Election Preferences in 2022 based on Gender and Education*

Rahul Gopeesingh

April 19, 2024

There are many factors that can influence someone's voting preference. Two of which that seem very relevant are a person's gender as well as their education level. This paper uses a logistic regression to investigate the relattionship between these factors and a person's political preference. We found that both females and highly educated people are significantly more likely to vote for Biden than for Trump

Table of contents

---

*Code and data are available at:https://github.com/Rahul-Uoft/election.git

1 Introduction

In this paper we predict the outcome of a given game of league of legends played by players of the pro teams T1 and Generation Gaming(Gen G) based on their creep score(CS) at 10 minutes into the game and the highest level lead they have against their opponent.

League of Legends is one of the most popular video games in 2024 and has held that stature for the last decade.[CITE TWITCH STATS] It is an extremely nuanced game but a brief overview shall be provided here but it will be explained with more depth in the Section 5. The game consists of 2 teams of 5 players each, in which both teams compete to destroy the enemy 'nexus.' The map is split into 3 lanes and a 'jungle' which is a dark area in between each of the lanes. While the primary objective of the game is simple, that is, to destroy the enemy nexus, there are a variety of strategies that players can implore to do it. For instance, one team may spend the majority of its time maximizing its gold income in order to buy items that make them stronger and once this is done, it will be easier to kill the enemy players and destroy the enemy nexus. Another popular strategy would be to sacrifice your own gold income to deny the enemy gold income as well as experience. While the gold may be even, you will have more experience points and will be stronger with higher levels. This strength disparity once again can allow you to kill your opponents easier and yup, you guessed it, destroy the enemy nexus.

The two gaming organizations, T1 and Gen G, have been battling at the top level for the last few years. They are both Korean teams, with Korea being arguably the strongest region in the world. At the time of writing this paper, T1 are the current world champions, and have some of the highest caliber of players. They were the favourites to win the 2023 world championship and they are home to the most renowned and respected league of legends pro player of all time, Faker. However there is a caveat to their apparent dominance. Domestically, they struggle to beat Gen G. Many fans coin Gen G as 'T1's Kryptonite.' In the 2024 Korean Championship,

Gen G beat T1 in a very close best of 5 that ended 3-2. Many fans also believe that T1 only won the world championship because Gen G got knocked out by a Chinese team.

This paper seeks to shed light on this historic rivalry by analyzing how good the players are at using their leads to close out games. The two types of leads that we will look at will be a high gold income as well as a straight out experience lead against their 'lane opponent'(more details are in the discussion)

We use R Core Team (2023) and (**tidy?**) in order to conduct this analysis

The remainder of this paper is structured as follows. Section 2 discusses the data, how it was collected and the purpose in which we use this data. It also gives a brief analysis of the data. Section 3 walks through the logistic regression model used and the justification for using this logistic regression. Section 4 shows the results and an intepretation of what they mean. Finally, Section 5 discusses the potential reasons behind these results and their implications.

2 Data

2.1 Data Source

The data used in this paper was downloaded directly fromt the Riot Games API, the company that owns league of legends. This is available to everyone for free(although an account must be created.) When a game of league of legends is played online, a large amount of data is recorded and stored in their database. These can then be accessed using their API, which allows this data to be queried. The data used in this paper comes from games played by each of the players on the pro teams, and were within the last 2 years. This allows it to be more indicative of their current performance, rather than their performance from 2016. It is taken from their solo queue games, which is essentially the regular games that anyone can play online by themselves and not the professional games in which all of the team's players team up and play against another full team. (As the name suggests, solo queue means you are queueing for a game alone.)

It is worthwhile mentioning that there are many third party data sources with the data already cleaned and presented in easily digestible formats however the accuracy of these can be called into question as many times they are incoherent statistics among them. Therefore the most reliable dataset is from the source itself.

2.2 Variables of Interest

There are 3 variables that we are concerned with, the player's creep score(cs) at 10 minutes into the game, the player's highest level lead against his opponent, and the outcome of the game. The implications of these are discussed thoroughly in the discussion but the overarching

belief is that the stronger you are, the easier it is to win the game. The higher the cs at 10 minutes is, the stronger you are as you have more gold to buy items. The same goes for the level disparity between you and your opponent. However, being stronger does not guarentee a win. Some players may be better at converting their leads into victories and therefore we analyze one of the richest rivalries in league of legends history, T1 vs Generation Gaming.
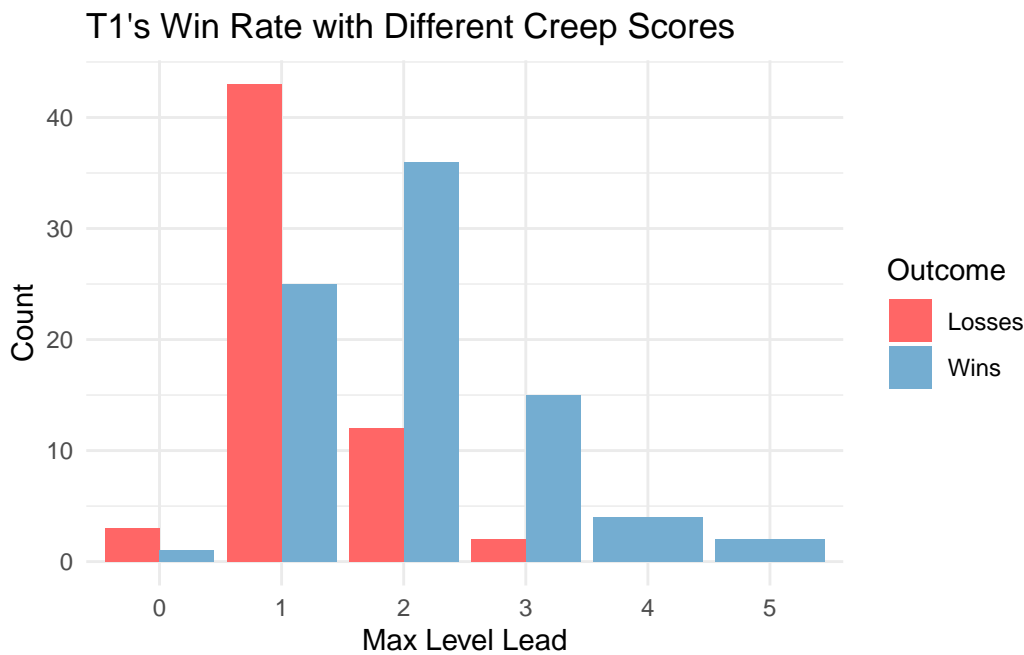

Figure 1: T1's CS at 10

Figure 1 shows the wins and losses based on different level leads from all the T1 players.

Figure 2 Shows the level and win rates of the T1 players without considering their support and jungler. These are 2 of the 5 players that do not benefit from killing creeps(increasing their creep score) and their level disparity is less significant. As such, this can skew the data if we include them in our analysis. We can see this by comparing these two figures, in which the proportion of wins to losses based on different level leads changes in the way one would expect, increases for higher level leads and decreases for lower level leads. These figures serve two purposes, showing how the win rate increases with increased experience advantages(level lead) and they also show why we should exclude both support and jungler from our analysis.

2.3 Data Summary

Figure 3 shows the wins and losses of all geng laners with their highest level lead while Figure 4 shows the same for the laners of T1. It is interesting to observe the subtle difference at the lower cs levels as these indicate differing play styles however this is not the topic of interest.
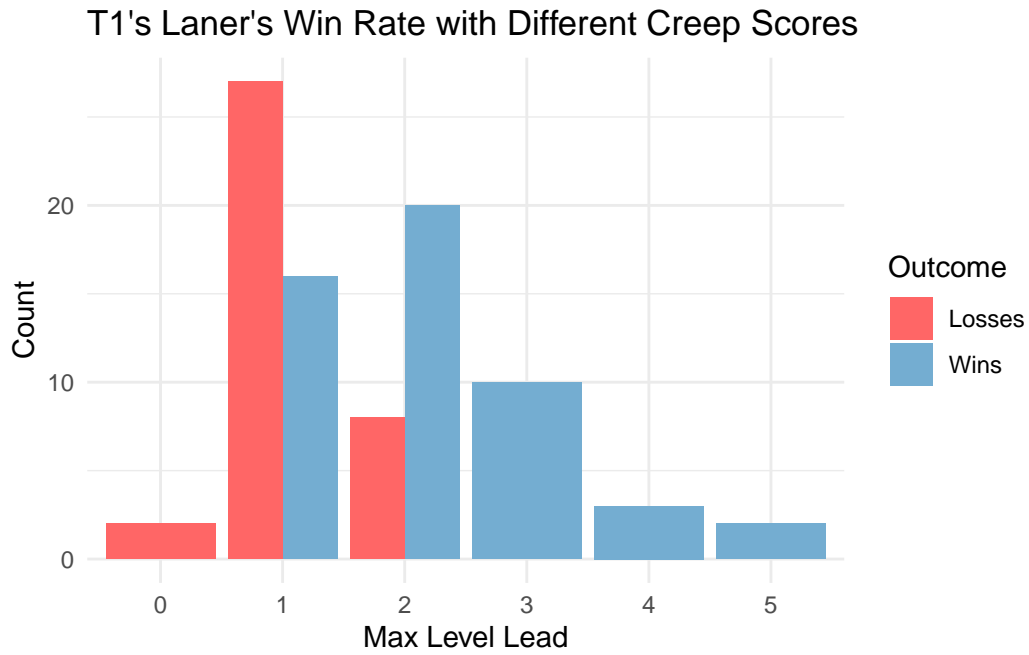
Figure 2: T1's Laner Level Lead and Wins

Finally, we should mention that the games of the players on each team were considered together to gain a better understanding of how the team functions rather than the individual players on the team.
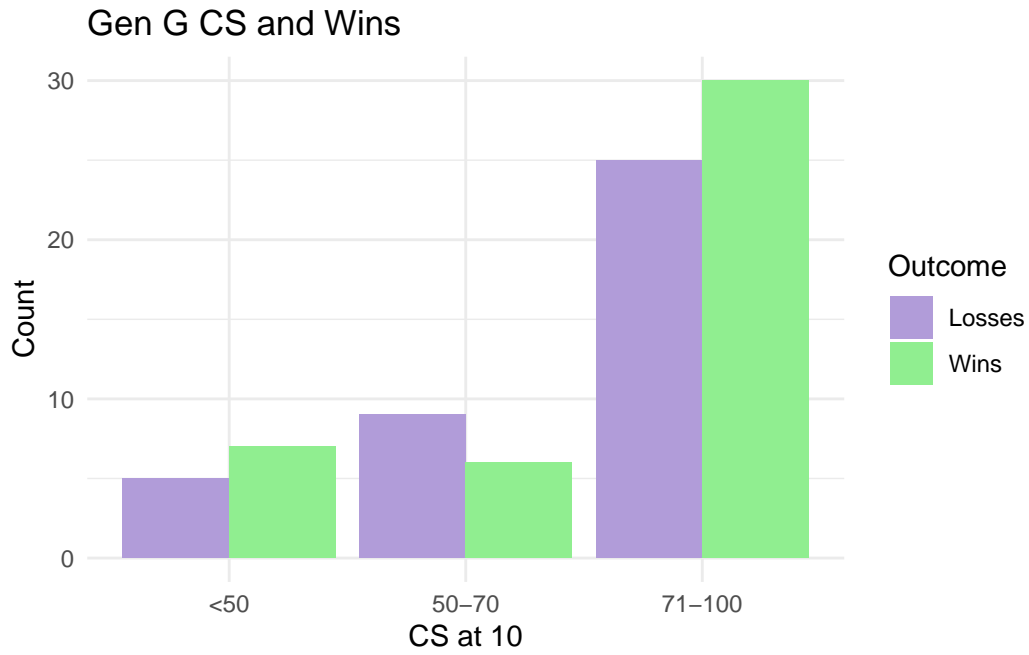
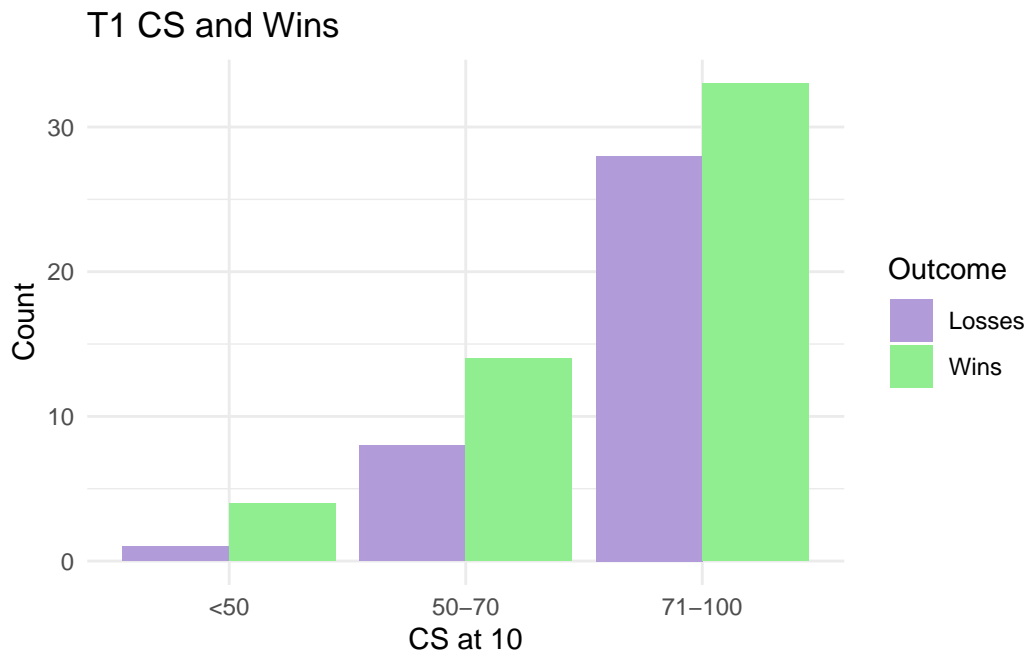Figure 3: Gen G's Laner's Win Rate with Different Creep Scores

Figure 4: T1's Laner's Win Rate with Different Creep Scores

3 Model

We use the Bayesian Analysis model to conduct this by defining a linear relationship between our outcome variable and our predictor variables. These predictor variables are then assigned a distribution and we use these to predict the outcome variable.

3.1 Model set-up

Define y_i as the outcome of the game and is 1 if the player wins the game and 0 if the person loses. Then CS_i is the player's creep score at 10 minutes and Level_i is the player's highest level lead against their lane opponent.

$$y_i | \pi_i \sim \text{Bern}(\pi_i) \tag{1}$$
$$\text{logit}(\pi_i) = \alpha + \beta_1 \times \text{CS}_i + \beta_2 \times \text{Level}_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

A logistic regression model was used in this analysis as the outcome variable is Binary. This is because we are considering whether they win the game or not. Whenever the outcome variable is Binary, a logistic regression model makes sense to help us understand it more. We used the default priors for our input variables.

A logistic regression model works by instead of considering an error value, which is done in linear regression, it considers a distribution for each of the inputs. The variability of these distributions inherently create the variability of the outcome which is associated to the error value we get in linear regression. The main advantage to this comes from the assumption made in linear regression models in which the error value is assumed to cancel out with each other(forming a normal distribution) In this case, consider the input of the cs at 10 minutes. In some games the players get more than their average and some games less, but they generally follow a normal distribution. The error is included by the variance of this normal. Assuming both of the inputs to follow a normal distribution makes sense because the number of games is high and this is the general trend of the data. (CITE FIGURE)

Each game can only result in a win or a loss so there is no other outcome that needs to be considered. For the inputs, we grouped the data into several categories as it is generally accepted that 100 cs in 10 minutes is a great creep score, anything above 70 is acceptable and anything under 50 is awful. For the max level lead we have not obtained any value greater than 5 as this is a massive level lead. The minimum for this is 0 as both players always start at level 1.

Another advantage of using a logistic regression model is that it does not require the independence of trials. This is great as burnout is a very real thing when playing league of legends. Players frequently go on both win streaks and loss streaks which would indicate that your recent games can impact your current one. This could easily make some of these trials dependent on some of each other but this won't affect the usability of the logistic regression model. We use the default priors in this model as default priors aren't very informative and therefore let the data about the games drive the model without much interference.

4 Results

Our results are summarized in Table 1 and Table 2. This table includes values for the intercept under various conditions as well as their error value.(The value in parentheses below each intercept.) Each value of the intercept indicates the estimated log-odds of the outocme variable. What this means is that the log of the outcome is given as the intercept. Therefore by taking the exponential function we can predict the likelihood of a certain outcome given a certain variable. For instance, we can see the intercept for T1's Max Level Lead of 2 being 1.37. Taking the exponential function of this gives a result of 4.0149. This means that in a random game, a T1 player is 4 times as likely to win than to lose if they have a level lead of 2.

The error value gives the standard deviation of these intercepts and we notice that this gets really high for the high extremes of level leads. This is probably because these level leads are an extremely rare occurrence and there aren't many games in which the model can use to accurately predict the outcome of a game with an extremely high level lead. In the dataset, both teams won all games with a level lead of 5 so even with a high standard deviation, the model predicts that T1 players are over a million times more likely to win a game where there are 5 levels ahead of their opponent.(A 5 level lead is massive, think Lebron James vs a 10th grader)

In comparing the two, the one that stands out the most is the max level lead of 2, in which Gen G are predicted to be 10 times as likely to win than lose whereas T1 is only 4 times as likely to win. This suggests that T1 is a lot worse at converting their leads into wins. The other intercept that stands out is the CS at 10'2 which refers to having a creep score between 50 and 70 at the 10 minute mark. The model predicts in this situation, T1 is 1.5 times as likely to win than lose whereas geng is predicted to be 4 times as likely to lose than to win. This likely indicates that T1 are better at playing chaotic games and do not need to have gold to create advantages. The second halves of the tables are not of concern.

Table 1: Logistic Regression Model Predicting Likelihood of T1 Winning based on Early Indicators

	T1 model
(Intercept)	−0.51
	(1.18)
CS at 10'2	0.39
	(0.65)
CS at 10'3	−0.07
	(0.55)
Max Level Lead'1	−0.15
	(1.26)
Max Level Lead'2	1.37
	(1.26)
Max Level Lead'3	7.65
	(3.95)
Max Level Lead'4	10.20
	(7.14)
Max Level Lead'5	13.96
	(11.10)
Num.Obs.	117
R2	0.260
Log.Lik.	−62.516
ELPD	−68.4
ELPD s.e.	4.6
LOOIC	136.8
LOOIC s.e.	9.2
WAIC	136.3
RMSE	0.43

Table 2: Logistic Regression Model Predicting Likelihood of GenG Winning based on Early Indicators

	Gen G model
(Intercept)	−0.41
	(1.19)
CS at 10'2	−1.38
	(0.75)
CS at 10'3	−0.95
	(0.53)
Max Level Lead'1	0.11
	(1.22)
Max Level Lead'2	2.30
	(1.23)
Max Level Lead'3	8.38
	(3.95)
Max Level Lead'4	18.76
	(15.08)
Num.Obs.	112
R2	0.328
Log.Lik.	−56.781
ELPD	−62.6
ELPD s.e.	5.4
LOOIC	125.2
LOOIC s.e.	10.9
WAIC	124.9
RMSE	0.41

5 Discussion

5.1 League of Legends

5.1.1 Player Roles

As mentioned previously, League of Legends has three lanes: top lane, mid lane, and bot lane. Usually, your tank goes in the top lane, which is the player that doesn't deal as much damage as the rest but takes a long time to kill. The mid lane has a mage who deals a different type of damage than the rest of their team and requires a different type of resistance to be nullified by the enemy team. The bot lane consists of the Carry and the Support. The carry is somewhat of a glass cannon, as they are very easy to kill but usually deal insane amounts of damage if left unchecked. The carry also usually follows a progression of being weak in the early stages of the game and strong in the late stages. This is why it is usually optimal to give the carry another player to support them, the support. The support's job is simple: babysit the carry and play for the team. The support champions are usually equipped with abilities that fulfill one of two roles: engage supports or enchanter supports. Enchanter supports buff the allied carry with a variety of shields, heals, and attack speed boosts. Engage supports have abilities that can lock down the enemy team and position them in such a way that it is easy for the carry to deal damage. The fifth player on the team is always the Jungler. They roam between lanes, and if any of the other four players on the enemy team are not careful, the Jungler can surprise them with a fight in which the advantage lies with the higher number of players.

5.1.2 Lanes

Each lane is filled with towers that deal abnormally high amounts of damage to the enemy, as well as lane minions that come in waves of 6 or 7 and are a simple aspect of the game that creates the most nuance out of any game mechanics. The towers only target the enemy if either of the following conditions are met: there are no enemy minions in the tower's range, or an enemy is dealing damage to an allied player who is under the tower. In this way, even if you are a lot weaker than the enemy, the tower provides a lot of safety. Of course, this does not make a player invulnerable, as coordinated teams can cycle the tower's targets by weaving in and out of tower range while trying to kill the player seeking protection from their tower. This is especially true at the highest levels of professional play where teams are extremely coordinated, which is the level this paper is investigating.

5.1.3 Lane Minions

The lane minions work in interesting ways. Firstly, experience (exp) is gained if an enemy minion dies close to you. Secondly, if you are the player that deals the killing blow (last hit) to an enemy minion, you gain a small injection of gold. At first glance, the amount seems

insignificant, but it adds up over time, and usually 20 more last hits than your lane opponent grants you a gold lead that is equivalent to killing them. Usually, the first concept beginner players looking to improve are told to work on is their last hitting. Not every minion is always worth last hitting, as sometimes, in the act of trying to last hit a minion, you expose yourself to deadly attacks from the enemy. A stronger enemy can use the threat of a deadly attack to prevent their lane opponent from getting last hits. In the top lane, it is also possible to deny your lane opponent exp by positioning very aggressively; however, this usually results in you not being able to secure all the last hits for yourself. The reason this doesn't work in mid is due to the shape of the map, in which the effective length of the lane is much shorter as towers are close to each other. Since the top lane is longer, one mistake from an unsuspecting opponent can usually lead to a kill, whereas in the mid lane, they lose health but usually not enough to kill them before they return to the safety of their tower. The bot lane, while the same length as the top lane, isn't usually able to zone the enemy from the exp range of lane minions as there are 2 players in each bot lane, so even with a strength lead, if you are less coordinated, you can lose the fight, and gold is usually worth more to a carry than exp. However, the bot lane has other options in which they can deny the enemy exp and gold. One way of doing this would be to stack a large minion wave and crash it under the tower. (No explanation of how to do this is provided as it is unnecessary, but essentially it means you have the lane minions from 3 or 4 different waves all under the enemy tower.) Then, they try to kill the enemy bot lane that is attempting to last hit the minions under the tower. Even if the tower kills you in the process, it is a net positive as you kill them so it evens out your death, but you also deny them 3-4 waves worth of gold and exp. Usually, when pro teams do this, they get assistance from their jungler to help kill under the tower. If they play it well and in the right situation, sometimes the enemy will die without trading back a kill.

5.1.4 Resources and How to Obtain Them

There are three main ways that you can become stronger throughout the game. The first is through gold, which allows you to buy items once in your base. The second would be experience, which lets you level up after a certain threshold is reached. Each level gives you more stats (health, attack speed, armor, etc.) Finally, you can kill neutral objectives (an enemy to both teams), and whichever team secures the final blow to this objective unlocks a buff for their team. Gold and experience are straightforward to obtain. Last hitting minions was already discussed earlier and is the primary way of obtaining both gold and exp. Killing an enemy also grants a large amount of gold and experience, but it also puts them on a respawn timer in which they cannot last hit any minions, so they go further into a gold and exp deficit depending on the situation in which they died. Finally, killing enemy towers and neutral objectives also grant large amounts of gold. The neutral objectives spawn in increments and get progressively more powerful, but there are some stacking bonuses, so playing for these in the early game is a very viable strategy. Of these three, the objectives usually have the least impact in the early game and do not provide a great indicator as to whether a team is ahead, as many times teams will choose to trade an early neutral objective to secure more last hits,

deny exp, or hit towers. As such, we do not consider the neutral objective statistics in our prediction of the match outcome.

5.1.5 Strategy

These gold and experience advantages make your player (your champion) stronger, allowing you to win fights more easily and contest neutral objectives in the late game. However, if you play imperfectly, you can easily lose a fight even with a gold lead; for instance, if you miss your abilities. There is also the concept of "scaling," in which some champions are a lot stronger in the late game but balanced by a weak early game. This sometimes means an early-game champion with a 60 cs lead (approximately 900 gold) may still be weaker than a late-game champion at 20 minutes into the game. A lot of newer players tend to overvalue the power of scaling, as even the most powerful late-game champions need to be executed with great precision to find success in the late game.

Another way a team at a gold deficit can win the game is by doing a "split push." A split push, as the name suggests, is one that splits the map by pushing waves on both sides (the top lane and the bot lane). For instance, consider a scenario where one team (team A) is stronger than the other (team B) in a fight with all 10 members. However, team B has a top laner that is much stronger than any one of team A's members. Team B can send their top laner to hit the enemy turrets in the top lane and their other 4 members to hit the enemy turrets in the bot lane. If team A only sends less than two players top lane, team B's top laner can easily take team A's top tower. If team A sends more than 1 player top lane, then team B can push the enemy bot lane towers as 4, and team A can only respond with 3. This favors team B in either scenario. While this is just a really surface-level analysis of a split push as there are many other factors to consider, it serves to show that through careful decision-making and precise play, the team that is behind on resources can still progress the game in their favor.

5.2 Implications of Findings

This paper analyzes the ability of T1 players and Gen G players to close out games after having certain leads. We consider two inputs: the cs at 10 minutes, which refers to how many enemy minions the player last hit before 10 minutes, and the highest level lead they had against their opponent. The maximum level lead helps the model predict a player's win rate of games in which they have a massive advantage over their opponent, as having a level lead of 2 or more is a significant advantage and usually means you also have a much higher cs lead. For this model, we do not consider the support players of either team, namely "Keria" and "Lehends," as it is usually better for support players to let the carry last hit the enemy minions instead. Also, sometimes they can leave their carry to hit minions alone and surprise the enemy mid laner. This means they get less exp even though their team gets a net positive play. Also, levels and gold are best used on champions that deal damage as this is the best way to increase the damage output, whereas supports that give shields or lock down enemy champions do not benefit that much from gold or exp. In fact, support players usually try to let their teammates get the kills because gold is least effective in the hands of a support.

The model predicts Gen G to be way more likely to win a game with a 2 level lead against a lane opponent compared to T1. This is arguably the most significant type of level lead as a 1 level lead can come due to random fluctations. For instance, if both players need to kill 7 minions to level up once, if you kill the 7th minion 1 second before your opponent, you technically had a 1 level lead. A 2 level lead is really significant as it means you denied your opponent massive amounts of exp and are much stronger than them. While it is true that a 3 level lead is even more powerful, it is so much more uncommon, which can be observed in Figure 1. This argument extends further to 4 level and 5 level. This disparity means that T1 is probably less clinical in the way in which they finish out games. They are unable to convert a significant level lead into a win at the same rate as Gen G can.

The flipside of this is that the model predicts T1 to be much more likely to win a game in which they have a creep score of 50-70 at 10 minutes. This can be explained by various factors. Firstly, this level of cs is considered low for good players, which means these games probably had lots of early fighting in which you are required to outplay your opponent as you have no gold or exp advantage yet. Maybe they are more skilled and therefore better at these early fights or maybe they know the ins and outs of the champions and therefore they know which champions can win early fights against other champions. This can also be explained by Gen G having a player called "Chovy" who is known for his insanely high cs levels. If a game occurs where he is not farming at these levels, it may be unfamiliar territory and therefore difficult to play. While it is inconclusive as to whether these are due to stylistic differences or sheer skill, it is clear that the players of two of the world's best teams don't receive the same benefit from the same advantage.

5.3 Weaknesses and Next Steps

5.3.1 Weakenesses

A major weakness of this analysis is that the the games under consideration were the solo queue games, with random teammates of a similar level. This version of league of legends has many differences to the professional play where team coordination is much more important than any individual player's skill level. These games were used as they are readily avaailable and generally speaking, the best professional players do really well in solo queue. Their solo queue characteristics are also apparent in their professional games. An example of this would Chovy, Gen G's mid laner. He is renowned for his extremely high cs and this shows up in both solo queue and professional matches. In this way, solo queue data are good approximations for analyzing certain aspects of player performance however the professional game data would be much more relevant for predicting professional games. Another weakness is the relatively small dataset. These players play thousands of games every year and we can analyze all of them to get a better understanding of how they play without cyclical shifts such as meta changes.

5.3.2 Next Steps

A possible next step would be to obtain a product key from riot games which allows for more api requests and therefore we can analyze data from thousands of games at a time. Another potential next step would be to look at different input parameters as predictors of the match outcome. Maybe early neutral objectives as well as the type of champion selected as these both speak to the strength of a team at a certain point in time. These are all available on from the apis however there is no available data from a primary source about the professional games.
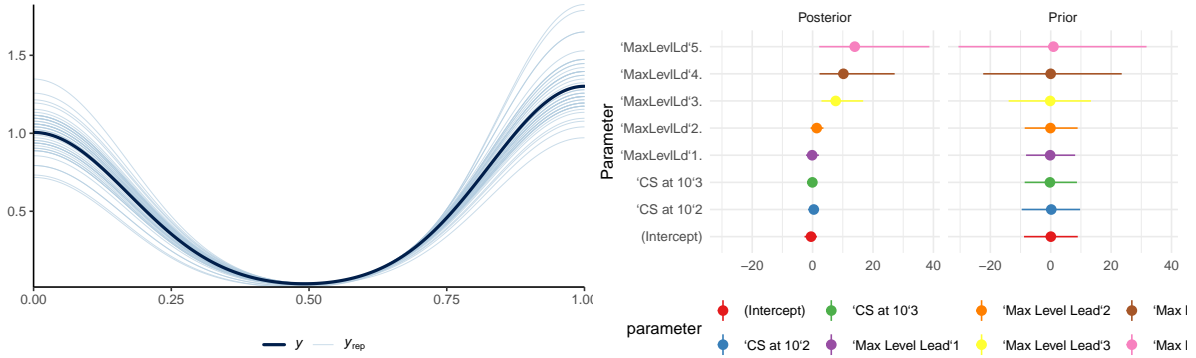
Appendix

A Model details

A.1 Posterior predictive check

In Figure 5a we implement a posterior predictive check. This shows the outcome variable with simulated variable from the posterior distribution.

In Figure 5b we compare the posterior with the prior. This shows how much the estimates change once the data is taken into account.

(a) Posterior prediction check   (b) Comparing the posterior with the prior

Figure 5: Examining how the model fits, and is affected by, the data

In Figure 6a we implement a posterior predictive check. This shows the outcome variable with simulated variable from the posterior distribution.

In Figure 6b we compare the posterior with the prior. This shows how much the estimates change once the data is taken into account.

A.2 Diagnostics

Figure 7a is a trace plot. It shows lines that appear to bounce around and are horizontal. This suggests nothing wrong with the model we used.

Figure 7b is a Rhat plot. It shows every value close to 1 and nothing greater than 1.1 This suggests the model we used is fine.

Figure 8a is a trace plot. It shows lines that appear to bounce around and are horizontal. This suggests nothing wrong with the model we used.
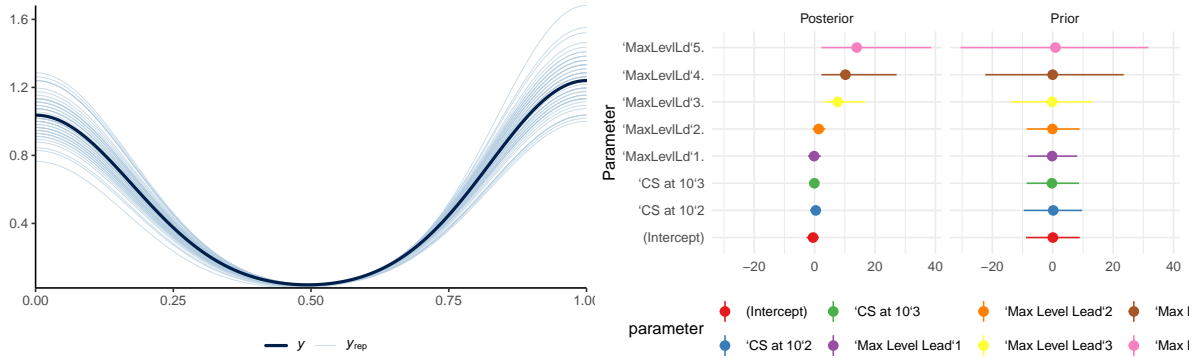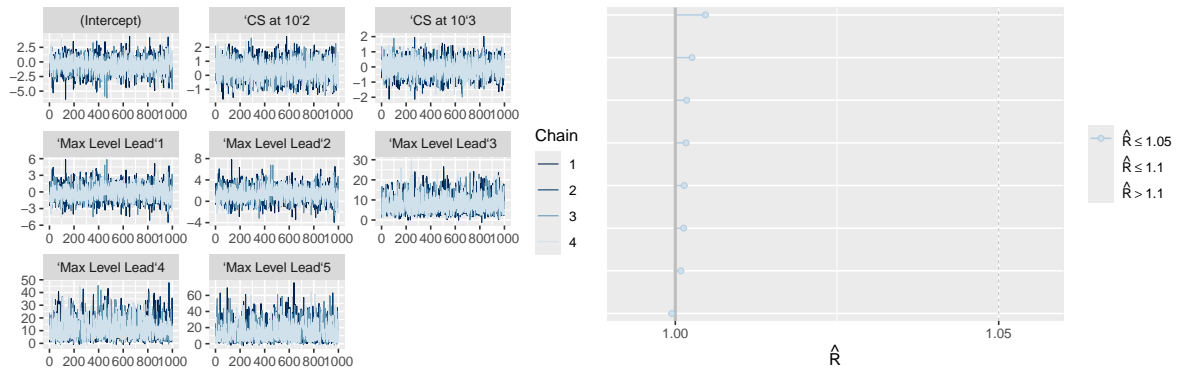
(a) Posterior prediction check

(b) Comparing the posterior with the prior

Figure 6: Examining how the Gen G model fits, and is affected by, the data

(a) Trace plot

(b) Rhat plot

Figure 7: Checking the convergence of the MCMC algorithm for the T1 model

Figure 8b is a Rhat plot. It shows every value close to 1 and nothing greater than 1.1 This suggests the model we used is fine.
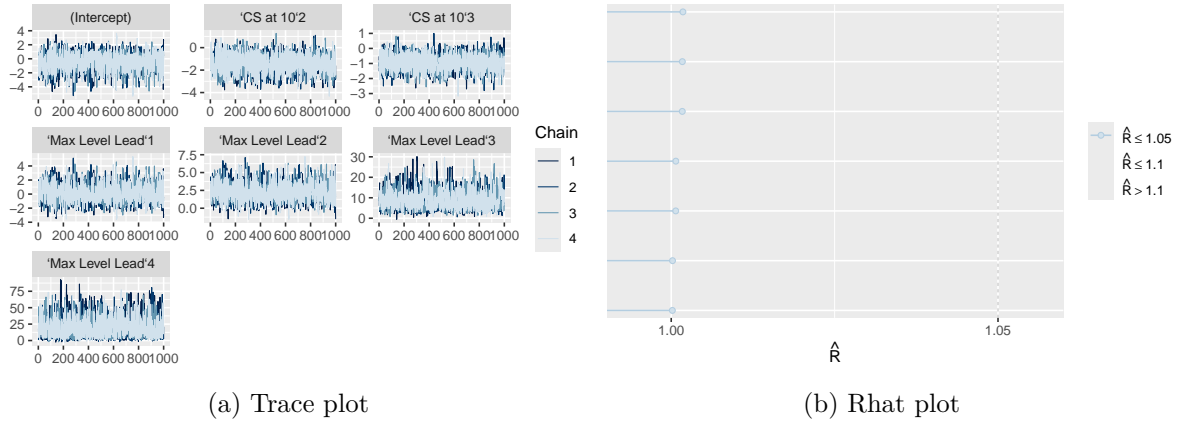

(a) Trace plot

(b) Rhat plot

Figure 8: Checking the convergence of the MCMC algorithm for the GenG model

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.