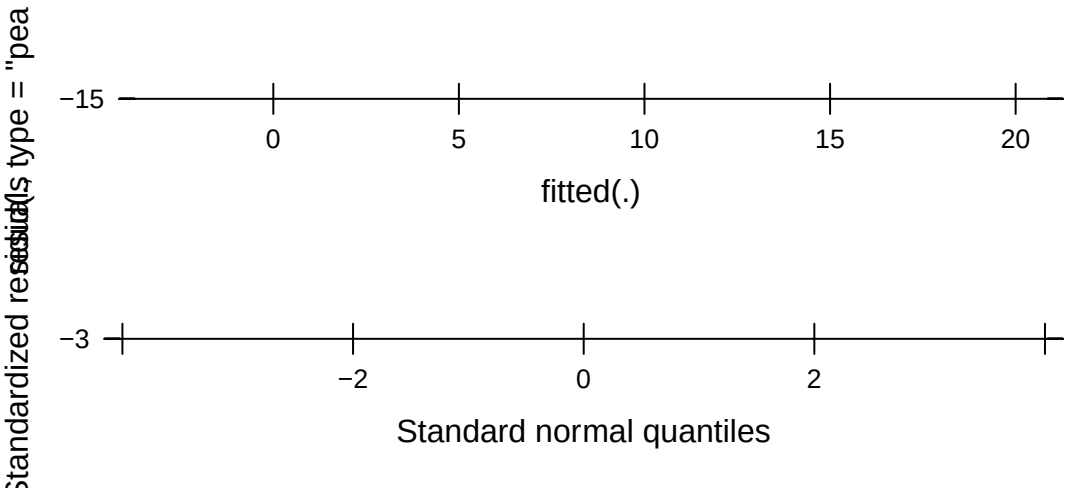


Figure 1: Residual plot and Q-Q plot.

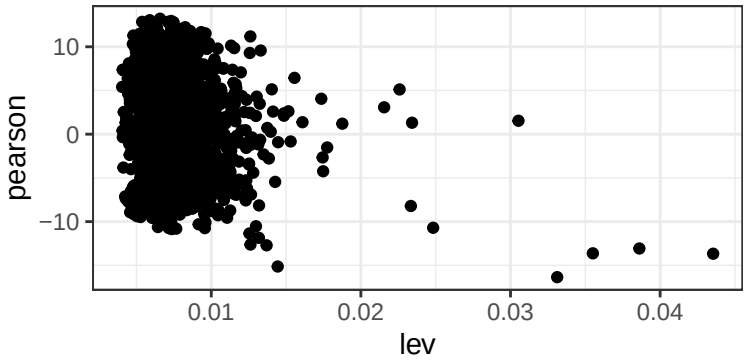


Figure 2: Residuals vs Leverage.

Appendix

More EDA