

Project 1 158

Tesfa Asmara and Kevin Loun

2/7/2022

Hello data

The dataset for this project contains 10,000 League of Legends ranked matches from the North American region with 775 variables offered through the Riot Games API, provided on Kaggle. The reference page for the Riot Games API can be found at: <https://developer.riotgames.com/apis>. The dataset can be found on Kaggle at: <https://www.kaggle.com/jamesbting/league-of-legends-ranked-match-data-from-na>. This dataset will be referred to as `lol10`.

Each row, or observational unit, in the table represents a single match. The columns represent characteristics, or variables, of each match. In this project, we will consider the following variables: lane, champion, map side, first summoner spell, kills, deaths, gold earned, and total damage dealt. A figure including all the relevant variables and their description is attached at the end.

Since the responses can take on a wide range of numerical values, kills, deaths, gold earned, and total damage dealt are numerical variables. Because the responses themselves are categories, lane, champion, map side, and first summoner spell are categorical variables.

Summarizing the data

In this section, we will consider the measures of central tendency and the measures of dispersion of the numerical variables for summoner 1 on the blue side of the map. Measures of central tendency include mean and the median, while the measures of variability include standard deviation, variance, and the interquartile range.

```
summary(lol10$b_summoner1_kills)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.000	2.000	5.000	5.469	8.000	31.000

```
summary(lol10$b_summoner1_deaths)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.000	3.000	5.000	5.531	8.000	21.000

```
summary(lol10$b_summoner1_gold_earned)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	665	7883	10213	10387	12659	27393

```
summary(lol10$b_summoner1_total_damage_dealt)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0	60723	104292	109334	147933	684420

From the output, we can infer that, for summoner 1 on the blue side of the map, the average number of kills is 5, the average number of deaths is 5, the average gold earned is 10,386, and the average total damage dealt

is 109,333. For 50% of the matches for summoner 1 on the blue side of the map, the number of kills is above 5, the number of deaths is above 5, the gold earned is above 10,213, and the total damage dealt is above 104,292. The standard deviation provides a measure of the amount of variation of the data from the mean. To see this consider the `b_summoner1_kills` variable; 6,800 of the 10,000 matches (68%) had a number of kills within 1 standard deviation of the mean, and 9,600 of the 10,000 matches (96%) had rates within 2 standard deviations. Moreover, the IQR for the number of kills is 6, the number of deaths is 5, the amount of gold earned is 4,776, and the total amount of damage dealt is 87,210.

Histograms

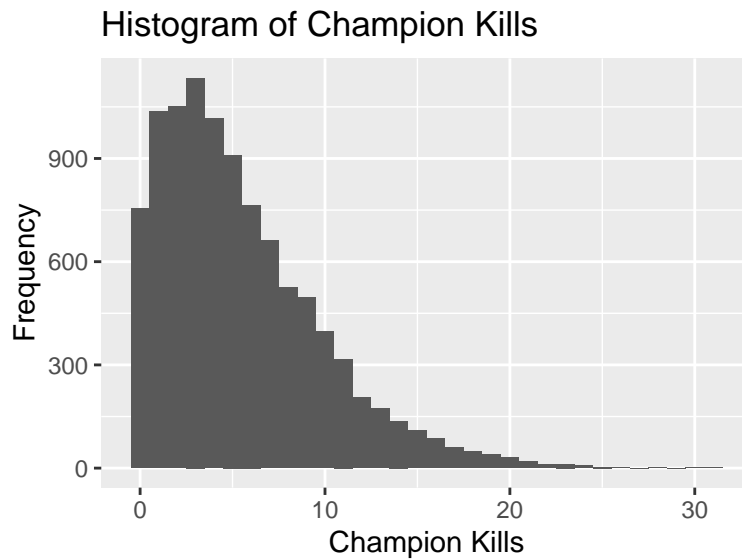


Figure 1: In League of Legends, the primary objective is to destroy the enemy team’s base, or nexus, getting kills on members of the enemy team provides experience and gold which allow the player to buy items that help provide an advantage over the enemy team in this objective. The Histogram shown plots the distribution of kills that players earned over the course of ten thousand games

This histogram reveals that the distribution of kills per game is right skewed indicating that it is common to get a low number or no kills but it is uncommon to get a large amount of kills. It should be noted that there is an outlier at $X = 30$ that may cause extra skewness in the data.

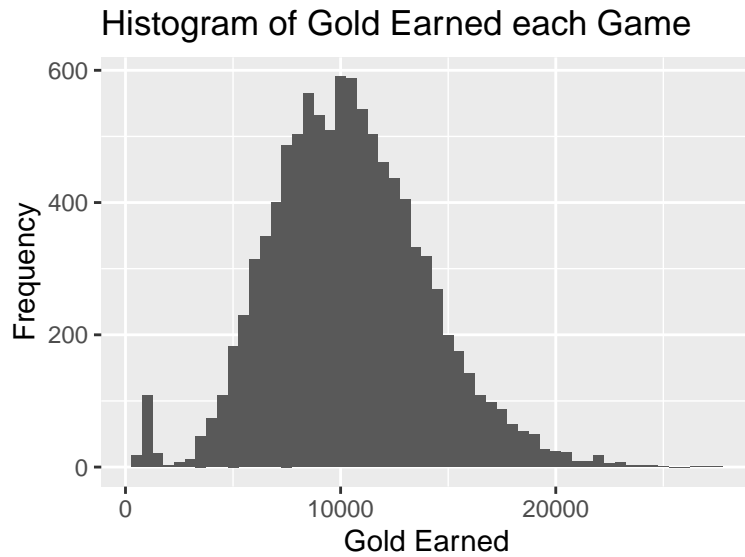


Figure 2: In League of Legends, Gold is the primary currency earned. It can be earned passively or actively by completing objectives and defeating enemies. Gold allows players to buy items that increase their strength and give them advantages against the enemy team. This histogram plots the gold earned by players over the course of ten thousand games.

This histogram reveals that the amount of gold earned per game is slightly skewed but very similar to a bell shape distribution. This means that in a game it would be common to earn around 10000 gold but rare to earn around 0 or more than 20000. One thing to note is that there seems to be an abundance of data near $X = 0$ that indicates there are many outliers in the data. # Scatter Plots Bellow are scatter plots constructed from some selected variables in order to see if there is any possible correlation between them. This can be useful as these variables are assumed to be correlated by many players already and can give insight on whether the data contradicts common sentiment.

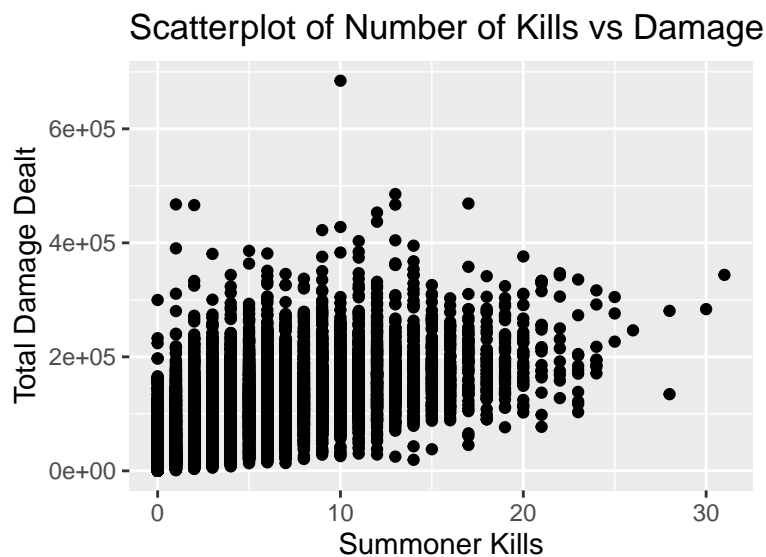


Figure 3: In League of Legends the general consensus is that the more damage you deal the more likely you are to accumulate kills, This histogram shows exactly that: Number of Kills vs Damage Dealt

This scatterplot compares the amount of kills and the total amount of damage dealt in a game. The scatter plot does not have a constant variance of errors due to the point at $X = 10$. Additionally, it seems that the plot has a positive trend indicating that number of kills may have a positive correlation with damage dealt.

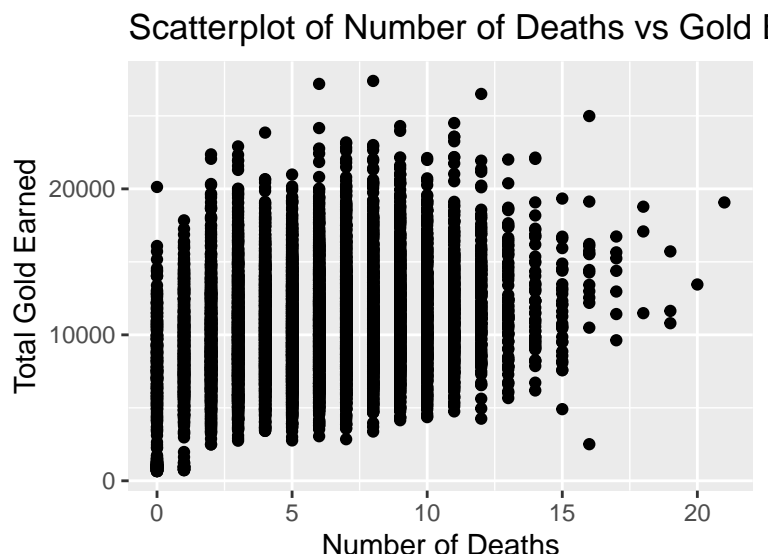


Figure 4: In League of Legends, deaths put the player at a disadvantage by providing their opponent with additional gold and experience. So it is generally assumed that more deaths would result in a player having a more difficult time earning gold. This Scatterplot shows Deaths vs Gold Earned

This scatter plot shows a positive trend between number of deaths and gold earned. It also appears that there is constant variance of errors in this plot. This plot implies that as the number of deaths increases so does the amount of gold earned.

Thoughts on Data

Most of the data met our expectations. However, plotting number of deaths earned and gold earned yielded surprising results. Generally, giving one's opponent an advantage would make the game more difficult. However, once the data was plotted, it seemed to indicate that deaths and gold earned had a positive correlation contrary to our expectations. The rest of the data met our expectations for what we would see based on our own knowledge of the game. We believe that sampling went well and that our sample is representative of our population because our data had ~10000 observations, a large sample size. This large sample size will help us generalize our findings to a general population of all league of legends matches.

Figure 5: Variables and their descriptions