Chapter 6 Assignment

Introduction:

I have been exploring and looking at data from 24 years of Major League Baseball World Series Champions(1992-2006) in an effort to be able to analyze if you can use certain data to determine if your team has a chance to win the championship in any given year. I am using ggplots to chart out certain data points versus each other to see if there is any correlation to better make that determination of if your team does stand a chance based upon historical information.

Data:

The data I have chosen to analyze for this project are what I feel are key factors in a Championship winning team and play an important part in making a team successful. I have chosen the following data fields to analyze:

Win Ratio (team win % for entire season)

Payroll Rank (Ranking of team’s payroll in relation to all 30 teams)

Wins above Replacement{WAR} (team total of WAR for entire season)

-Wins Above Replacement or WAR, is a non standardized [sabermetric](https://en.wikipedia.org/wiki/Sabermetrics) [baseball](https://en.wikipedia.org/wiki/Baseball) [statistic](https://en.wikipedia.org/wiki/Baseball_statistics) developed to sum up "a player's total contributions to his team".[[1]](https://en.wikipedia.org/wiki/Wins_Above_Replacement#cite_note-1) A player's WAR value is claimed to be the number of additional wins his team has achieved above the number of expected team wins if that player were substituted with a [replacement-level player](https://en.wikipedia.org/wiki/Replacement-level_player): a player who may be added to the team for minimal cost and effort

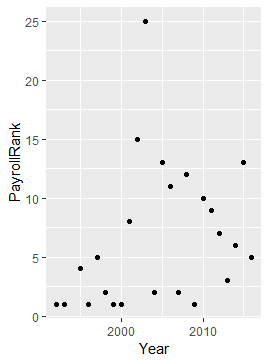
Wins over Replacement Rank (Ranking of teams WAR in relation on all 30 teams)

I feel that these 4 categories will help determine if your team has a chance to win the World Series in any given year. To clean the data down to these categories and years, I had to make a determination of what year I wanted to start with and delete the rest of the data I originally found. I made the year choice as that 25 years from the end of the data set (1994 didn’t have a World Series due to Labor Stoppage, that’s why there is only 24 Champions in data set). I also had to import from Baseball Reference all the WAR data and only limit it to World Series Champions.

Exploration and Findings:

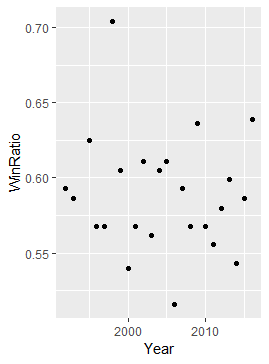
I have run 2 base graphs and 4 different comparisons in graph form to show my findings.

1. Payroll Rank



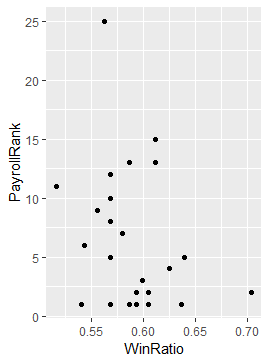
This shows that for the most part (outside the 2003 Florida Marlins) that you need to be within the top 15 (or top 50%) to win the World Series.

1. Win Ratio



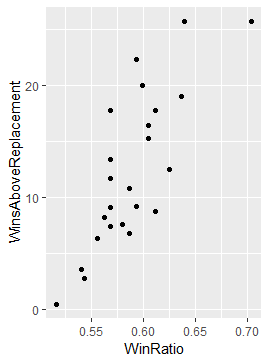
This shows that you need to be above 54% win % (exception is the 2006 St. Louis Cardinals with 51.6%) to have a chance to become World Series Champions.

1. Win Ratio vs Payroll Rank



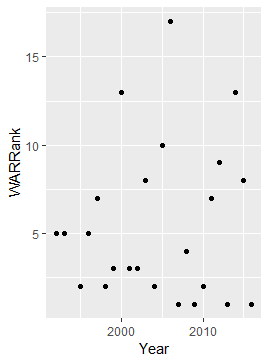
This graph shows that for the most part past world series champions have a higher win ratio the higher their payroll rank is. As you can see the win ratios over 0.6 mostly are within the top 10 in payroll rank, meaning that usually you have to spend in the top 3rd of the league and win more than 60% of your games to stand a chance at becoming world series champion.

1. WinRatio vs WAR



This graph shows the correlation between wins above replacement performance increases your win ratio. These values coexist, meaning higher WAR equals higher Win Ratio.

5)War Rank



This graph shows that you need to be within the top 10 in the whole league in WAR rank to win a World Series for the most part.

Conclusion:

From this analysis it shows that to be World Series champion, you need to follow the simple steps of being in the top 3rd in league in WAR Rank, top half of league in Payroll Rank, and win above 54% of your games and you should have a good statistical chance of playing for the title.