IN4086: Data Visualization

Analyzing and Visualizing Player Behavior in DOTA

Ka-Ping Wan   Carlo van der Valk

4111443   1308378

December 16, 2015

1 Introduction

This assignment is about visualizing data originated from DOTA2. DOTA2 is a multi-player online battle arena game, where two teams fight against each other in an arena environment. The data contains information about the positions of the players, zone of the map and the inter-team distance.

2 Goals

The main goal of the assignment is to analyze the data and visualize the results to give the players valuable insights. We have identified 4 sub-goals that will support this.

2.1 Movement

Knowing how each player move is important. One goal is to show the movement of each player over time. It is also valuable to know his or her walking path together with frequently visited points in the map. This will help in understanding their own strategy, but it is also useful for the opponents in plotting their strategy.

2.2 Inter-team distance over time

The data also contains information about the inter-team distance. It would be interesting to see how the distance changes during a match and if there is correlation between the distance and match victories.

2.3 Opening strategies

Team members have to work together to win using some kind of strategy. The beginning of a game is important. Therefore, players have to perform some kind of opening

strategy. Seeing which winning opening strategies were used in games can help players improve their own strategy.

2.4 Zone changes

During a match of DOTA2, players might frequently move across the map going from zone to zone. Looking at how frequent players change from zones during different stages of a match and showing from which zones players move might provide useful insight.

3 Results

Below are some screenshots of each visualization and a brief explanation of the design choices.

3.1 Movement
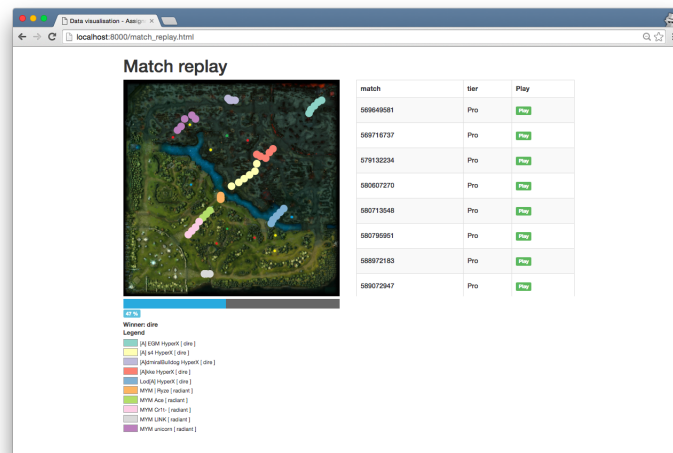
For the movement we first listed all available matches on the right side. By clicking on "play" the movement of each player will be displayed on the map. Each dot represents a player and has a distinct color. Together with the legend, each player can accurately re-watch his movement.

3.2 Inter-team distance over time
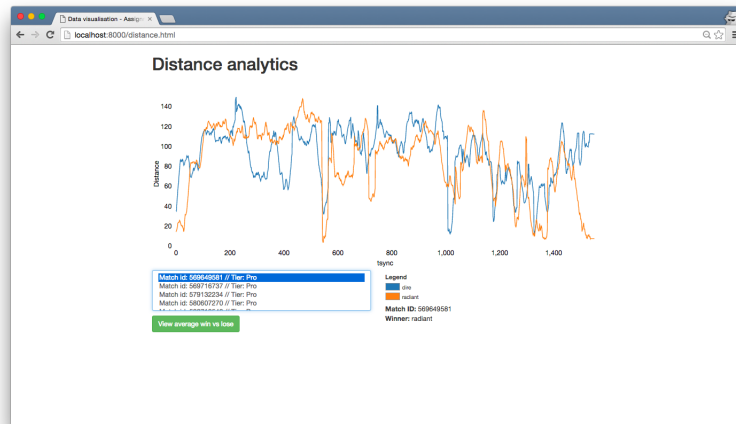
For the distance we have graph the inter-team distance over time for each team. By giving each team a distinct color, the visualization can easily be used to compare the distance of each team.

3.3 Opening strategies
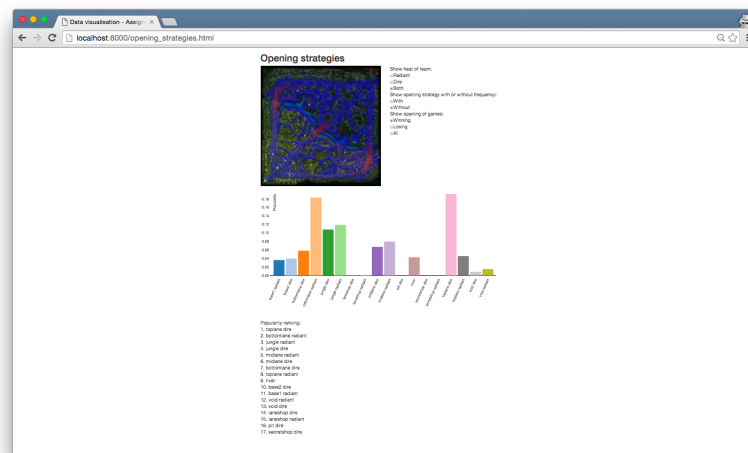
For the opening strategies we show a heatmap along with a bar chart showing the zone popularity of the first 10% of the game. With the heat map and bar chart players can easily see with places and zones of the map are used in opening strategies.

3.4 Zone changes
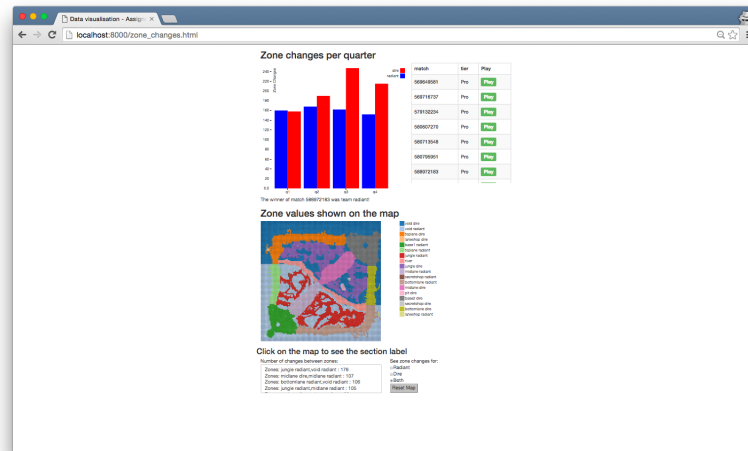
For the zone changes we used a grouped bar chart to display the number of zone changes per quarter of the match. The grouped bar chart was used so that the number of zone changes for each team could be compared with ease. Also displayed is a map of the game with the different zones marked by distinct colors. This map has a legend for the zone colors, also a point on the map can be clicked and zone name will be displayed. Below the map is shown a list of zone changes that occurred along with the number of times it occurred. These can be selected so that the corresponding zones will be highlighted on the map. With this chart and map players can see clearly the significance of zone changes.

4 Technical restrictions and instructions

Below are some remarks about the project concerning technical restrictions and running the project itself.

4.1 Libraries

For the project we used D3.js for the data visualization, jQuery for a few DOM-manipulation and Bootstrap CSS for some basic styling and document markup. The NumPy package was used in the Python preprocessing code.

4.2 Data pre-processing

The original dataset was too large to process at once. Therefore we have split the `master-zones.csv` and `master-distance.csv` in separate match files:

- `/preprocessing/available_matches.csv` list the matchID and tier of the available matches.

- `/preprocessing/master-zones` contains files from `master-zones.csv` separated based on the matchID.

- `/preprocessing/master-distance` contains files for each matchID.

- `/preprocessing/heatmaps` list amount of activity on a position of the map aggregated from all the matches in the dataset. There are two types of heatmaps; one heatmap where the amount of times a point on the map is visited within a single game also counts towards the heat(we refer to this as "with frequency"), and another heatmap where the visitation of a point on the map is only counted once per game(without frequency).

- `/preprocessing/popularity` list the normalized popularity of each zone. Here using the section labels of each point of the map and the data from the heat map, the popularity of each zone is obtained. Because the heatmaps have two variations, and are used to calculate the popularity, there are also two variations of the popularity list.

Due to the large size of the data we have limited the total match to 15. In order to recreate the pre-processed data do the following:

- Place the `master-zones.csv` and `master-distance.csv` in the `original_data` folder.

- Run the python and R scripts located in `python scripts` and `R scripts` respectively.

4.3 Running the project

In order to easily run the project, open `Terminal(Mac)` or `PowerShell(Windows)` and change the directory to the project. Start a python server with the following command: `python -m SimpleHTTPServer`. The project should be live at `http://localhost:8000`.