Byungjin Kim
Kyle Zhou
Adam Hoerger

Introduction/Background: We have decided to pursue looking into a League of Legends dataset. Those already familiar with the game can likely skip this first paragraph. League of Legends is an online MOBA (multiplayer online battle arena) game that consists of two teams, distinguished by the colors red and blue. One team will start from the blue side of the map and vice versa for the remaining team. There are 5 players on each team, each filling a specific role within the game: Top, Jungle, Mid, Bot (ADC), and Support. Each player then focuses on killing minions that spawn every 30 seconds, which is a source of income (gold). Players can also earn gold by killing an opponent character (aka champion). The primary objective of each team is to destroy various structures belonging to the enemy team, with victory being achieved once a structure termed the "nexus" is destroyed. Windows of opportunity to freely attack these structures can be obtained either by superior map positioning or by reducing the health of your opponents' champions to zero, which will remove them from the map for a period of time. Which team holds the advantage at any point in the game is typically dictated by the relative strengths of the various champions, the items (bought with gold) held by each, and the skill of the players involved.

While a demonstration of individual skill is part of what makes League of Legends a viable esport, the various strategies employed by teams adds further depth to competition. A "meta" is a term used to describe a collection of strategies, many of which revolve around choice of champions, items, or playstyle, which are widely utilized by the general community or competitive scene at a given time. For example, a champion, or even a combination of several champions, that is picked in a large proportion of games would be considered to be "meta" or "in the meta." As the developers of the game add or change content in the game, the meta evolves to adapt to these changes. Even at the same point in time, different regions may have differing philosophies on what is most effective, resulting in different metas across regions.

Question of interest: We want to be able to characterize the various metas that have emerged across different regions of competitive League of Legends, exploring how metas vary across regions and try to pull out trends for meta performance and recurrence over time.

Dataset: This dataset includes all competitive matches between 2015 to 2017 for League of Legends. The dataset is limited to match history from 5 of the competitive leagues: North America, Europe, South Korea, Taiwan, and Brazil. We're given all of the information about each match such as the winning team, the player who got a kill, the player who got killed, the amount of gold difference between the teams*, etc.. This information is split into individual csv files, requiring us to use the "Address" column in order to relate all the information with specific

games. The “Address” column, besides allowing us to match specific games between csv files, holds the website address where the match information is from, meaning that we can look up specific matches in our browser based on the address. By utilizing the address column, we can also look to create our own database/dataframe by querying the information we specifically want to look at from the various tables, making it easier for us to use the algorithms and methods learned in class.

One downfall of this dataset is that we would have liked to have a more complete view of the final statistics for the match. The dataset is missing things like final kill count and gold amount for each team, which is unfortunate since this information can help quantify how the metas stack against each other. This can be remedied by putting in a bit more work into the data cleaning. The kills are listed individually: for each instance of a kill, the dataset lists out the killed player, the player who killed, and players who assisted with the kill. As a part of our cleaning step, we can go through all the kills in a given match to figure out how many total kills happened. A similar thing can be said for gold, as the dataset shows gold per role for each team per minute. We would just need to take the last entry in each array for the gold per minute for each player then total it for the team’s total gold.

Another downfall would be the age of the dataset. Unfortunately we could not find an up to date data dump of the competitive scene which would allow for more relevant exploration and analysis. If we really wanted to, we could apply for access to Riot Games’ developer API, but we were unsure of how long/tedious that process would be.

- Link to dataset: https://www.kaggle.com/chuckephron/leagueoflegends
- *Note: The gold difference is relative to the team on the blue side, meaning negative gold difference would represent the team on the red side having more overall gold than the blue side team and vice versa for positive gold difference.

Methods and hypothesis: After cleaning the data set and determining what parameters differentiate one meta from the other, we can separate each game into two different points of data, one for each team, and generate coordinates based on the aforementioned parameters. We then intend on leveraging higher dimensional clustering to try to detect and quantify the occurrence of different metas within the dataset, testing various methods to see which gives the most interesting results. Once we have our meta clusters defined, we can analyze their spread across regions; potentially being able to characterize the play style of each of them. We can then do some more interesting things with the clusters such as ranking them in terms of performance, using market basket analysis to determine which metas/champions tend to appear alongside each other, or visualizing how the competitive scene has changed over time. Then we can make some pleasing infographics, post on r/leagueoflegends and reap that sweet sweet karma ezClap.