# League of Legends Research Paper

Blake Casey, Sophie Panagrossi, Benjamin Tisinger, Devene Bellamy, Cristobal Gavaldon

IST-687

December 12, 2022

**Introduction:**

League of Legends is the subject of the dataset, which is a team based strategy game consisting of two teams each containing five players/people. There are over 100 characters (champions) to choose from, and each team will fight one another, and pursue objectives to gain advantages to assist in achieving victory.

The original dataset contained extraneous variables that were not necessary for the data analysis, so was subsetted, resulting in the columns below.

| Fields | Description | Example |
|---|---|---|
| GameID | Session ID or each Game ID represents a single game | 4035079604 |
| champExperience | How much experience a champion has. The higher the experience, the higher the champLevel | 15453 |
| champLevel | Level of champion at game end. (18 max) | 16 |
| championName | Name of the champion | Sett |
| deaths | How many times the champion/player died | 6 |
| gameEndedinSurrender | Whether the game was surrendered (as opposed to losing) | TRUE |
| individualPosition | Which role the champion played | TOP |
| kills | How many other champions were killed | 15 |
| magicDamageDealtToChampions | Magic damage dealt to champions | 678 |
| MagicDamageTaken | Magic damage taken | 4518 |
| physicalDamageDealtToChampions | Physical damage dealt to champions | 20578 |
| totalDamageDealtToChampions | Total damage dealt to champions | 23814 |
| totalDamageTaken | Total damage taken | 29672 |
| wardsPlaced | Wards placed | 7 |
| win | Whether the team won. (TRUE = win) | TRUE |

**Objective:**

The objective of this investigation is to use statistical analysis to help us determine whether certain champions are inherently better picks than others.

**Methods:**

A method we used for answering what is the correlation between damage received and the probability of the game ending in surrender, and what is the correlation between total damage and win rate is a linear regression. It is used to study the linear relationship between a dependent variable Y (champTotalDam, champEndedInSur) and the independent variable X (win, champTotalDamTaken). The form of the model would be:

champTotalDam = b0 + b1*win

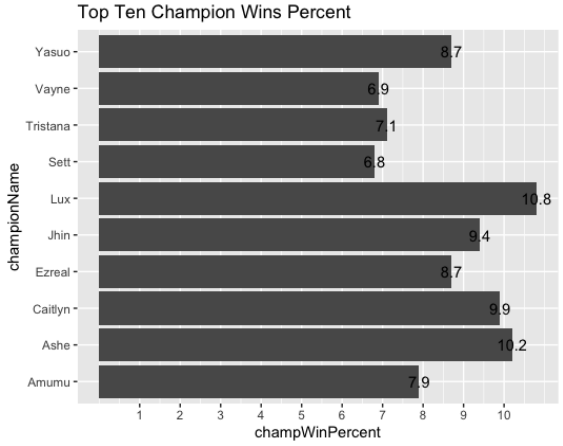
champEndedInSur = b0 + b1*champTotalDamTaken

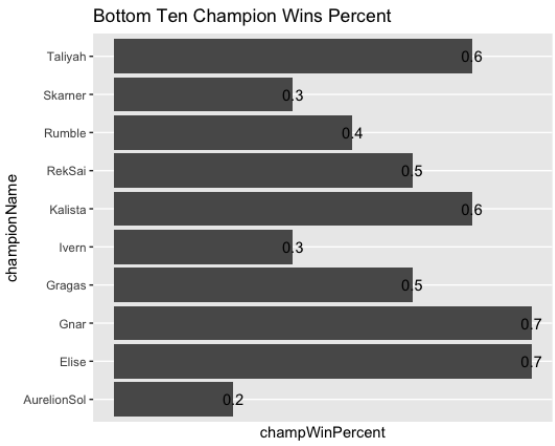
As well as using a correlation test to answer the questions, what is the correlation between damage received and probability of the game ending in surrender, and what is the correlation between total damage and win rate? The correlation analysis measures the strength of a linear relationship between the two variables and calculates their association. We are specifically using the Pearson correlation coefficient; this test ranges from -1 to +1. -1 means there is a negative association and +1 means a positive association.

We used the library, Tidyverse, to answer the rest of our questions. Tidyverse is used to clean and tidy data. Using the Tidyverse pipe (%>%), we were able to manipulate and create new datasets from our original data. By doing so, this allowed us to find a win percentage, What utility champions have the highest win rate at level 18, and which level each player made it to win or lose. This allowed us to then use ggplot() to create all of our graphs. Our first question is what is the win rate per champ? The champion with the highest win rate is Lux with a 10.8% win rate, and the champion with the lowest rate is AurelionSol with 0.2%. The top ten champion win percentages range from 6.8% to 10.2%. The Bottom ten percent were all below 1%. We can use this with our future experiments to see which champions are better picks.

**Results/Discussion**:

Question 1: What is the win rate per champion?




The champion with the highest win rate is Lux with a 10.8% win rate, and the champion with the lowest rate is AurelionSol of 0.2%. The top ten champion win percentages range from 6.8% to 10.2%. The Bottom ten percent were all below 1%. We can use this with our future experiments to see which champions are better picks.

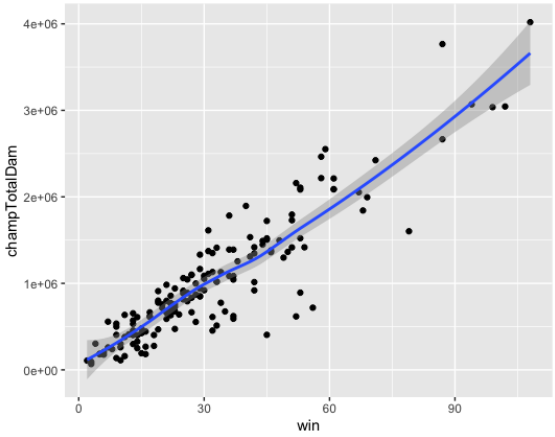
Question 2: What is the correlation between total damage and win rate?

```
lm(formula = champTotalDam ~ win, data = champstats)

Residuals:
     Min       1Q   Median       3Q      Max
-1064501  -112660     8978   149546  1004796

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    15719      44966    0.35    0.727
win            31560       1176   26.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 306300 on 154 degrees of freedom
Multiple R-squared:  0.8238,    Adjusted R-squared:  0.8227
F-statistic: 720.2 on 1 and 154 DF,  p-value: < 2.2e-16
```




```
        Pearson's product-moment correlation

data:  champstats$champTotalDam and champstats$win
t = 26.836, df = 154, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8753656 0.9318807
sample estimates:
      cor
0.9076536
```

This linear model translates to, with every increase of win, the total damage on average increases around 31560. The R-squared of the model is 0.8; meaning 80% of the variation in the output variable is explained by the input variables. When testing Pearson's correlation test, the results show that the p-value is 2.2e-16, and the correlation coefficient is .908. The p-value of the test is less than the significance level alpha = 0.05. We can conclude that champ total damage given and win percentage are significantly correlated.

The graph above shows the relationship between the champion total damage and the win percentage. It shows that the higher the damage the champion does, the win percentage increases, as well as the higher the win percentage the higher the damage the champion does to their opponents. It does support our conclusion that there is a relationship.

We looked at the top 5 champions from question 1, and put their stats into the linear regression. We want the player with the highest summary output because we want the player with the highest damage given to other players. Lux had the highest with a summary output of 3424199.

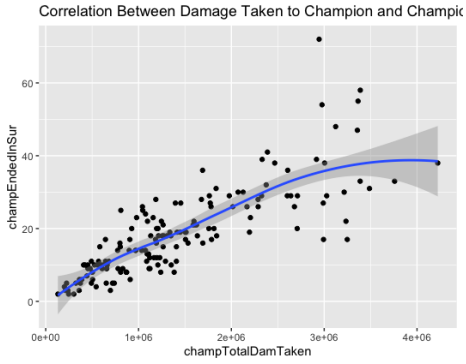
<u>Question 3: What is the correlation between damage received and probability of game ending in surrender?</u>

```
Call:
lm(formula = champEndedInSur ~ champTotalDamTaken, data = champstats)

Residuals:
    Min      1Q  Median      3Q     Max
-21.385  -3.969  -0.449   3.184  36.895

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.221e+00  1.062e+00   3.033  0.00284 **
champTotalDamTaken 1.083e-05  6.266e-07  17.277  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.025 on 154 degrees of freedom
Multiple R-squared:  0.6597,    Adjusted R-squared:  0.6574
F-statistic: 298.5 on 1 and 154 DF,  p-value: < 2.2e-16
```




```
        Pearson's product-moment correlation

data:  champstats$champTotalDamTaken and champstats$champEndedInSur
t = 17.277, df = 154, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7508824 0.8596188
sample estimates:
      cor
0.8121895
```

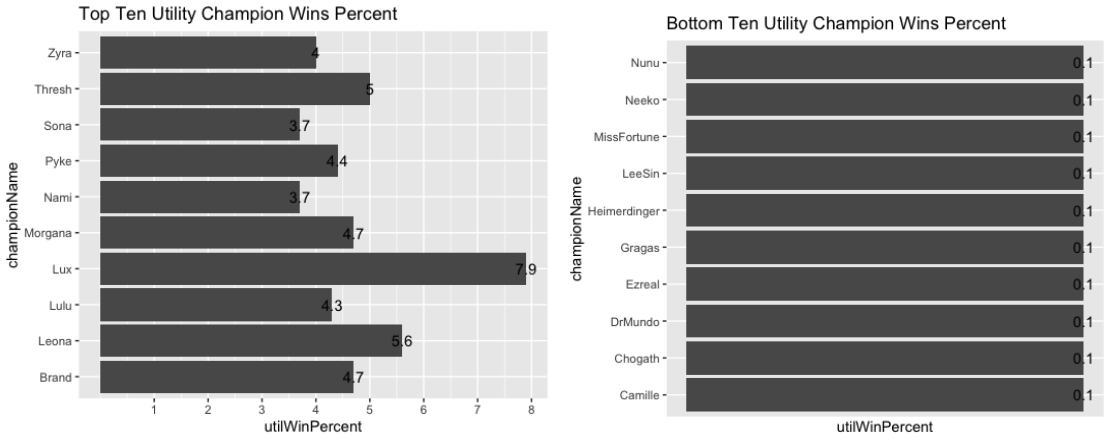
With every damage taken to the champion, the champion ending in surrender increases by 1.083e-05. With an R-squared of 0.6574, showing a moderate effect size- 66% of the

variation in the output variable is explained by the input variables. Looking at the correlation test, we can see that the p-value is 2.2e-16, and the correlation coefficient is 0.812. Because the p-value is less than 0.05, we reject the null hypothesis, and we can confirm there is a significant relationship between the two variables.

The graph for question 3 seems to have a linear relationship in the beginning; as the champ's damage taken increases the more of a chance the champion ends in surrender. But towards the end, around 3e-6 damage, the graph seems to level out at the maximum. We can conclude from the graph that there is a moderate correlation between the two variables.

We looked at the top 5 champions from question 1 to see who has the lowest summary output. We want a small summary output because damage taken to the champ gives the champ a lower chance of winning. Lux had the lowest summary output of 35.111.

Question 4: What utility champions have the highest win rate at level 18?

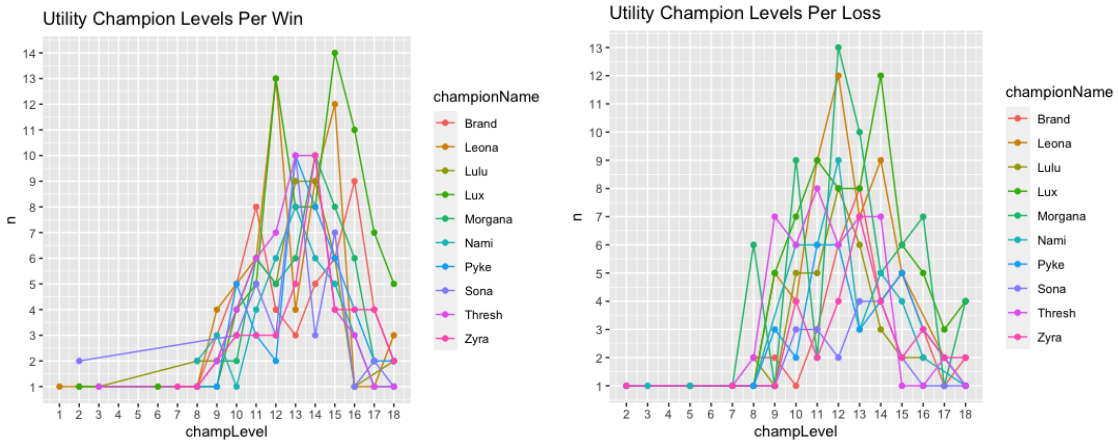



The first graph has the top ten champions and their win percentage. We are looking at the champions who identify as Utility. Lux has a win rate of 7.9%, followed by Leona, Thresh, Morgana, Brand, Pyke, Lulu, Zyra, Sona, and Nami.

Our second graph shows the bottom ten champions that identify as Utility based on their win rate. Every single champion has a 0.1% win rate.

Lux has shown that she has the top winning percentage in champions looking at all positions, and when looking at just the utility position Lux still has the highest win rate.

Question 5: What is the win/loss ratio of utility champions per level




Looking at the top ten Utility champs from the previous question, the graph shows how many wins or losses they got per level. Each level gets increasingly harder. We can see that the higher the level, the higher chance the player has to win or lose. Lux has the highest number of wins after level 15. We can conjecture that since Lux wins more than the other players at the higher levels, she is a better player as the game progresses through the levels than others.

**Conclusion:**

The majority of our data propose that Lux is statistically better than other champions not only in the class in which she belongs, but also ranks the highest win rate amongst all champions of all roles. There are certain stages of the game where there are better picks.

The scope of our investigation didn't look at how all champions did throughout the game, as this was only focused specifically on utility champions. For a future study, if we were to expand the scope and include all champions and all roles, it may reveal how teams could strategically compose their roster to emphasize certain points of the game (such as aiming to win early in a short game or play a longer game and win late).

More forward-looking ideas would include normalizing the dataset to only include games that the specific champion participated in. In the current state, the statistics do not accurately reflect whether a champion was present in a game or not. Additionally, adding another dataset for the sake of comparison could yield some valuable insights. One of the problems with adding in another dataset would be the lack of any DateTime data or timestamps. This was not in scope

for this analysis because it would involve more time than the project allows as it would likely have us collecting our data or calling out to an API.