

DOTA 2 Gameplay Project

Tao Cheng and Sijia Chen



Fig. 1. DOTA2 International Tournament in 2018

Abstract—An analysis of the balance of Dota 2 gameplay. The game Dota 2 is a famous in the world. There are many players found that the two factions in the game are not balanced. With the help of visualization, we will find the reason why and what caused such an imbalance, and then present those reasons with some visual diagrams to help people understand how and where may cause the imbalance. Lastly, we will try to find a way to solve the problem by analyzing the data we found

1 INTRODUCTION

1.1 Topic Background

Dota 2 is a multiplayer online battle arena (MOBA) video game developed and published by Valve. This game has two teams of five players compete to collectively destroy a large structure defended by the opposing team ('Radiant' and 'Dire'). This game is presented on a single map in a three-dimensional isometric perspective. Every player controls one character, which we called 'heroes'.

1.2 Project overview

The project is about the balance of the DOTA2 game. We collected around 90,000 gameplay stats, from January to March, 2019, to analyze the ratio of two different sides ("radiant" and "dire"). After working with the gameplay stats, we found that the win ratio of the radiant is around 52%, while the win ratio of the dire side is 48%. This means the balance of the game is "broken". Therefore, my team is looking for what factors or attributes can cause this problem, and we want to show which part can improve the game by using data visualization.



Fig. 2. Win Ratio Presents by Sketch

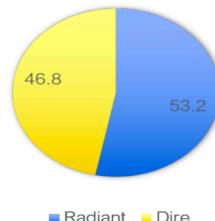


Fig. 3. Win Ratio Presents by pie chart

1.3 Research question

Firstly, our team is aim to explore the win rate between the two factions "Radiant" and "Dire". Next, we try to use the visualization to show why there is such a difference. For examples, the visualization may represent different 'hero' picks, different lane choices, and different play styles etc. After we found the result and visualization, we would like to see if there is a way to balance the game. Finally, we want to know whether the research results will compound our expectations.

1.4 Motivation

As the history of video game evolved, e-sports became increasingly professional. And DOTA is one of the great contributions to e-sports. Also, DOTA is a game that accompanied many young people through their childhood. DOTA2 is the inheritance of DOTA, it witnessed the growth of many people as well. As we proceed with the project, we found that DOTA2 is still one of the most popular game till now. However, the number of people who play this game is showing a downward trend. We want to figure out what factors cause

this situation, because it not only may save many ‘old school’ games life, but also we can find the rule of the life cycle of many games. After we finished the project, we hope to see that we can help more DOTA2 players to enjoy the game.

2 DATA & DATA QUESTIONS

We used the dataset we found on Kaggle.com, link: <https://www.kaggle.com/magnundenis/dotaset>. There are 90,614 rows by 42 columns. It is relational and number series data. From the dataset, we know the win ratio of the two fractions "Radiant" and "Dire". Then, according to the mini map and other data, we found that there are multiple reasons that would affect the result of a game. We would like to use the data based on the games to know if the win ratio really affect the game balance. Also, which part of the data attributes can influence the result most? Our team would like to try to see if there is a way that we can fix the equality of the game. Finally, we will present the results of the research in a visualization way to help people understand the research we did and find if there is something, we missed or wrong.

3 RELATED WORK

3.1 Other Solutions

We try to use curve line graph to represent the win ratio for both sides from the gameplay stats. It will show that every detailed change from January to March, 2019. However, this visualization can only solve the question that the game is imbalanced. We try to figure out that what is the main factor causes this result. Probably, we still need more visualization such as bar chart or line chart to illustrate some detail part and exact numerical value. After that, we can have the answer that which element “ruined” the game. Thus, we will provide our solution to improve the game balance by showing the data visualization.

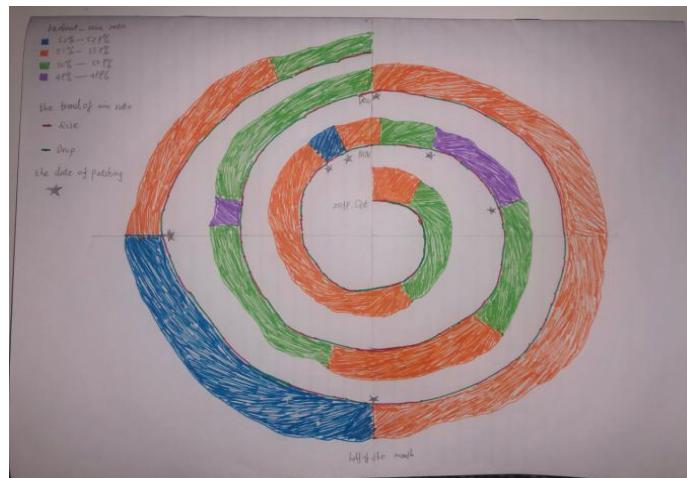


Fig. 4. Changing of Win Ratio by Sketch

3.2 Related Work

As we working with the project, we found there is another data project relevant ours. It is named “Predição de vitória em Dota2” from Kaggle. That project shows the same result, namely “Radiant” side got a higher win ratio, by using bar chart.

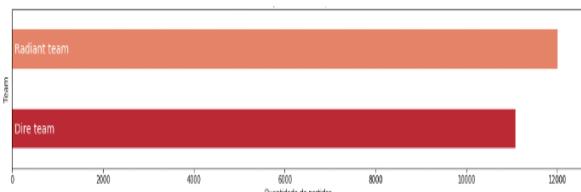


Fig. 5. ‘Radiant’ and ‘Dire’ Win Ratio via Kaggle

His project is focusing on the prediction of the game win. He analyzed the data of thousands of games to get the attributes and conditions for predicting the outcome of the game. This also helps us save time to filter less important data, and lead us finding the most useful attributes to complete our own project. As we observed, we found two conditions can be our objectives, namely ‘hero picks’ and ‘map designs’. According to his result, there are top 15 heroes have high win ratio through the whole gameplay stats. This is a useful information that we can take “hero” as one of our potential elements which will cause imbalance in our project. However, we still want to dig more data to prove whether the map location will influence the game or not.

4 DESIGN JUSTIFICATION

4.1 Choice of visual encodings

We plan to use Python to make the data visualization, because we have experienced some python coding projects before. And this will help us with designing these visualizations.

4.2 Layout

For the first visualization, we want to show a pie chart for comparing two teams’ win ration, because it will be easier to figure out. For the second one, we think adding dot and area to a mini map of the DOTA2 image will be the best choice. This can make readers know the map layout. And we can point out the important data within the map. The readers can know what is the highlight of the visualization. For the last one, we want to show the most heroes picking in the game. As we know, the heroes have their unique name in the game. Therefore, we want to show the name in the visualization.

4.3 Alternatives Discussion

For the second visualization, it is like the mixture of the scatters using and map using. As we read the article written by Karl Sluis, his viewpoint is there are three levels of the organization of a visualization namely quantitative, ordered, and qualitative. So we could not just use scatters for representing the value of the data, but also some bars or curved lines can also show the value. Another thing we want to point out is we use massive color elements to show the differences. However, in Robert Kosara’s article called ‘How the Rainbow Color Map Misleads’, he believes that we often seem to be content with default colors or with an arbitrary selection that we prefer. But we may cause more damage to our visualization than what we thought. Therefore, the size of the marks, or the shape of the marks can also help us to complete the visualization that focuses on comparing.

5 IMPLEMENTATION

```
In [66]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as dates
import seaborn as sns
import random
import warnings
from wordcloud import WordCloud, STOPWORDS
from matplotlib.colors import ListedColormap
import sklearn
import xgboost as xgb

from sklearn.linear_model import LinearRegression, LogisticRegression
from sklearn.svm import SVR
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import RandomizedSearchCV
from sklearn.metrics import classification_report
import scipy.stats as stats
import statistics as stat

import os
print(os.listdir("."))

['player_time.csv', 'heroes.csv', 'test_players.csv', 'yasp_sample.json', 'teamights_players.csv', 'item_ids.csv',
 'DS_Store', 'test_labels.csv', 'chat.csv', 'picks_ban.csv', 'ability_upgrades.csv', 'matches.csv', 'purchase_log.csv',
 'match.csv', 'chatter_regions.csv', 'heroes_pick.png', 'player_matches.csv', 'Seng474_Project.ipynb',
 'players.csv', 'project_pic', 'hero_names.csv', 'ability_ids.csv', 'match_outcomes.csv', 'player_ratings.csv',
 '.ipynb_checkpoints', 'match_patch.csv', 'teamfights.csv', 'objectives.csv', 'patch_dates.csv']
```

Fig. 6. Import the Datasets

First of all, we import the DOTA2 game stats dataset, which we downloaded from Kaggle, into the Jupyter Notebook by using

python and some additional packages. These packages are includes Numpy, Pandas, and Matplotlib. Numpy can help with adding support for large, multi-dimensional arrays and matrices. For example, we transfer the game data arrays to matrices. Pandas offers data structure and operations for manipulating numerical tables and time series. Matplotlib is a plotting library for Python. With the help of these tools, we were able to convert the digitized data into visualized one.

However, we did face some challenges when convert the numerical data to visual data. For example, we are not familiar to using Matplotlib package. This means we spent much time on self-learning the tools. Also, the charts are too ‘simple’ to treat as fantastic diagrams. So we have reviewed tons of examples on website for improving ours.

6 DATA INSIGHTS

Firstly, as we can see the chart, the higher the score gained, the better the chance of winning game. However, this is not always true, because we can see some special situation has shown in the diagram.

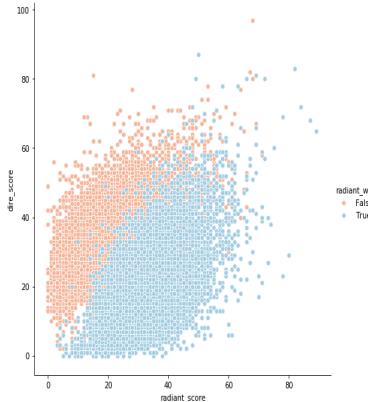


Fig. 7. Win Ratio by Scores

Secondly, this visualization shows some the most popular heroes. The scale and position of the name of the hero is represents the popularity. We are surprised that ‘Sand King’ and ‘Rubick’ are so big. We think this visualization is the most intuitive to see the data in it.



Fig. 8. Most Popular Heroes’ Name in Game

7 RESEARCH FINDINGS

Again, our research questions revolve around two main objects. First, if the map designing influence the balance of the game via three different lanes. Second, if there exist some heroes with absolute advantages through the game.

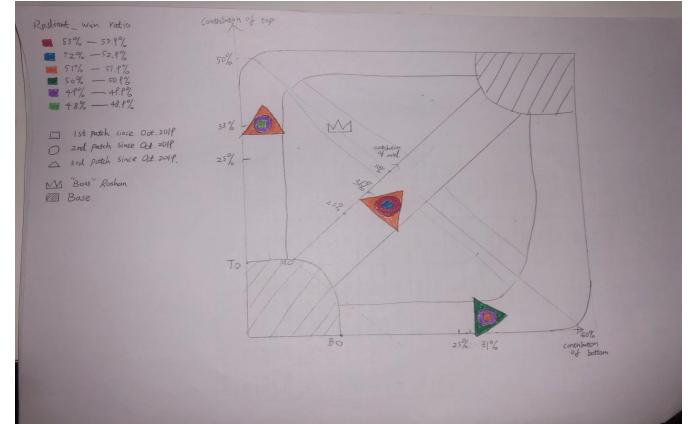


Fig. 9. Win ratio for different lane in 3 versions by sketch

For the first question, we have already known that ‘Radiant’ has higher win ratio. But we want to find out which lane (top lane, mid lane, bot lane) has the greatest impact on win ratio, so we know whether the map designing will also affect it. We used three versions of the data. It was found that almost every version of mid lane has the highest win ratio.

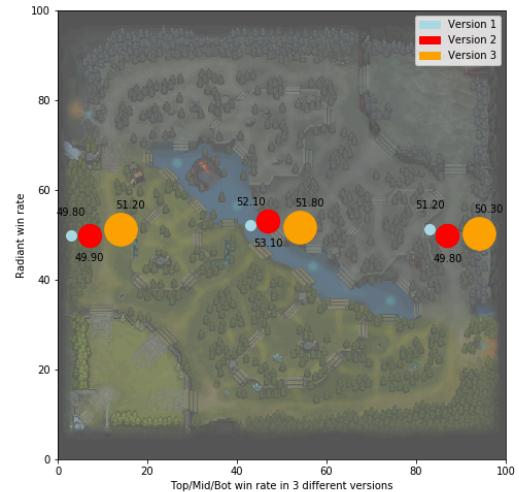


Fig. 10. Win Ratio for Different Lanes in Three versions

What surprised us is that the highest win ratio fluctuation is the top lane, followed by the bottom lane. In fact, no matter what version, the map of mid lane is almost unchanged. However, every time a major version is updated, the maps on the top and bot lane will be adjusted. This visualization just proved that the change of the map design indeed causes a certain impact, which achieved the effect of changing the balance of the game.

For the second one, we built the visualization for top ten heroes with the highest win ratio all the time.

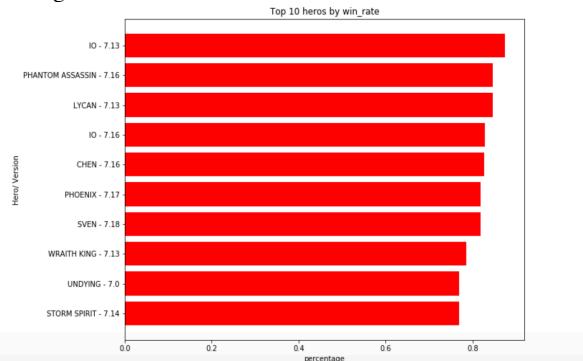


Fig. 11. Top 10 Win Ratio Heroes before 2020 Versions

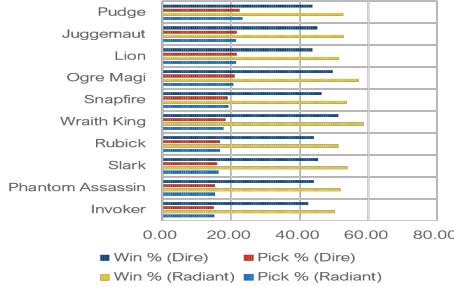


Fig. 12. Top 10 Win Ratio Heroes after 2020 Versions

We can clearly see that the hero with the highest win rate in game history can achieve a win rate of over 90%. And the win ratio of 7 of them is more than 80%. For a multiplayer competitive game that pursues fairness, such data is very shocking. Because of this visualization, we can be more convinced that our approach is in the right direction. That is to choose some specific heroes will definitely make this game unbalanced.

8 DISCUSSION

For the strengths, firstly, because our dataset comes from the official website game data, so with the data updated, our visualization can be updated at the same time. For example, we just need to download the latest version gameplay data, then import them into the same function we have. The other advantage is Python can import so many packages for visualizing the data. Because it has such potential, as our experience increases and coding skill gets better, we can create more innovations and ideas on the basis of the original. This means we can keep developing the project.

For the weaknesses, to be honest, our visualizations are not really charming comparing to the visualizations designed by D3 or other professional tool. For example, the second sketch (fig. 3) is the ideal visualization for showing the change of the win ratio in recent game versions. We spent too much time on how to designed the spiral plot in python or D3. However, we failed to build a diagram like that. Another regret is that we wanted to make a visualization that can contain all the information (such as 2 team's win ratio, 3 lane's win ratio, and specific heroes' win ration). But we found that the visualization would be messy and some data would present incorrect.

When we searched for articles related to our project on the websites, we found that many people focused on predicting the outcome of the game, rather than simple analyzing the balance of the game. Perhaps when predicting the game, we can indirectly find the imbalance in the game. Therefore, we can also include the prediction of the game result in future project, which can improve the balancing of the game as well.

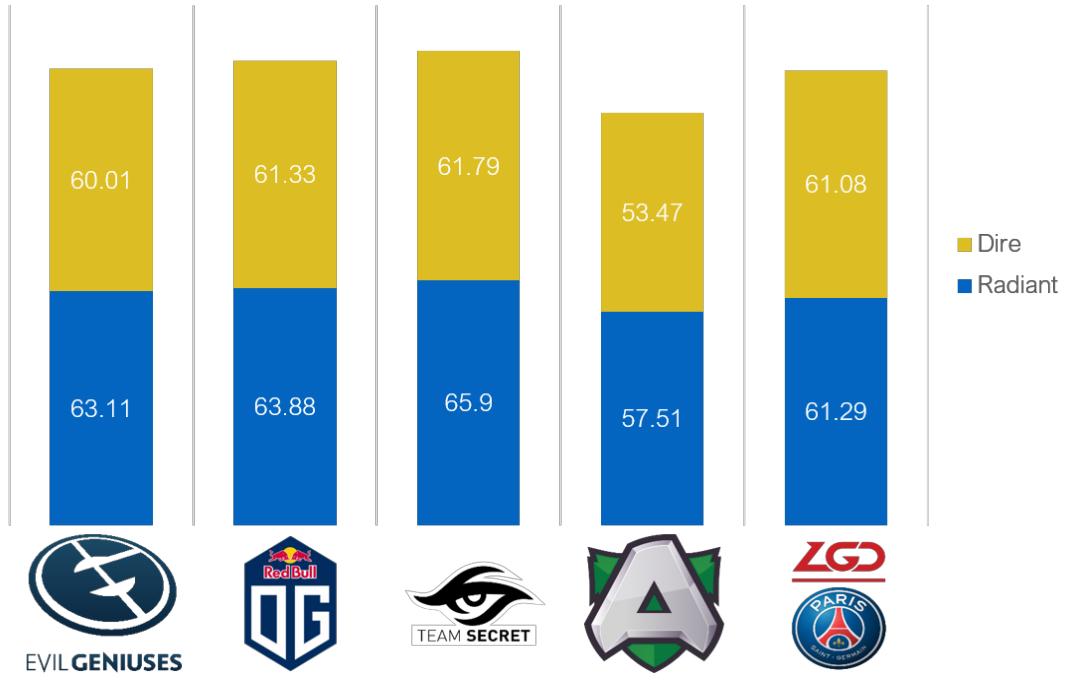
- [2] K. Sluis. *Before Tufte, there was Bertin*. Jacques Bertin and the Graphic Sign-System. February, 2016. (URL: <https://medium.com/@karlsluis/before-tufte-there-was-bertin-63af71ceaa62>)
- [3] R. Kosara. *How The Rainbow Color Map Misleads*. eagereyes. July, 2013.(URL: <https://eagereyes.org/basics/rainbow-color-map>)
- [4] N.P.Rougier, M. Droettboom, P.E. Bourne. *Ten simple Rules for Better Figures*. Plos Computational Biology. September, 2014.

REFERENCES

- [1] Magnun. Predição de vitória em Dota2. Kaggle. June, 2019

1. Sketch 1

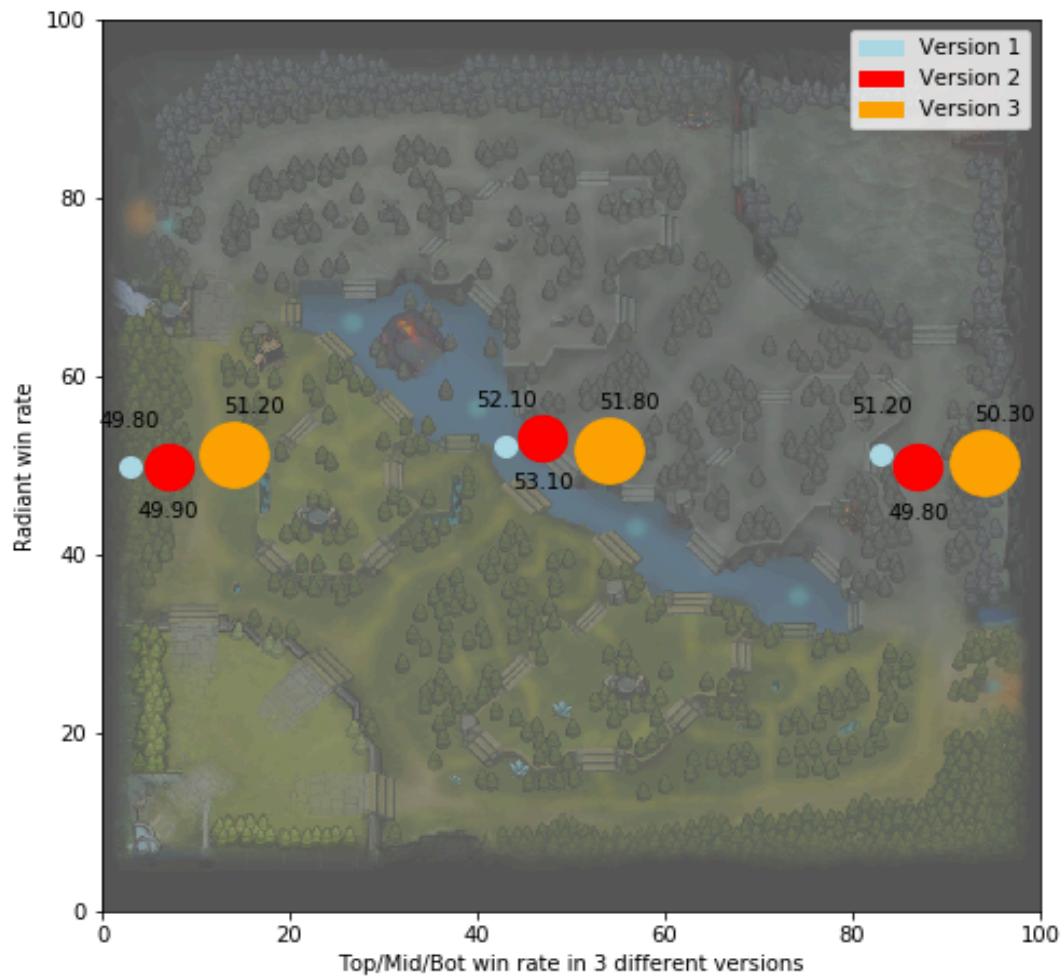
Number of words: [38] / 100



This visualization shows top 5 DOTA2 professional Team's win ratio. The yellow bars represent the win ratio when they are in team 'Dire', while the blue bars illustrate the win ratio when they are in team 'Radiant'.

2. Sketch 2

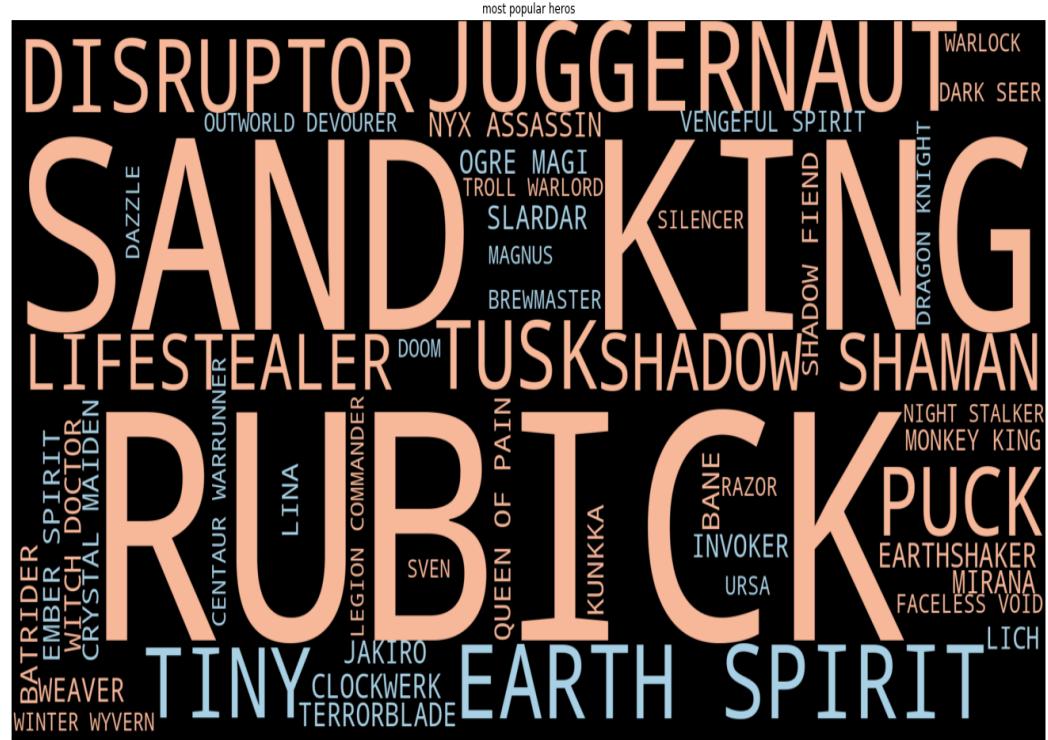
Number of words: [46] / 100



This is the visualization for the team 'Radiant' win ratio in last three versions. Three different position represent top, mid, and bot lane players' win ratio. And three different colors and size represent three different versions win ratio data. And the background is the DOTA2 minimap.

3. Sketch 3

Number of words: [45] / 100



This visualization is showing the most popular heroes in DOTA2. The larger the hero's name, the more times the hero is selected. And the name is red color means the hero win ratio is over 50%, while the blue color means the hero win ratio is under 50%.