



Age of Empires 2: Exploratory Data Analysis

Alana Willis, Raj Dasani



The Game

History:

- Released in 1999 by Microsoft.
- Can be played on Windows, Mac, and PS2.
- Game peaked in April 2020, reaching 60k for their average player count.

Concept:

Strategy game where players gather resources in order to build their towns and prepare for battle by building up their armies and attacking other rival bases and towns.

Variables

1. **Profile_id**: unique number for each player account
2. **Name**: player username
3. **Rank**
4. **Rating**
5. **Country**
6. **Games**
7. **Wins**
8. **Losses**: includes games dropped
9. **Drops**: lost connections or forfeits
10. **Game_type**: 1v1 or Teams

6,539 Observations

Structure of the Data Set

Important Notes

- 10 games played per game type is required to earn an official ranking.
- Team ratings and 1v1 rating are ranked separately.

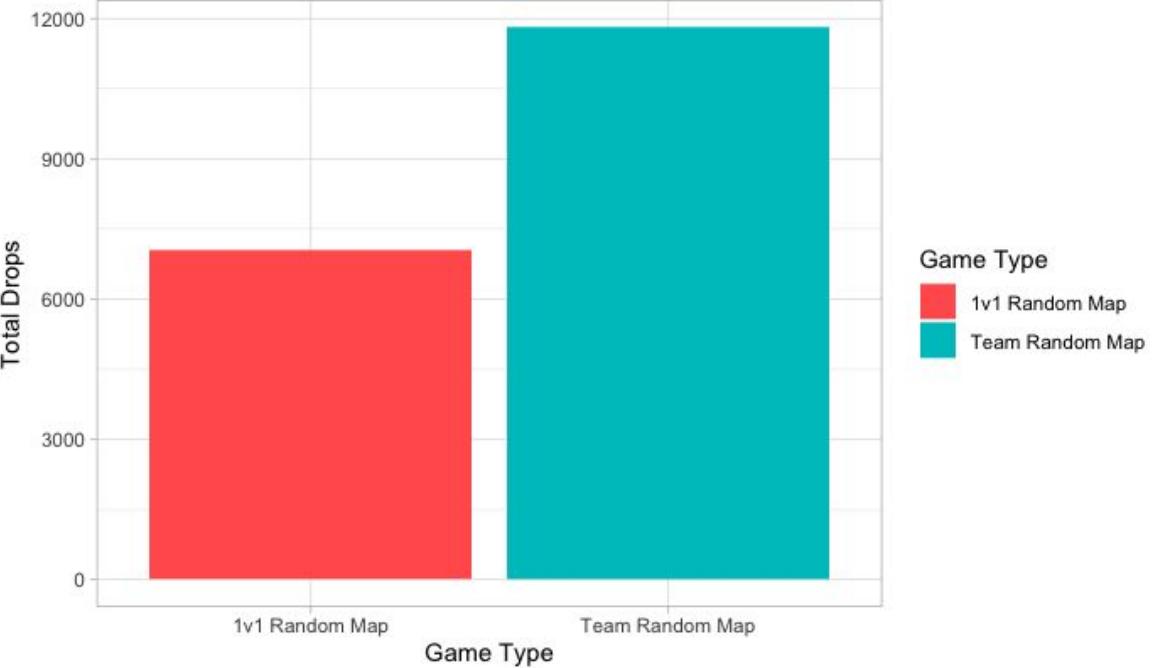
Teams vs lvl - Drops

Hypothesis I:

A player is more likely to get dropped from a Team Match.

Drops

Amount of Drops per Game Type



Thought Process:

Visualization shows that more drops occur during Team Matches.

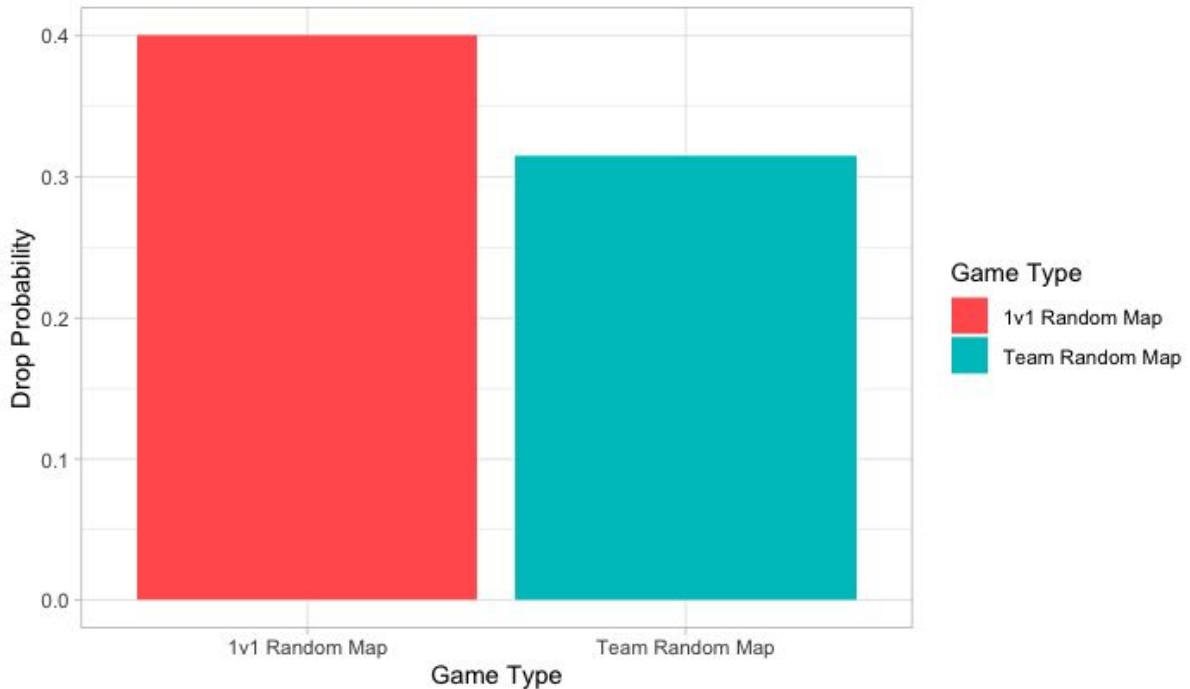
1v1 Random Map: 7037
Team Random Map: 11805

However, this wasn't enough...

Drops

Conclusion:

The Probability of Drops



Our hypothesis proved to be incorrect.

Although there are more drops in Team Matches, the probability shows something different.

1v1 Random Map: 32%
Team Random Map: 40%

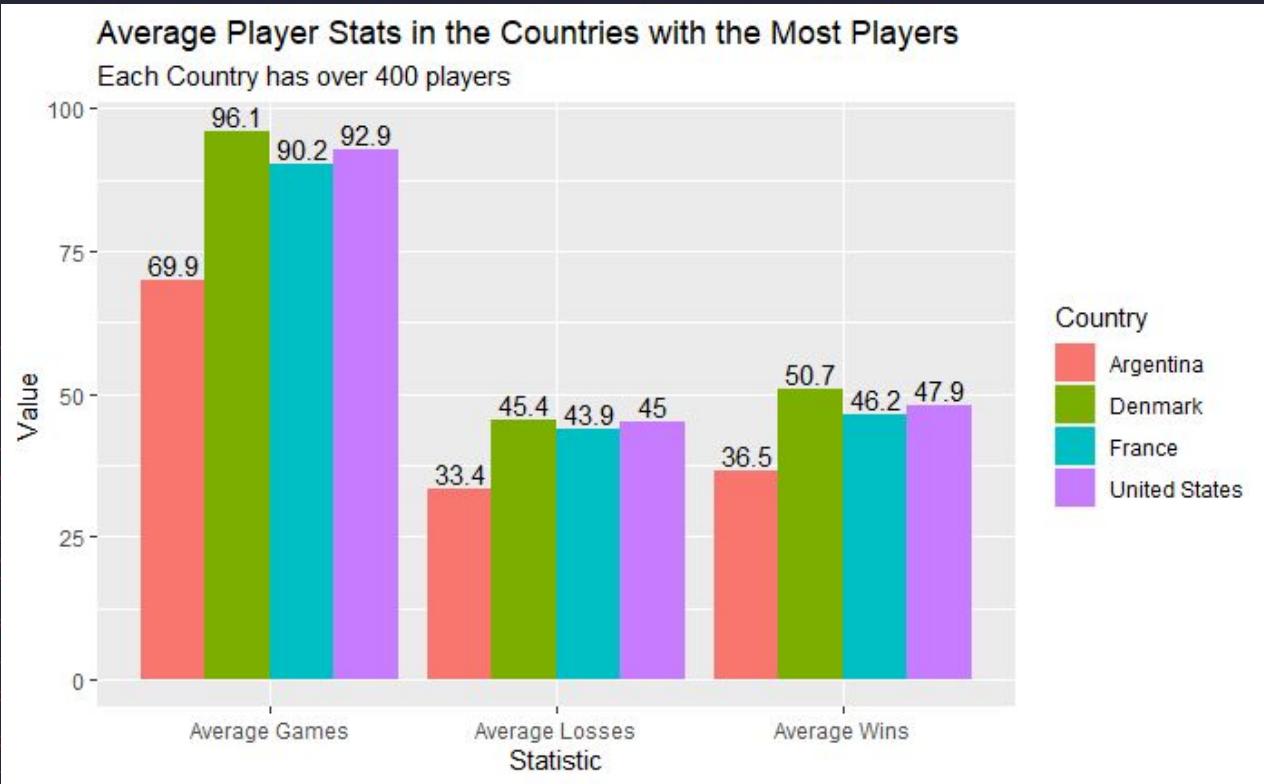
Drops are **more likely** to occur when an Individual Map is played.

Country vs Country

Hypothesis 2:

The average US player is better than other countries.

The US Falls to Denmark: Popularity Breeds Competitiveness



US players, despite playing **6 less games** on average than **Denmark** players, lose at almost the **same rate**.

Denmark barely rises to the top with each stat.

The biggest gap being average games (3.2) and wins not far behind (2.8).

Indicating no clear dominance on average by any country.

Wins and Losses Correlation

The Unique Beauty of E-Sports

Hypothesis 3:

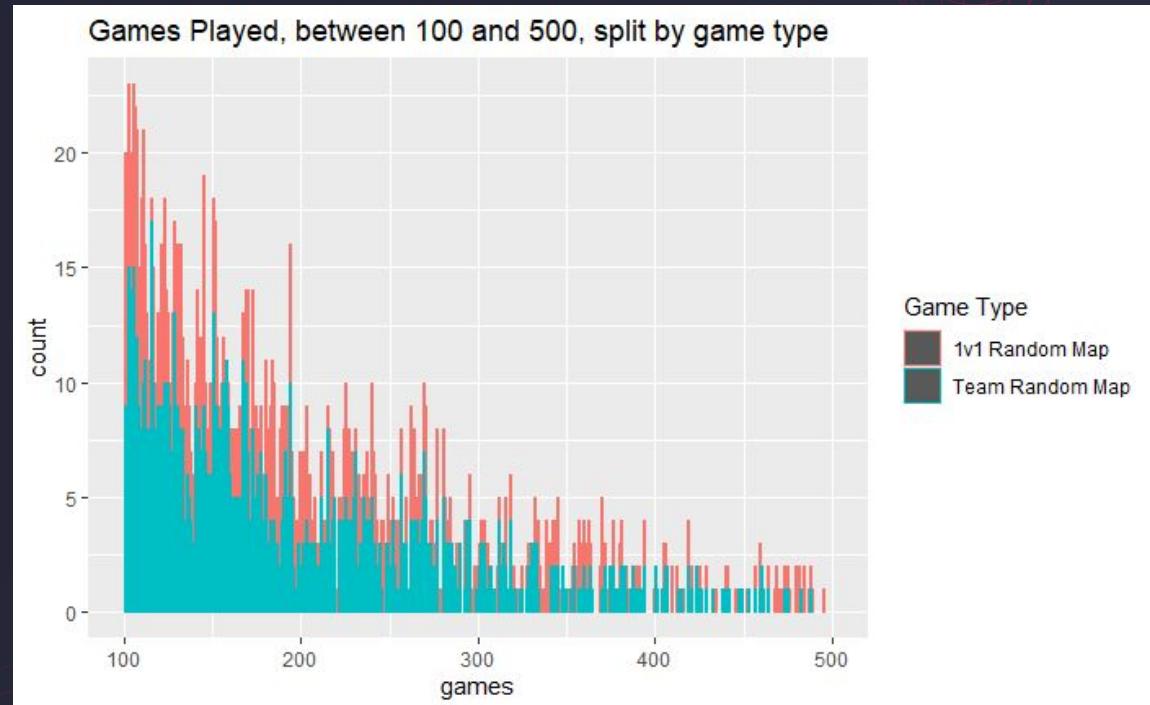
Wins and Losses are in fact correlated.

Context: Number of Games per Player

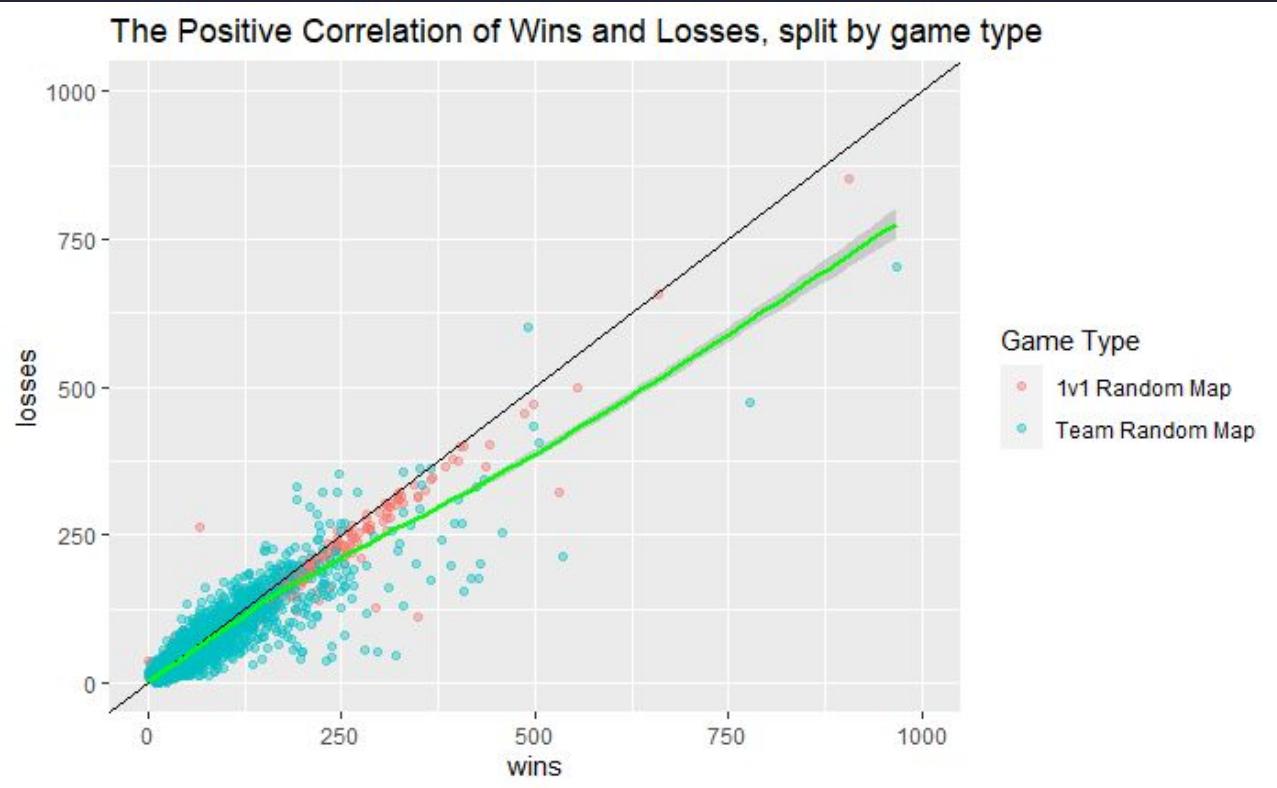
In most competitive sports, each team plays around the same amount of games.

- 82 in the NBA
- 82 in the NHL
- 16 in the NFL
- 162 in the MLB

For E-Sports, it's not set.



Wins and Losses Scatterplot, split by Game Type



Black Line (45° line):
Number of wins =
Number of losses

Green Line:
Closest regression with
the slope decreasing as
players gain more wins.

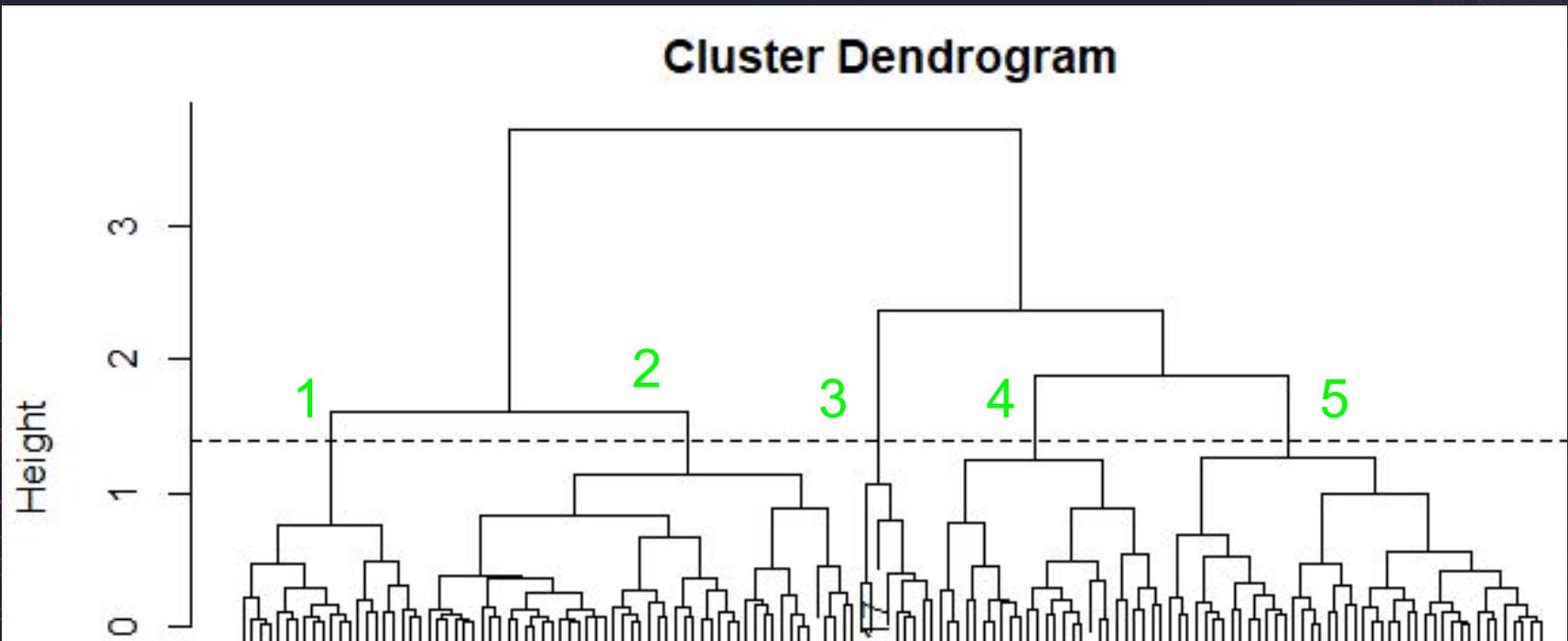
Cluster Analysis:

The Best Players

by Rank and Percentage of Wins

** Only Team Random Map

Hierarchical Modeling: Deciding Number of Clusters



Clustering

Group 2: (Orange)

All Stars

- Play a lot, win a good amount
- 70-75% win perc

Group 3: (Blue)

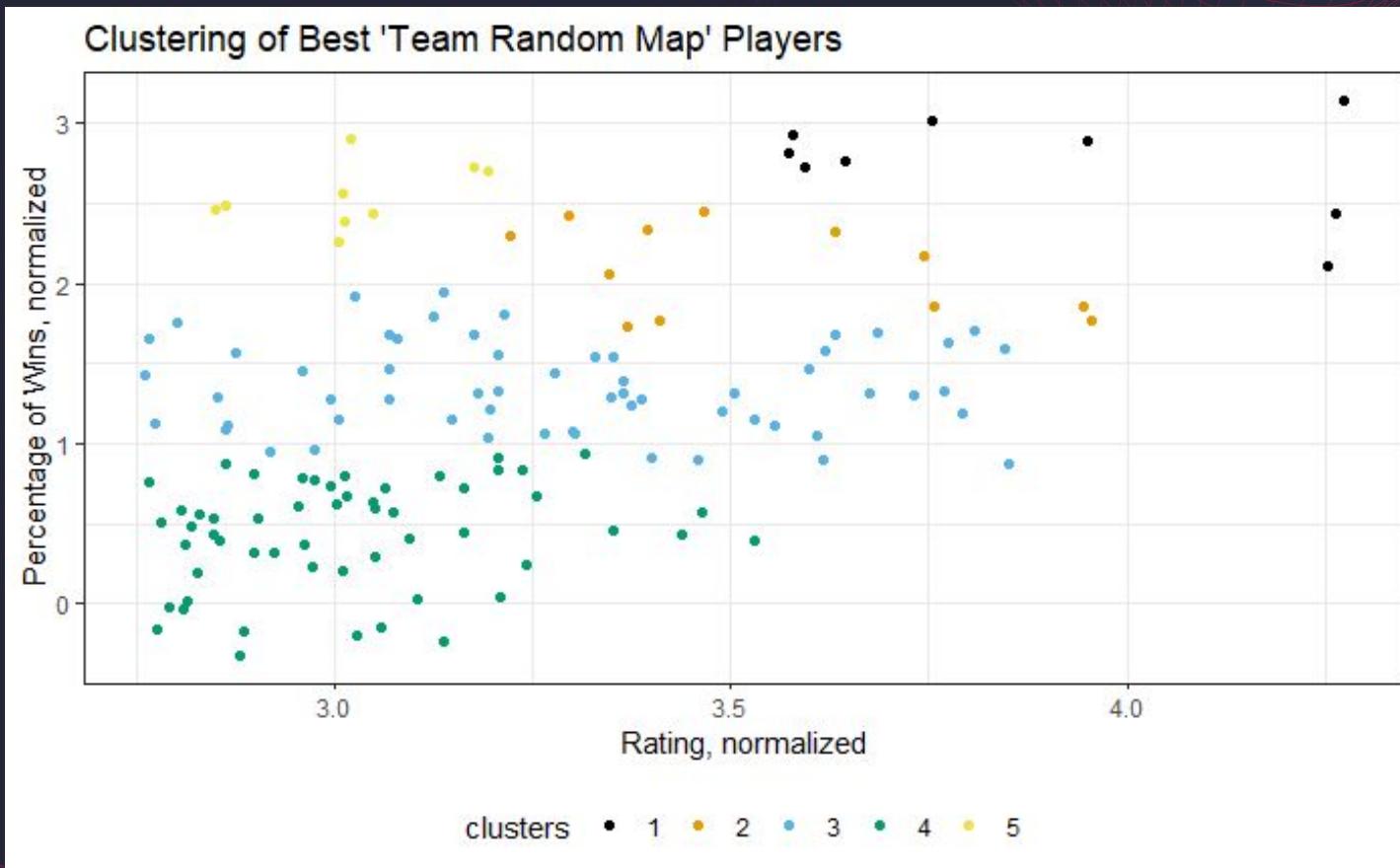
Competitive Players

- Big range of games played and rating
- 65-70% win perc

Group 4: (Green)

The Casuals

- Lowest rating
- 50%-65% win perc



Clustering

Group 5: (Yellow)

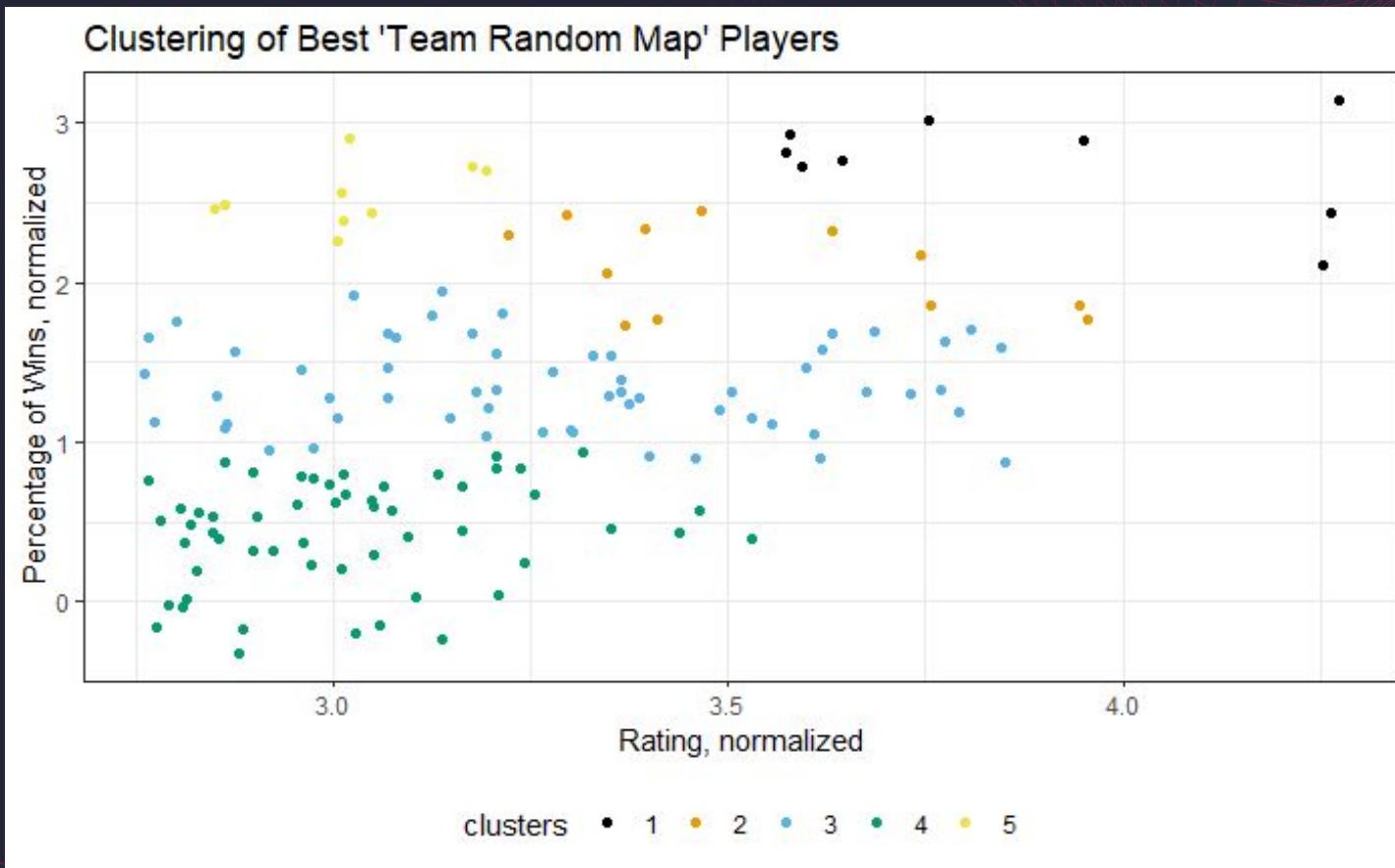
Underrated, Great Players

- Not many games played/lower rating
- 75-80% win perc

Group 1: (Black)

The Greats

- Most games played / highest rating
- 80% + win perc



Cluster Averages

Cluster	Average Rank	Average Rating	Average Games	Average Wins	Average Losses	Avg Perc of Wins
1: Greats	118.1	2590.2	276.33	230.4	45.9	83.1%
5: Underrated	1012.9	2292.4	94.4	76.4	18	80.7%
2: All Stars	318.9	2475.0	295.3	218.5	76.8	75.2%
3: Competitive	904.9	2321.2	270.1	178.4	91.7	66.7%
4: Casual	1176.3	2263.1	395.2	220.3	174.9	55.9%

Rounded to the nearest tenth and Sorted by win percentage

The background features a dark navy blue gradient with a subtle texture. Overlaid on this are several large, undulating wireframe mesh patterns in a bright red color. These meshes resemble the surface of a sphere or a complex 3D object, with their grid lines curving and flowing across the frame.

THANKS!
