Independent Project: How to Win A League of Legends Match

Minwoo Kang

December 16, 2022

1. Introduction

League of Legends (LoL) is a game of multiplayer online battle arena (MOBA) genre developed by Riots Games and published in 2009. As of 2022, there are currently 180 million League of Legends players with the game's popularity rising consistently over the years.

League of Legends have the following tier progression: Bronze, Silver, Gold, Platinum, Diamond, Master, Grandmaster, Challenger. In a regular League of Legends ranked match - official games that increase or decrease your tier points (to essentially move up in tiers) depending on the game outcome of win or loss - players are put into two even teams of five members comprised of a jungle role, support role, bottom role, top role and middle role where the bottom and support roles are paired up to be in the bottom lane together. "Blue" team starts from the bottom-left side of the map while "Red" team begins their game from top-right side of the map. The objective of the game is to destroy the enemy "Nexus" while cooperatively and strategically progressing through the "turrets" in the way to the Nexus.

Each player is able to grow and gain more in-game strength during the game by obtaining game objectives that give special abilities for the players (Dragons, Rift Heralds and Baron), obtain gold - the currency used in League of Legends - through killing minions that yield experience points as well as gold and most importantly the enemy champions in order to buy stronger items that would give a player significant advantage in reaching the enemy base and ultimately in destroying the Nexus.

Previous literature states that higher kill and lower death counts in fact increase the chance of winning a game [1], showing that the capturing of kills and having less deaths would positively affect the game outcome. Additionally, another study has found that kills at every role lead to a worse chance to winning the game except for Top and Bottom positions [2], indicating that it is important to be aware of who within the team mainly consumes the experience points and gold.

In this case study, I aim to examine the in-game factors associated with winning the game. With each game generally lasting for 25 - 40 minutes on average, I want to better understand what predictors throughout the game are associated with higher probability of winning. To do this, I will specifically be looking into the in-game profiles and the performance statistics of Challenger-ranked players to better understand how the highest ranked players in the world go about winning a ranked game.

1.1 Dataset Used

For this study, we will be using League of Legends Challenger solo/duo ranked matches dataset that was originally published on Kaggle. The data set was gathered using Riot Games API and contains match history of all Challenger-rank matches that took place in EUW, KR, and NA servers in January, 2022. The data was collected such that each row entry in the data represents the information of one randomly chosen player from each game and the game ID - a designated ID that uniquely identifies each game - was never repeated in the data set, meaning that the data only strictly extracts information of one player from each ranked match.

The data set provides information with regards to the basic in-game player profile including champion played, side, role and level as well as in-game player performance statistics such as damage done to in-game structures and opponents, number of kills, deaths, and assists, gold earned and lastly whether or not the player has won the game.

1.2 Variables of Interest

In investigating the predictors that are associated with the winning game outcome, the response variable used in this study is a binary variable indicating whether or not the player has won the game: 0 if the player has lost the game and 1 if the player has won the game.

With regards to predictor variables, this study firstly includes the variables that effectively inform us of in-game player profiles. To do this, I made use of variables that tell us about which side the player started the game in (blue or red) and the player role (jungle, middle, support, bottom or top). The variable representing the champion played by the player has been excluded from the analysis mainly due to the fact that the trend - or the "meta" - of the champions used in a particular season (resets yearly) or a patch (updated every 2 - 3 weeks) changes on a frequent basis, implying that the data related to champions used may not be consistent with future updates.

Next, I have chosen to include the variables that effectively reflect a player's in-game performance during the match. The amount of gold earned has been included and mean-centered due to the generally large magnitude of the variable compared to other predictor variables in the model. Additionally, KDA Ratio is included to holistically represent the player's performance in terms of kills, assists and deaths. Based on our findings from EDA (Appendix B), high KDA ratio, which is given by calculating $\frac{\text{Number of Kills + Number of Assists}}{\text{Number of Deaths}}$, is generally associated with high number of kills, high number of assists, and low number of deaths, which is usually the ideal performance expected from players. Hence KDA ratio will be used to assess the player's capability to achieve high number of kills and assists while maintaining low death counts and therefore the variables representing the exact number of kills, assists and deaths have not been included. Lastly, the model further includes the player's in-game level (ranging from 1 - 18), total time spent applying crowd-control (CC) effects to opponents, damage dealt to opponents and taken, total creep score (minions killed) and the vision score (obtained when installing a ward or destroying enemy ward) to better understand the association between performance statistics and game outcome.

2. Exploratory Data Analysis (EDA)

We see from the EDA (Appendix B) that the proportion of won and lost games as well as the beginning side - Blue and Red - are similar and we can conclude that the data set has a fair distribution of game outcomes and the sides players started in. In addition, our EDA tells us that high KDA ratio is in general associated with high number of kills, high number of assists and evidently, low number of deaths.

3. Methodology

3.1 Description of Methodology

In order to investigate whether CABG or PCI is associated with more favorable outcomes in the data, we chose to fit a Cox proportional-hazards model with various patient demographic and health history covariates of interest. In addition, from our investigation of prior literature, we decided to add interaction effects between the type of procedure conducted and history of diabetes and number of diseased vessels. This is motivated by the fact that an overwhelming amount of papers published concluded that CABG was more effective for patients with a history of diabetes as well as patients who experienced more sever CAD. Thus, we were interested in understanding whether the effect of each surgery was perhaps dependent on certain underlying conditions.

In order to investigate the association between our predictor variables and the binary response variable representing whether or not a player has won the game, I decided to fit a logistic regression with various in-game profile and performance covariates of interest. Additionally, I have chosen to add an interaction term between a player's role and KDA ratio to further explore whether a winning game outcome is potentially

Variable	GVIF_DF2
Side	1.003400
Gold Earned	9.276028
KDA	6.131742
Level	6.797797
Crowd Control Time	1.361546
Total Damage Done	11.346082
Total Damage Taken	3.252756
Total Number of Minions Killed	17.534863
Vision Score	3.431328
Role	6.984143
KDA : Role	5.672935

dependent on certain roles and their kill, death and assist performance during the game. For the interaction term, I will be specifically focusing on bottom lane, which comprises of bottom and support roles as delineated by the research question.

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

The multicollinearity between the variables representing total damage done and total number of minions killed was initially a concern for this model. However, given that the total damage output as well as number of minions killed during a farming process for gold and experience points are two significant aspects of character growth in League of Legends and therefore I decided to include the two variables, tolerating the potential collinearity issue.

3.2 Model Specification

The following is the equation of the final logistic model: $\log(\frac{\pi_{Win}}{\pi_{Loss}}) = \beta_0 + \beta_{\text{Side}}x_1 + \beta_{\text{Gold Earned}}x_2$
$+ \beta_{\text{KDA Ratio}}x_3 + \beta_{\text{Level}}x_4 + \beta_{\text{Time Spent Applying CC}}x_5 + \beta_{\text{Total Damage Done}}x_6$
$+ \beta_{\text{Total Damage Taken}}x_7 + \beta_{\text{Total Minions Killed}}x_8 + \beta_{\text{Vision Score}}x_9 + \beta_{\text{Role}}x_{10}$
$+ \beta_{\text{KDA Ratio}\times\text{Role}}x_{11}$

3.3 Model Assumptions

Since logistic regression makes a few assumptions for the model, I will conduct model condition checks to make sure the model is valid.

Linearity

From our empirical logit plots (Appendix A), we can observe that there is a linear relationship between the empirical logits and the quantitative predictors that we have included in our model - gold earned, KDA Ratio, level, time spent applying CC, total damage done and taken, number of minions killed and vision score. Since the remaining predictors in our model are categorical variables, we do not have to assess the empirical logit for such variables.

Randomness

We do not have any way of knowing if the dataset used in this study is representative of all regular League of Legends games and therefore it is possible that randomness is not satisfied. However, given that our data source tells us that the sample size is large with around 17000+ entries, a player is selected at random from each unique game, and that Challenger players - the highest rank achievable - are commonly assumed to embody the ideal standards for high level of performance that users should try to adopt, we conclude that there is no reason for us to believe that the dataset used is significantly different from those from previous seasons and hence the randomness condition is satisfied.

Independence

As previously mentioned, each entry of the dataset represents one randomly chosen player from each uniquely identified game and the corresponding game ID was never used again to query for data, we can conclude that the independence condition is satisfied.

Cooke's distance

We can see from our Cooke's Distance plot for the final model (Appendix A) that there are no influential points. We observe that all of our observations fall below the threshold of 0.50 for Cooke's Distance and therefore all points can be left in the final model.

4. Result

I will be using a significance level of $\alpha = 0.05$ for the interpretation of results.

The odds of being high risk of heart disease for those with a college degree are expected to be 0.751 (exp(-0.286)) times the odds for those with some high school

With regards to the general in-game profile of the players, we observe from our model that roles of jungle, support and top are statistically significant predictors associated with a winning outcome. We are 95% confident that the expected odds of a player of jungle role winning the game are expected to be 0.039 to 0.097 times lower than those of a player of bottom role, holding all else constant. Similarly, we expect the odds of a player of support role winning the game to be 0.175 to 0.378 times lower than those of a player of bottom role with 95% confidence, holding all else. Lastly, we are 95% confident that the expected odds of a player of top role winning the game are expected to be 1.087 to 1.867 times higher than those of a player of bottom, holding all else constant.

The model also shows that with respect to the in-game performance of the players, the amount of gold earned, KDA ratio, in-game level, time spent applying CC, total damage taken, total minion kills and vision score were the statistically significant predictors associated with a winning outcome. Out of these variables, KDA ratio exhibited the highest increase in the odds of winning the game. We are 95% confident that for each additional one point increase in KDA ratio, the odds of winning the game are expected to be 2.189 to 2.545 times higher, holding all else constant. In addition, for each additional one point increase in in-game level of a player, we expect the odds of winning the game to multiply by 1.601 to 1.760 with 95% confidence, holding all else constant. However, it is interesting to note that for each additional one minion count increase, the odds of winning the game in fact decreases by a factor of 0.978 to 0.983 with 95% confidence, holding all else constant. A potential explanation for this could be that given that performance factors such as KDA ratio and level show a high significance in the model, such factors allow for a greater in-game lead and consequently a higher chance of winning over the advantages brought by total minion kills. Another variable that contrasted with my belief was vision score. We expect with 95% confidence that for each additional one vision score increase, the odds of winning the game decreases by the factor of 0.987 to 0.994, holding all else constant.

Regarding the interaction terms, I discovered that only the interaction between KDA ratio and role of support is statistically significant. Given that a player takes a role of support instead of a role of bottom, for each additional one point increase in KDA ratio, we expect the odds of winning the game to multiply by a factor of 0.789 to 0.960 with 95% confidence, holding all else constant. This was another interesting observation as it is generally assumed that higher KDA ratio would result in higher probability of winning the game and our results suggest that higher kills and assists combined with low deaths for support role has negative effect on the game, which is consistent with the previous literature [3].

5. Discussion

5.1 Conclusion

For this case study, my aim was to explore how different in-game profiles and level of the player performance is associated with a League of Legends game outcome. Through my experiences as well as previous literature, I expected KDA ratio, total damage done, total number of minion kills and vision scores to be the factors that would increase the odds of a player winning a game, holding all else constant. As expected, higher KDA ratio - assumed to have high kills, assists and low deaths - was statistically significant. This was unsurprising given that within the League of Legend system, higher kills and assists yield gold and experience points (EXP) to boost leveling growth while deaths yields the enemy a certain sum of gold and also penalizes with longer respawn time for the player as the in-game level increases. But it was shown by our model that both of total number of minion kills and vision scores lead to lower expected odds of winning the game, holding all else constant. As mentioned previously, such observations may be caused by the importance of higher kill and assist contributions paired with low death counts and its overall advantages overruling those achieved through higher vision scores and minion kills. Therefore, a potential takeaway from this is that players should focus on level growth and capturing kills and assists more so than applying CC, aggressively getting more minion kills or acquiring higher vision scores for higher chance of winning.

5.2 Potential Limitations and Future Directions

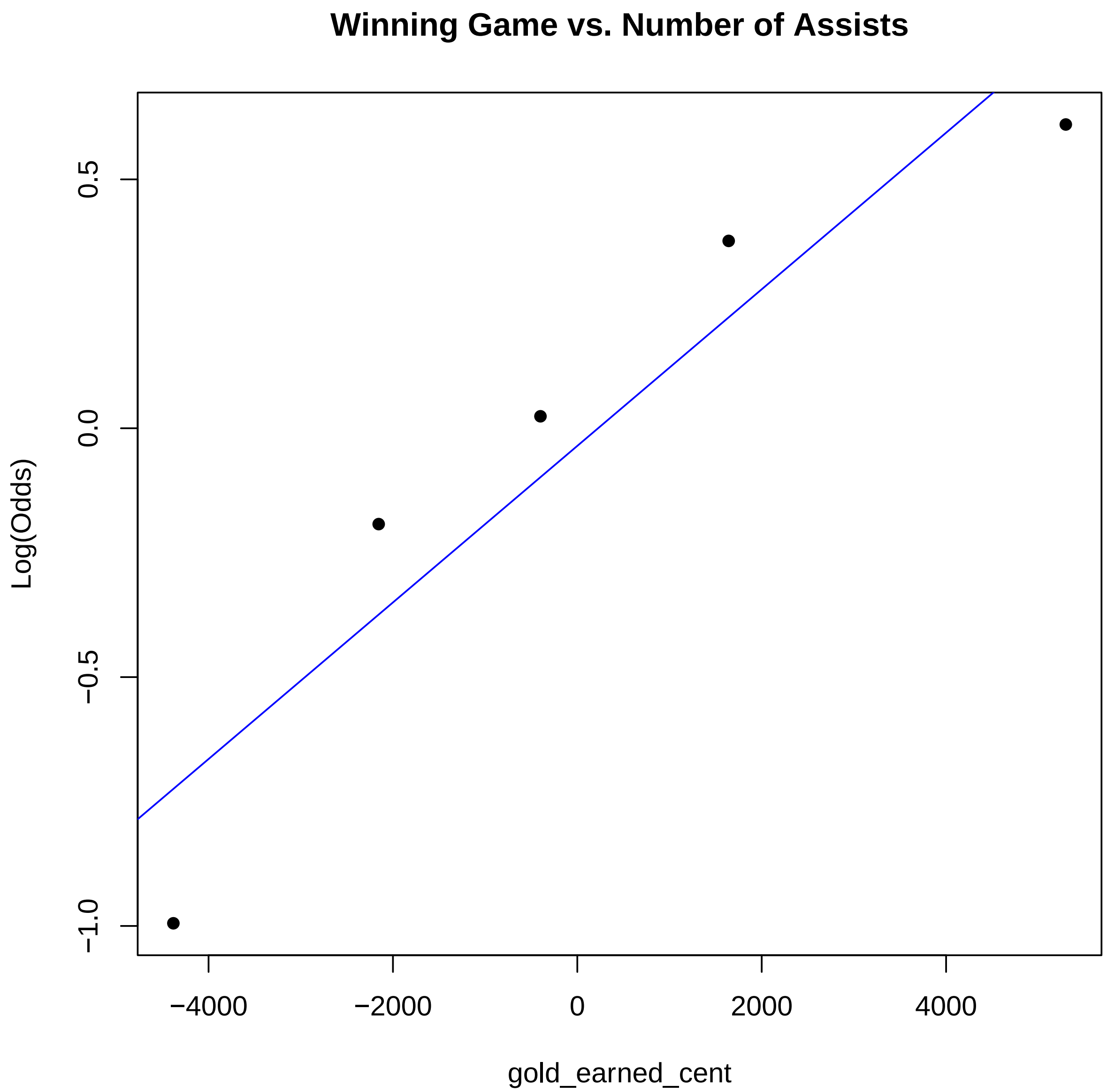
A potential limitation from this study is that the data only contains the match data from January, 2022. As mentioned before, League of Legends undergoes major changes every year - especially at the beginning of the year as the season resets - and also several patches during a season that may completely change the trend or meta of the game at the time. For instance, a champion widely used in January, 2022 may have gone through major balancing process during the patches and no longer be picked as often. It would be interesting to obtain data of different seasons over the years and compare the differing trends in player behaviors.

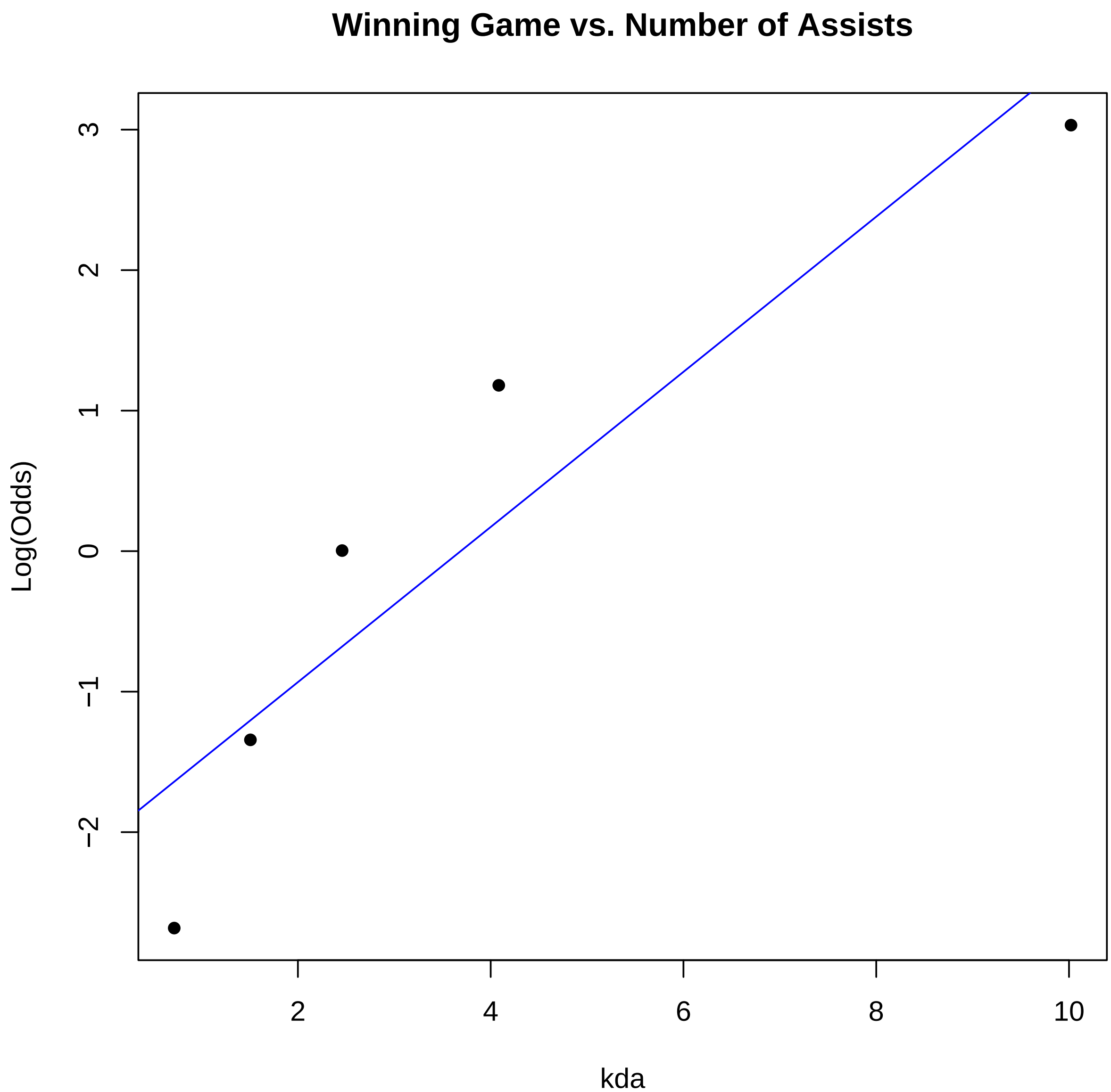
Another limitation that may have interfered with our results is the fact that the data set being used in the study is a combination of LoL game data from Europe West (EUW), Korea (KR) and North America (NA) servers. As previous literature suggests, the meta of these three servers differ significantly, not to mention that the bi-monthly patches and updates are entirely based on the statistics of NA server. This would be greatly improved if I could use a data set that has greater number of data points for each server regions or if I focus on one server to keep the underlying conditions consistent.

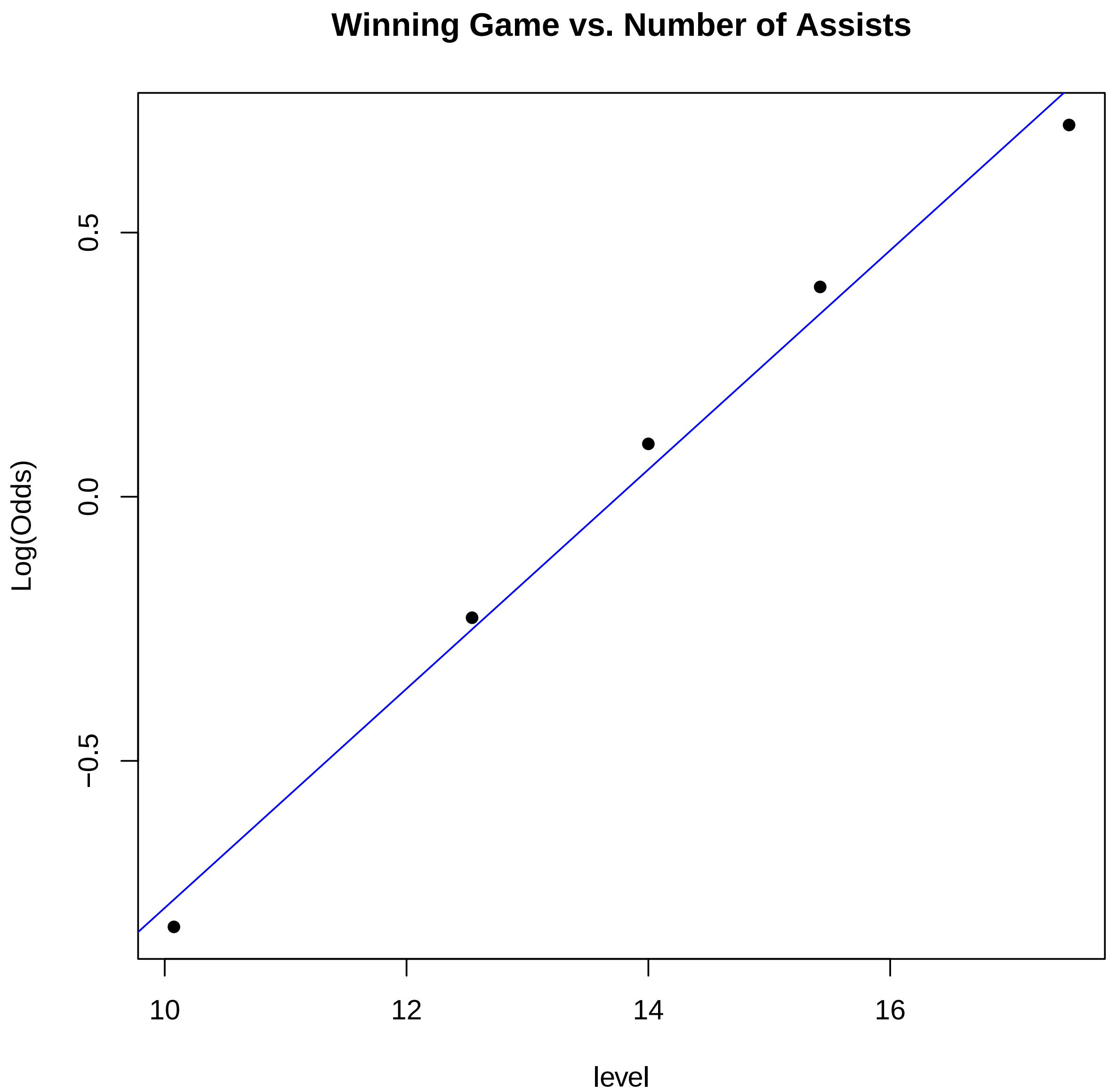
Appendix A

Model Diagnostics

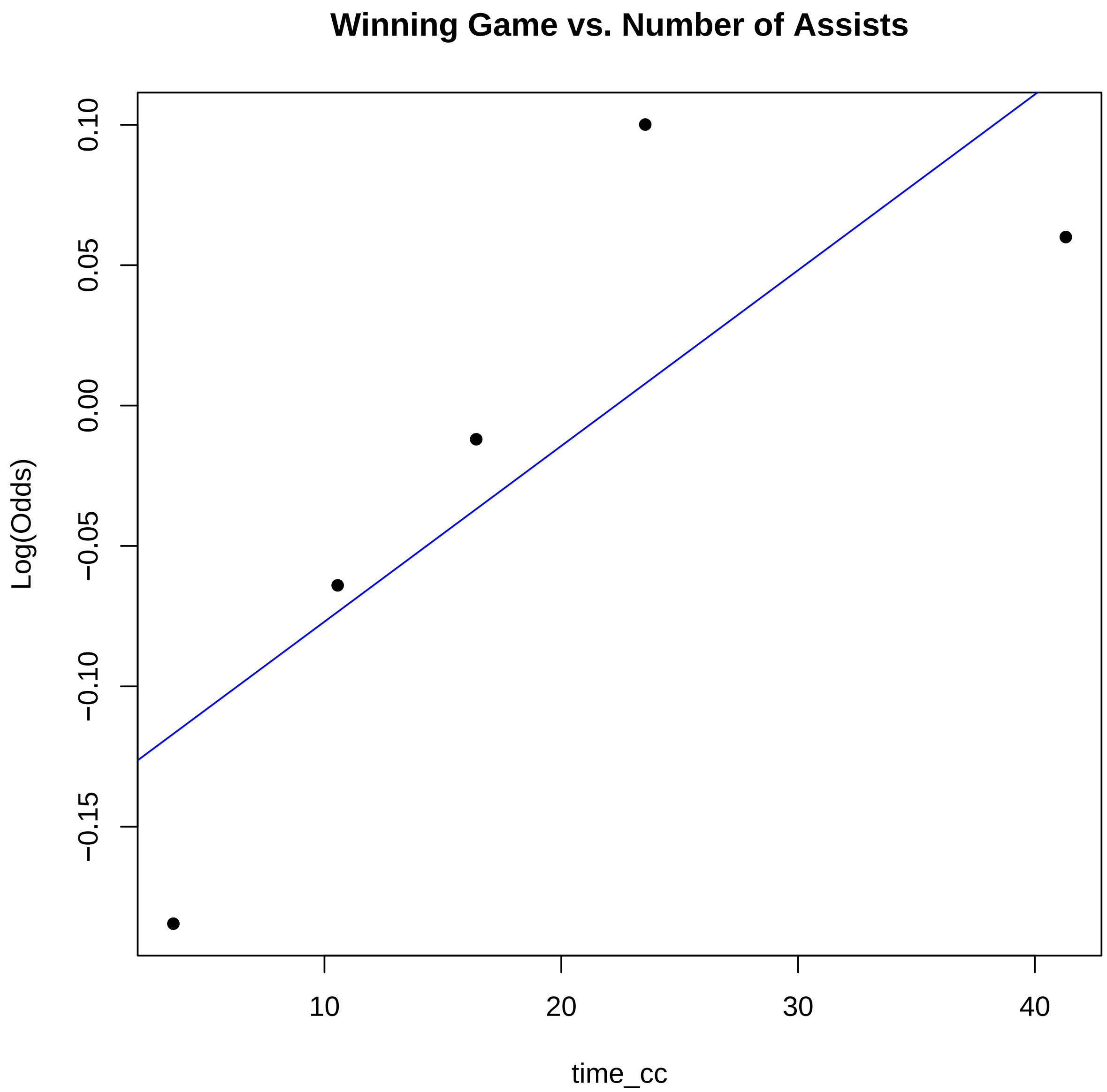
Linearity

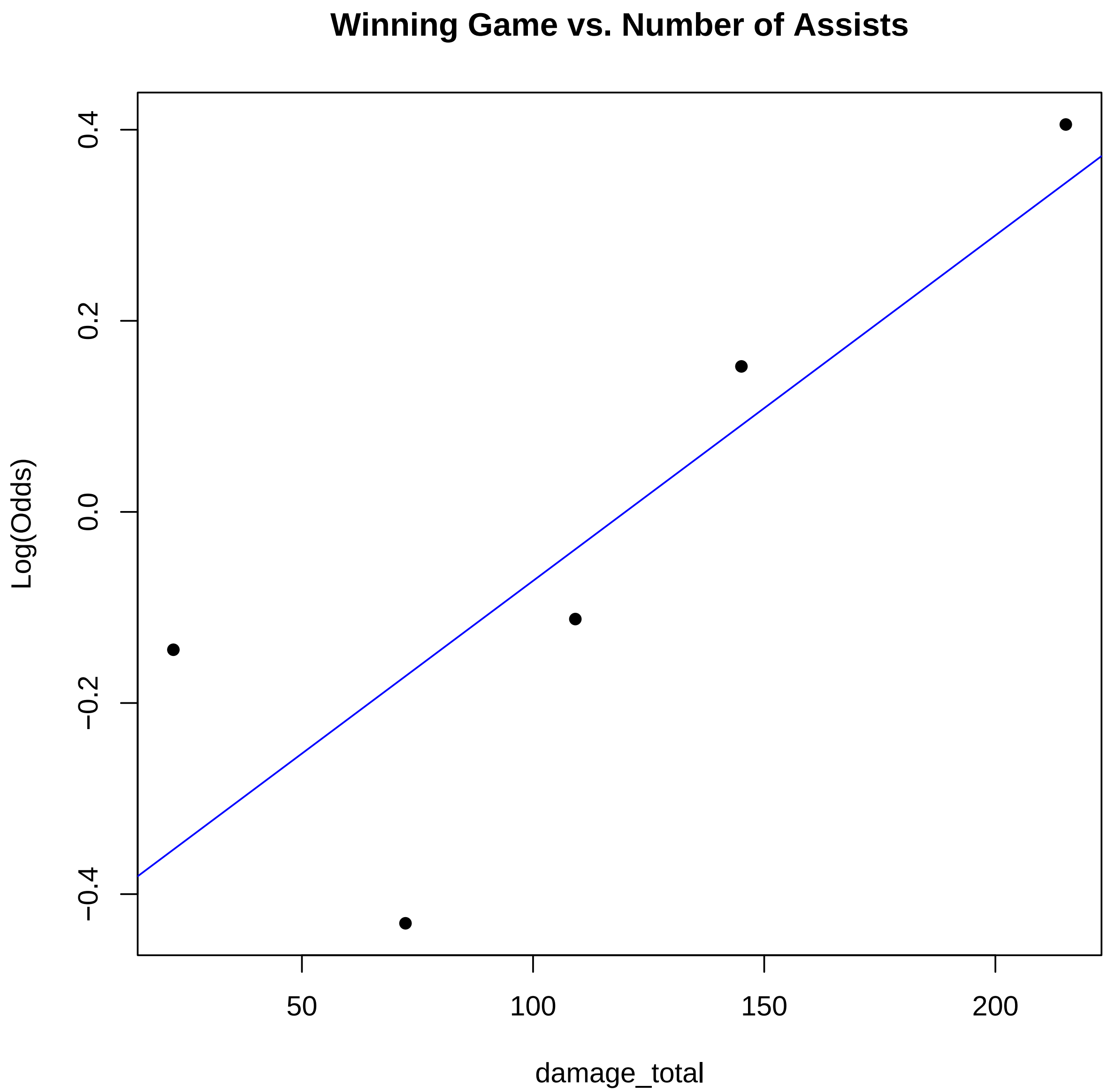

Winning Game vs. Number of Assists
Log(Odds)
0.5
0.0
−0.5
−1.0
−4000
−2000
0
2000
4000
gold_earned_cent


Winning Game vs. Number of Assists
Log(Odds)
3
2
1
0
−1
−2
2
4
6
8
10
kda

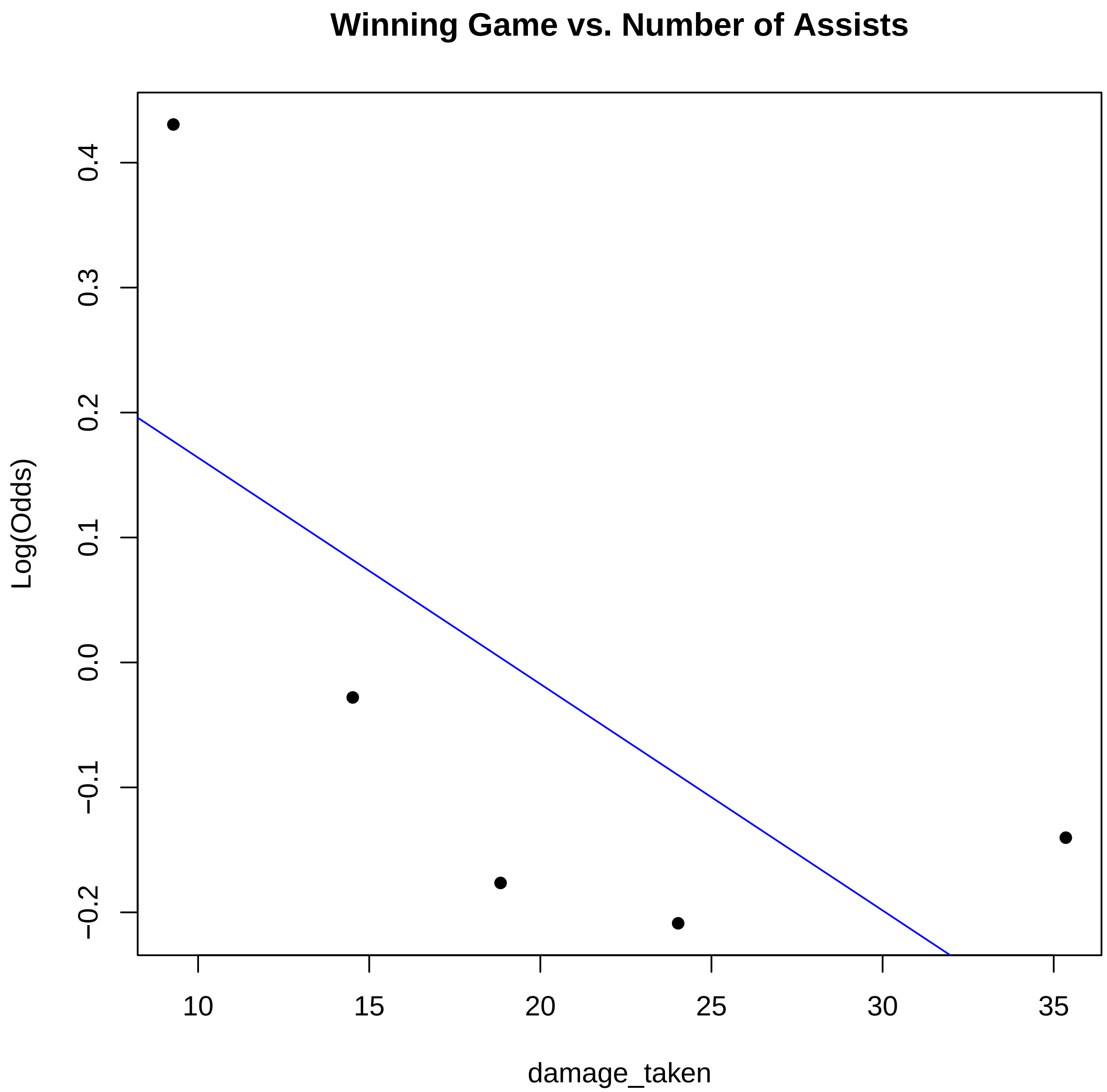


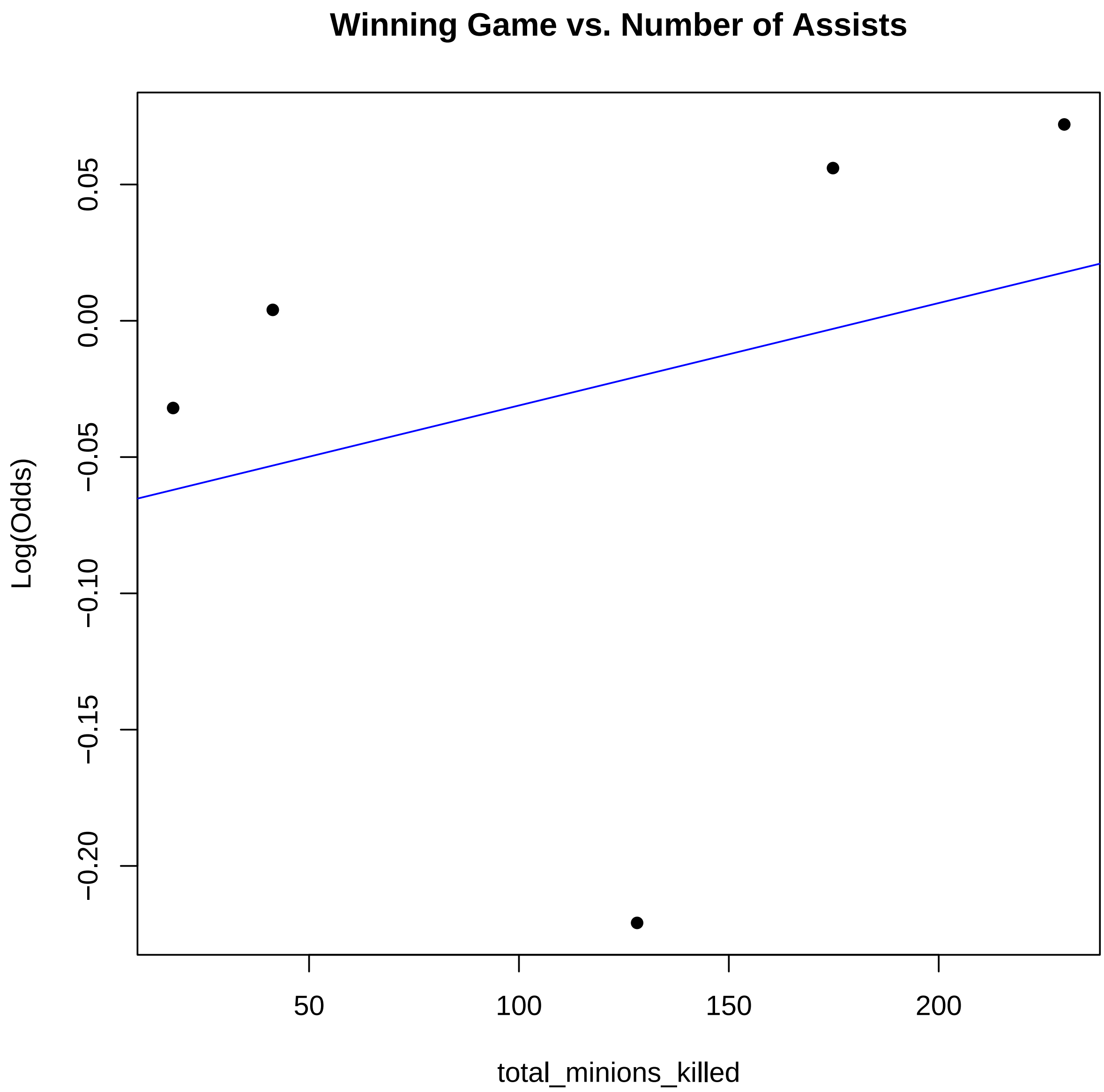
Winning Game vs. Number of Assists
Log(Odds)
0.5
0.0
−0.5
10
12
14
16
level

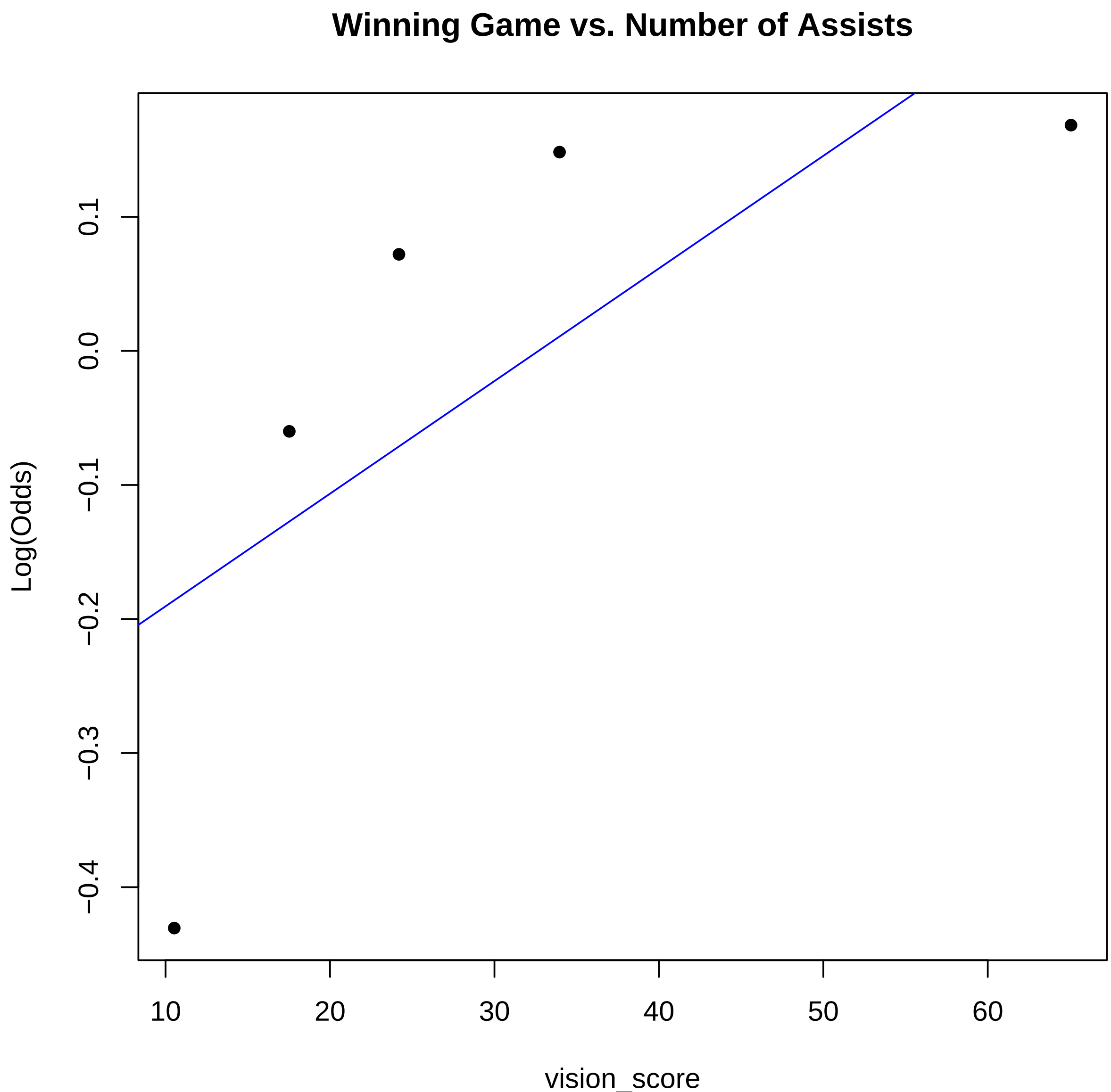

Winning Game vs. Number of Assists
Log(Odds)
0.10
0.05
0.00
−0.05
−0.10
−0.15
10
20
30
40
time_cc



Winning Game vs. Number of Assists
Log(Odds)
0.4
0.2
0.0
−0.2
−0.4
50
100
150
200
damage_total


Winning Game vs. Number of Assists
Log(Odds)
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
10
15
20
25
30
35
damage_taken


Winning Game vs. Number of Assists
Log(Odds)
0.05
0.00
−0.05
−0.10
−0.15
−0.20
50
100
150
200
total_minions_killed



Cooke's Distance

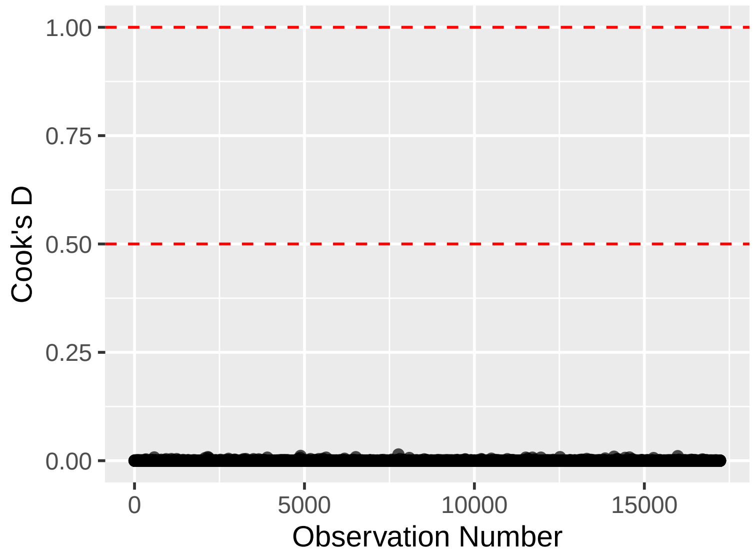
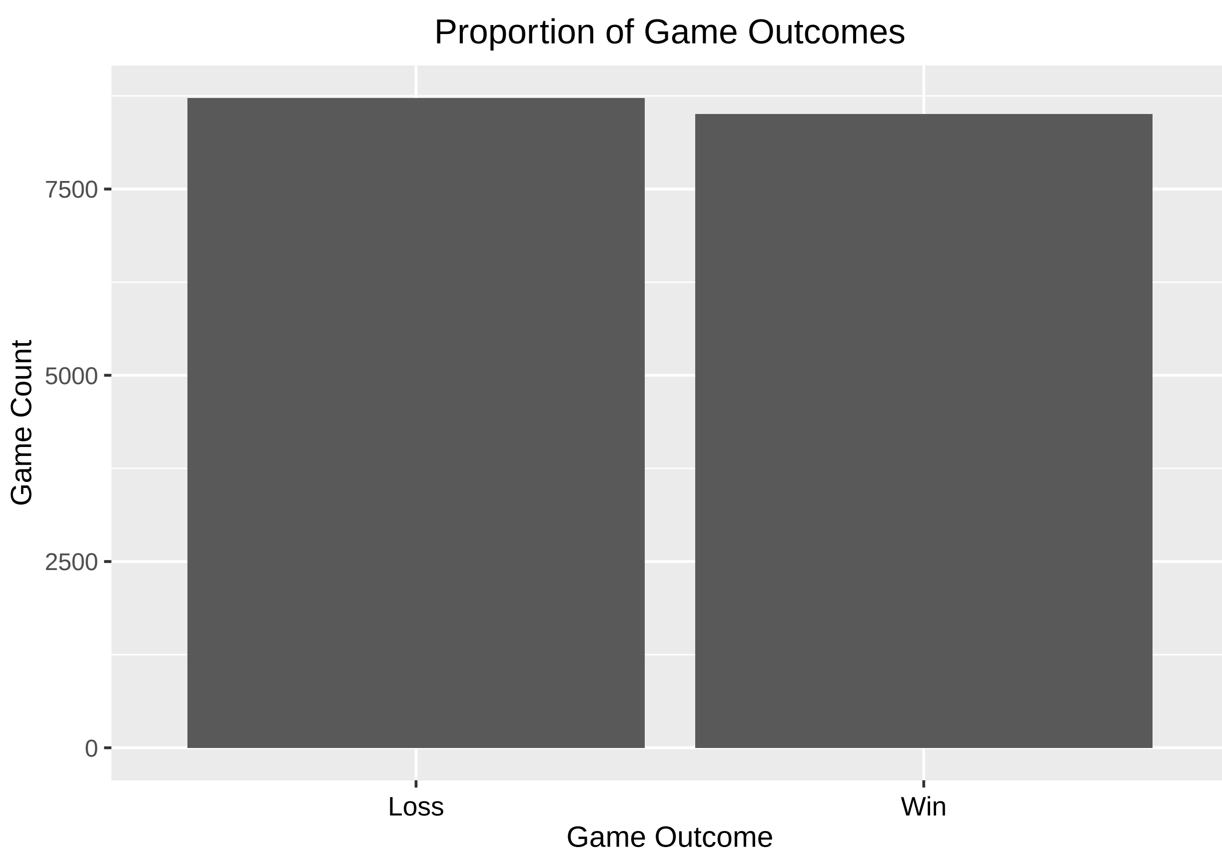
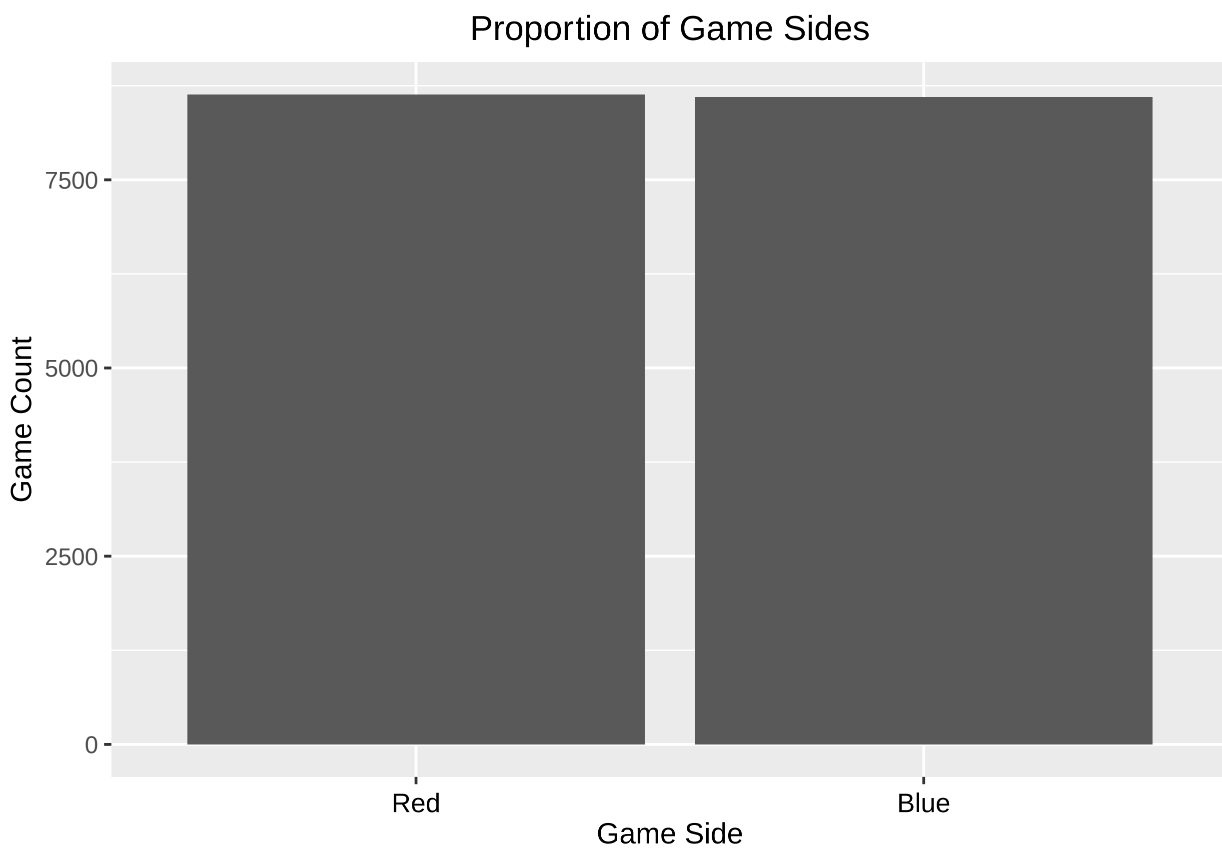


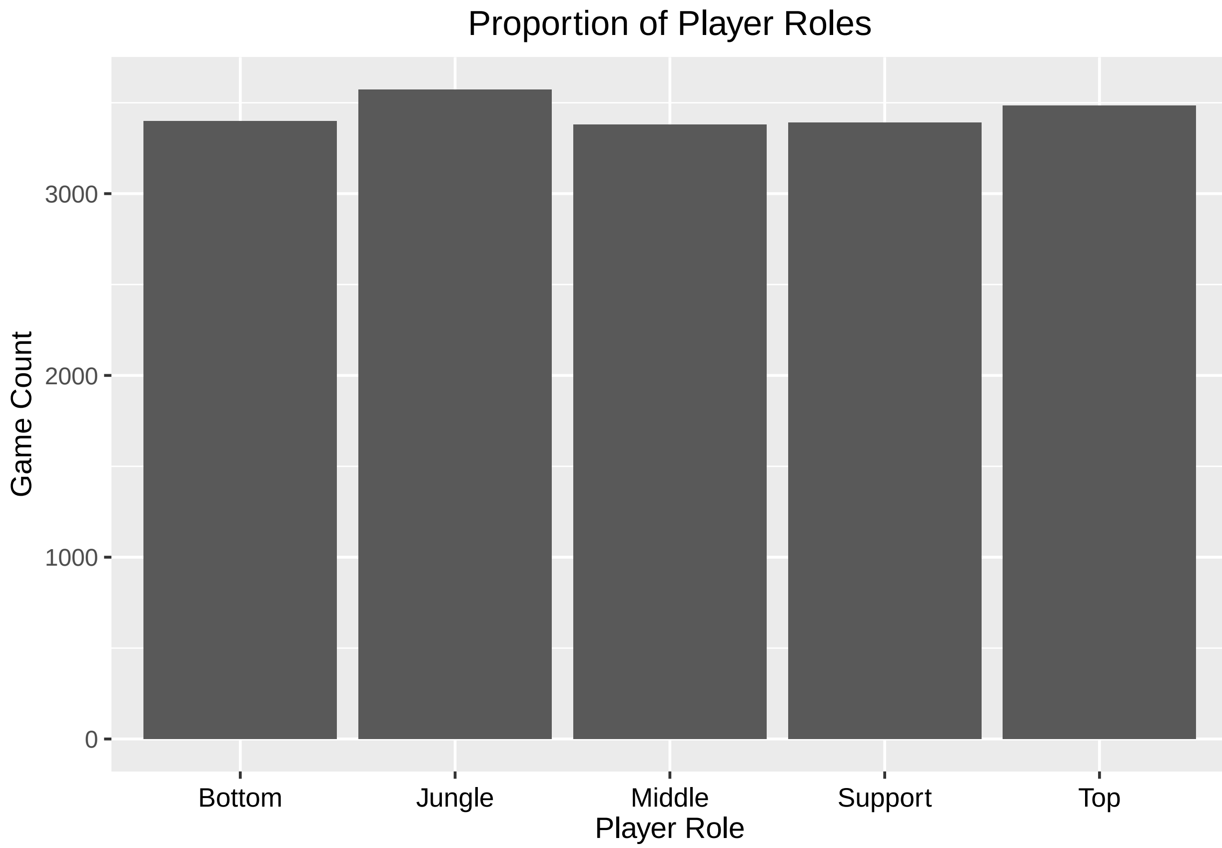
Figure 1: Cooke's distance plots

Appendix B

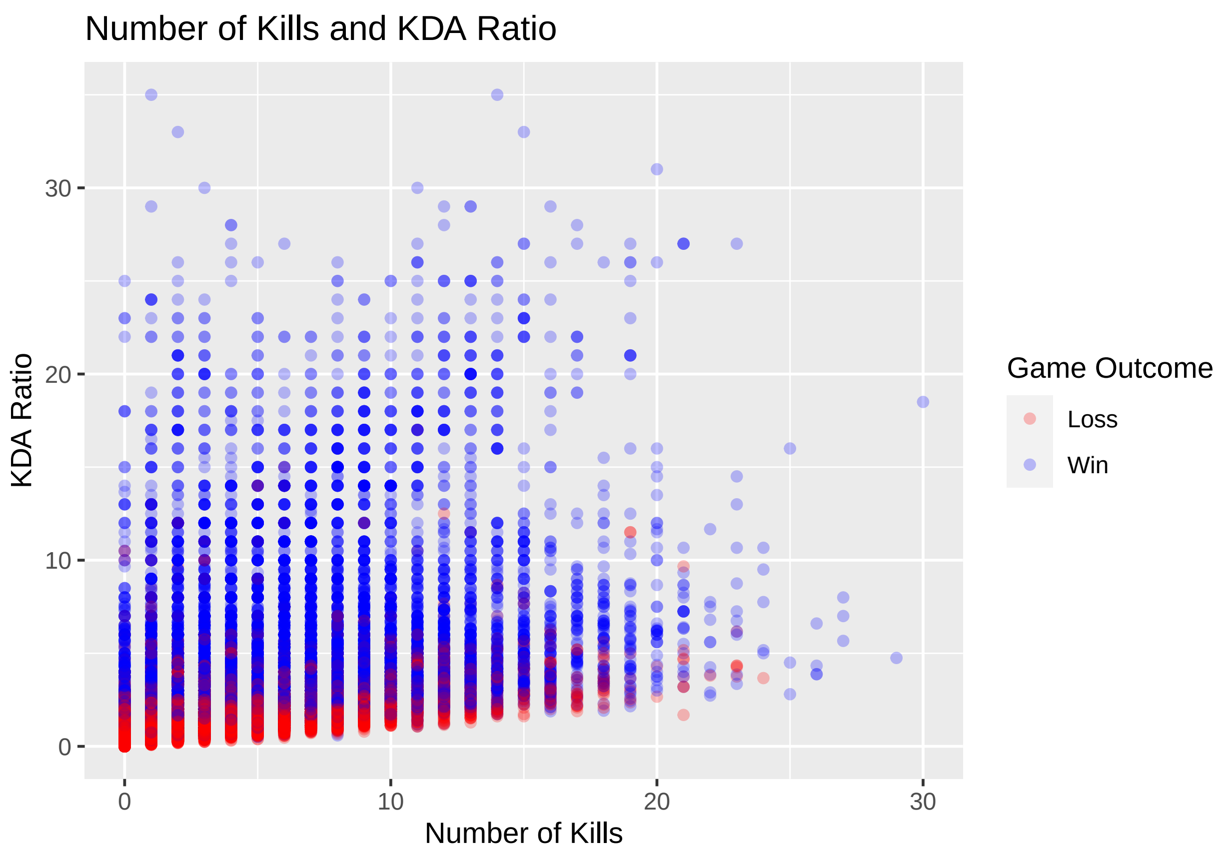
Exploratory Data Analysis



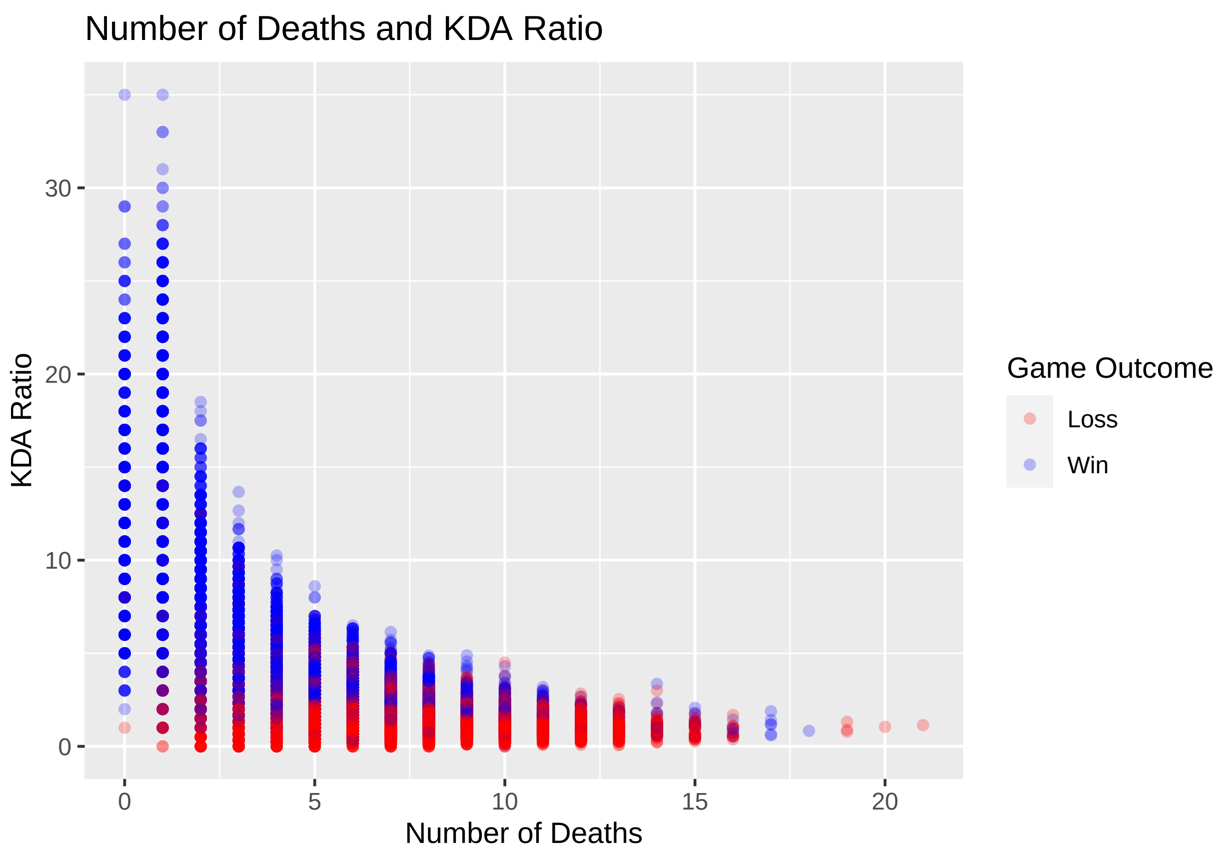

Proportion of Game Sides
Game Count
7500
5000
2500
0
Red
Blue
Game Side

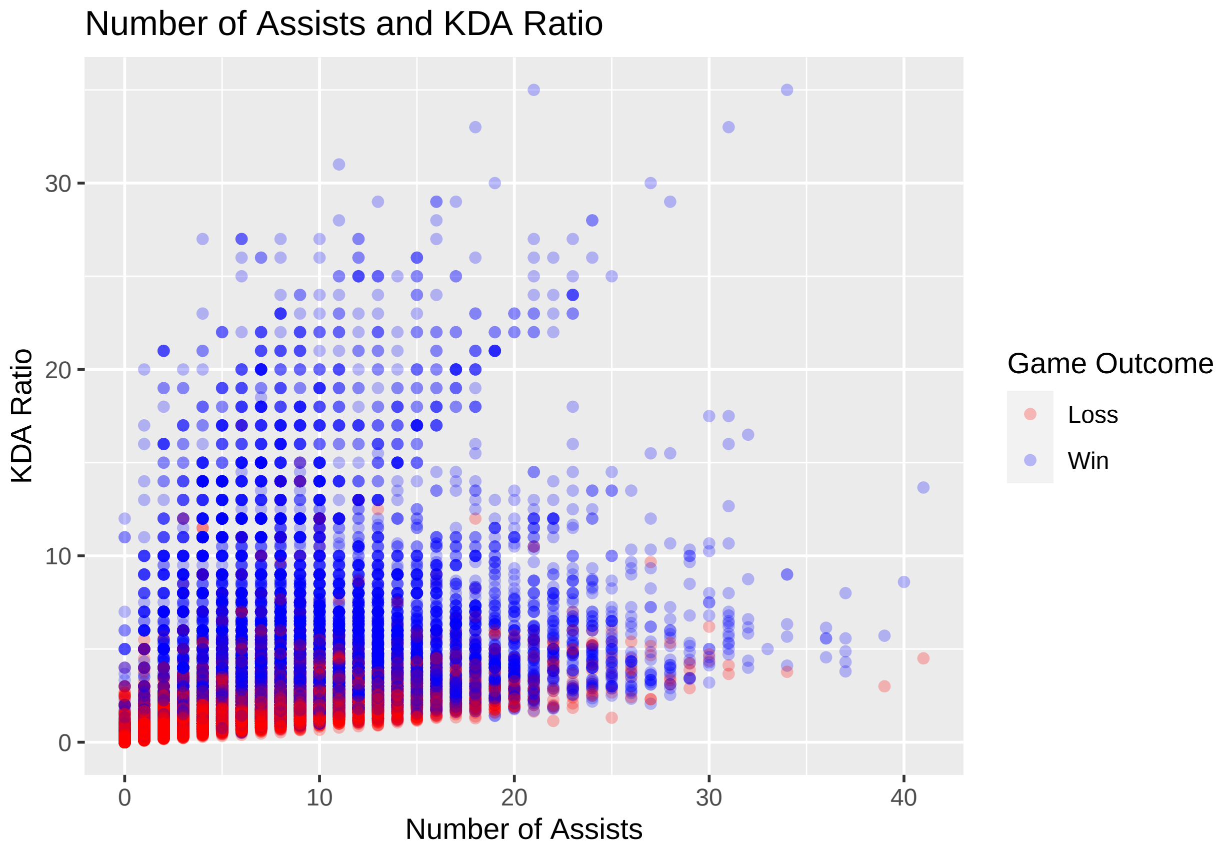


Note that for all plots of kill, assist show data points for 30+ kills or assists. These are unusually high values - by norm, high number of kills and assists will generally mean 10 - 15 kills and 15 - 20 assists respectively.



Number of Kills and KDA Ratio
KDA Ratio
30
20
10
0
0
10
20
30
Number of Kills
Game Outcome
Loss
Win


Number of Deaths and KDA Ratio
KDA Ratio
30
20
10
0
0
5
10
15
20
Number of Deaths
Game Outcome
Loss
Win



Sources

[1] Xia, Bang, et al. "What Contributes to Success in Moba Games? an Empirical Study of Defense of the Ancients 2." Games and Culture, vol. 14, no. 5, 2017, pp. 498–522., https://doi.org/10.1177/1555412017710599.
[2] Maymin, Philip Z. "Smart Kills and Worthless Deaths: Esports Analytics for League of Legends." Journal of Quantitative Analysis in Sports, vol. 17, no. 1, 2020, pp. 11–27., https://doi.org/10.1515/jqas-2019-0096.
[3] "Esports Win Probability: A Role Specific Look into League of Legends." Samford University, https://www.samford.edu/sports-analytics/fans/2020/Esports-Win-Probability-A-Role-Specific-Look-into-League-of-Legends.