



# 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.

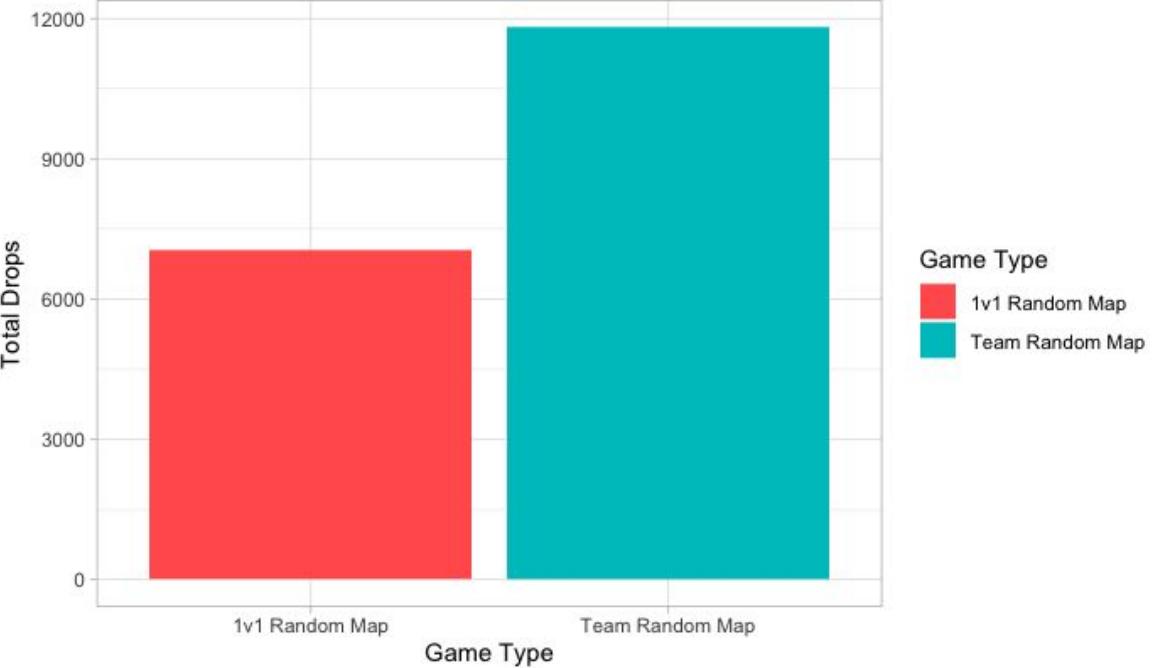
# **Teams vs lvl - Drops**

# Hypothesis I:

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

# Drops

Amount of Drops per Game Type



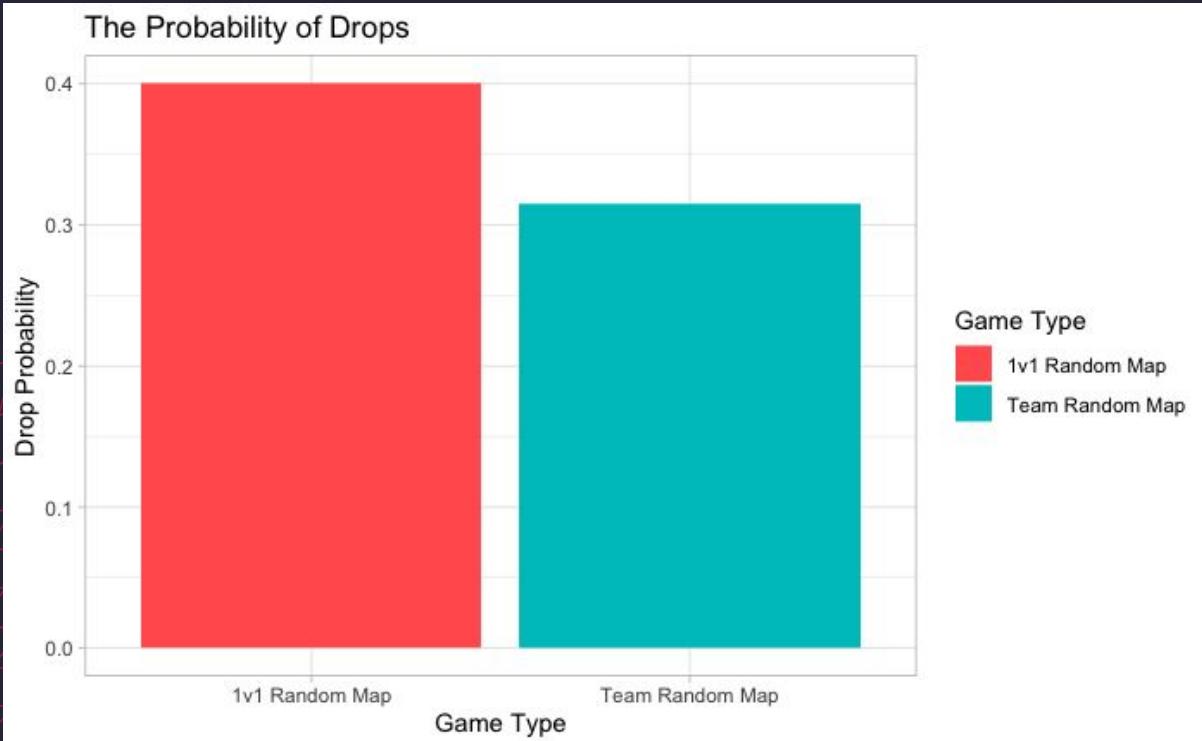
Thought Process:

Visualization shows that more drops occur during Team Matches.

1v1 Random Map: 7037  
Team Random Map: 11805

However, this wasn't enough...

# Drops



## Conclusion:

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.

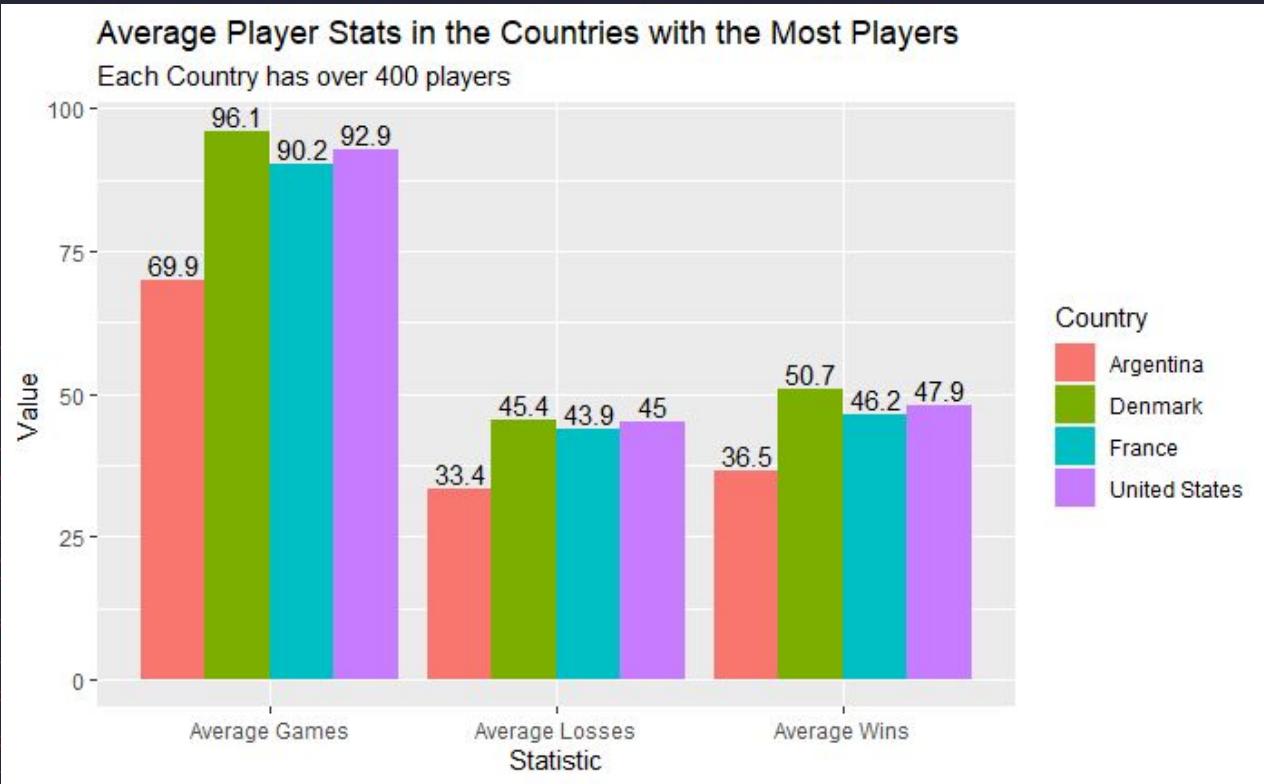
# Country vs Country

---

# 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:

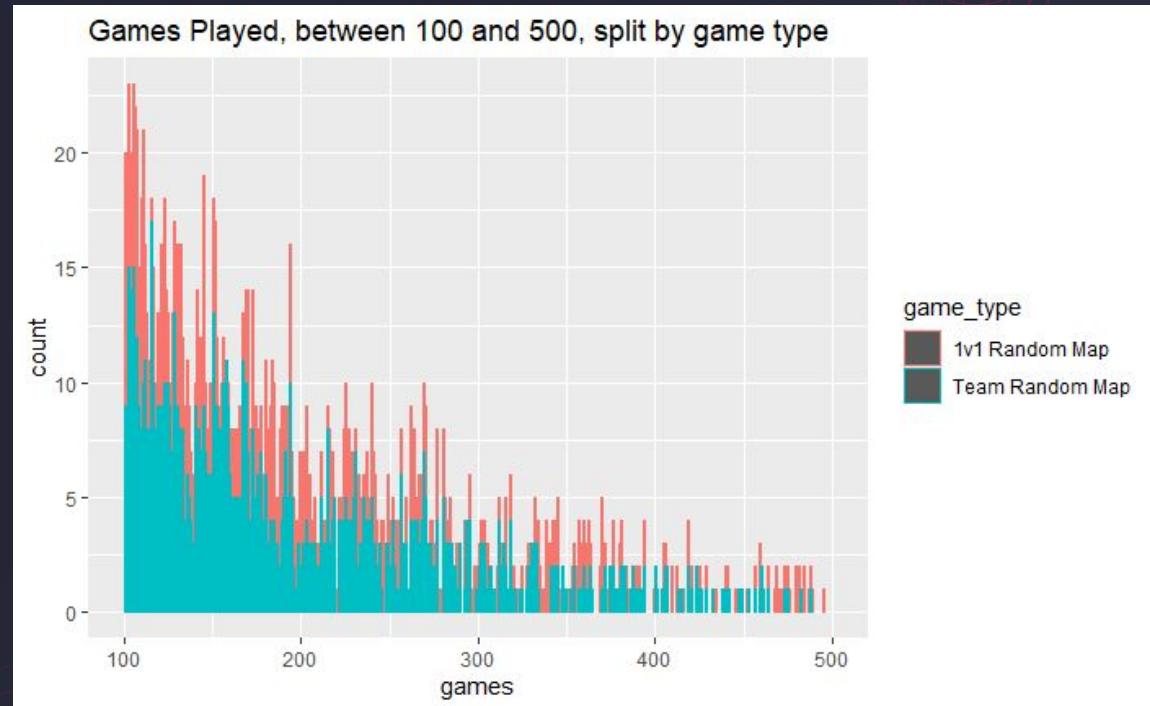
Wins and Losses are in fact correlated.

# Context: Number of Games per Player

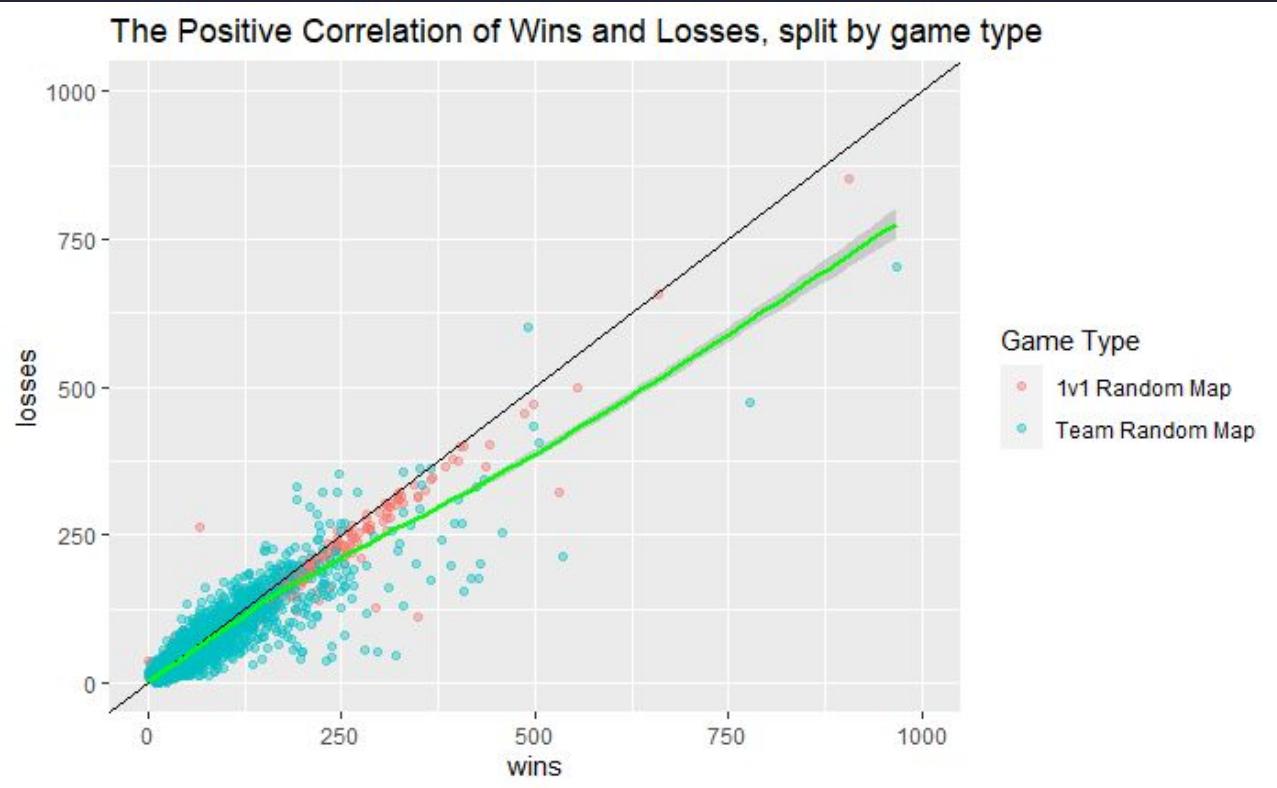
In most competitive sports, each team plays around the same amount of games.

- 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

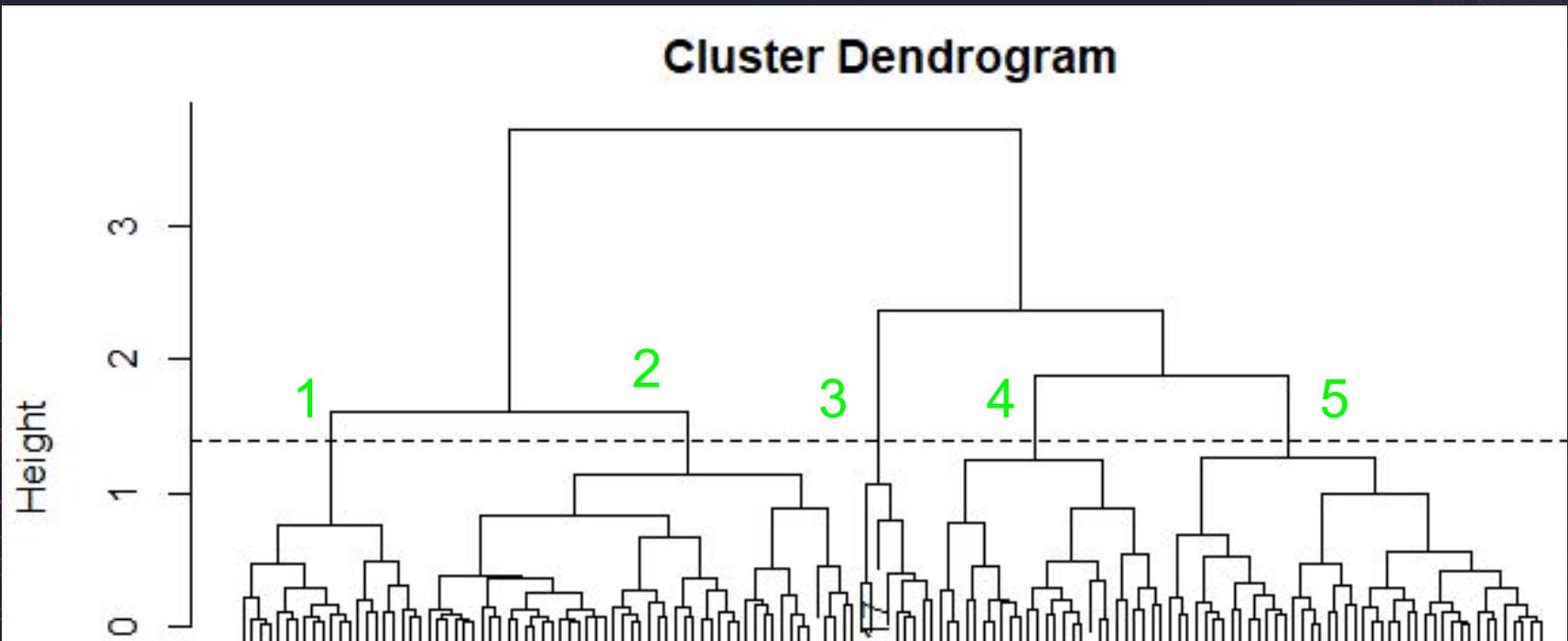
**Green Line:**  
Closest regression with  
the slope decreasing as  
players gain more wins.

# **Cluster Analysis:**

## **The Best Players**

by Rank and Percentage of Wins

# Hierarchical Modeling: Deciding Number of Clusters



# Clustering

## Group 2: (Orange) Competitive Gamers

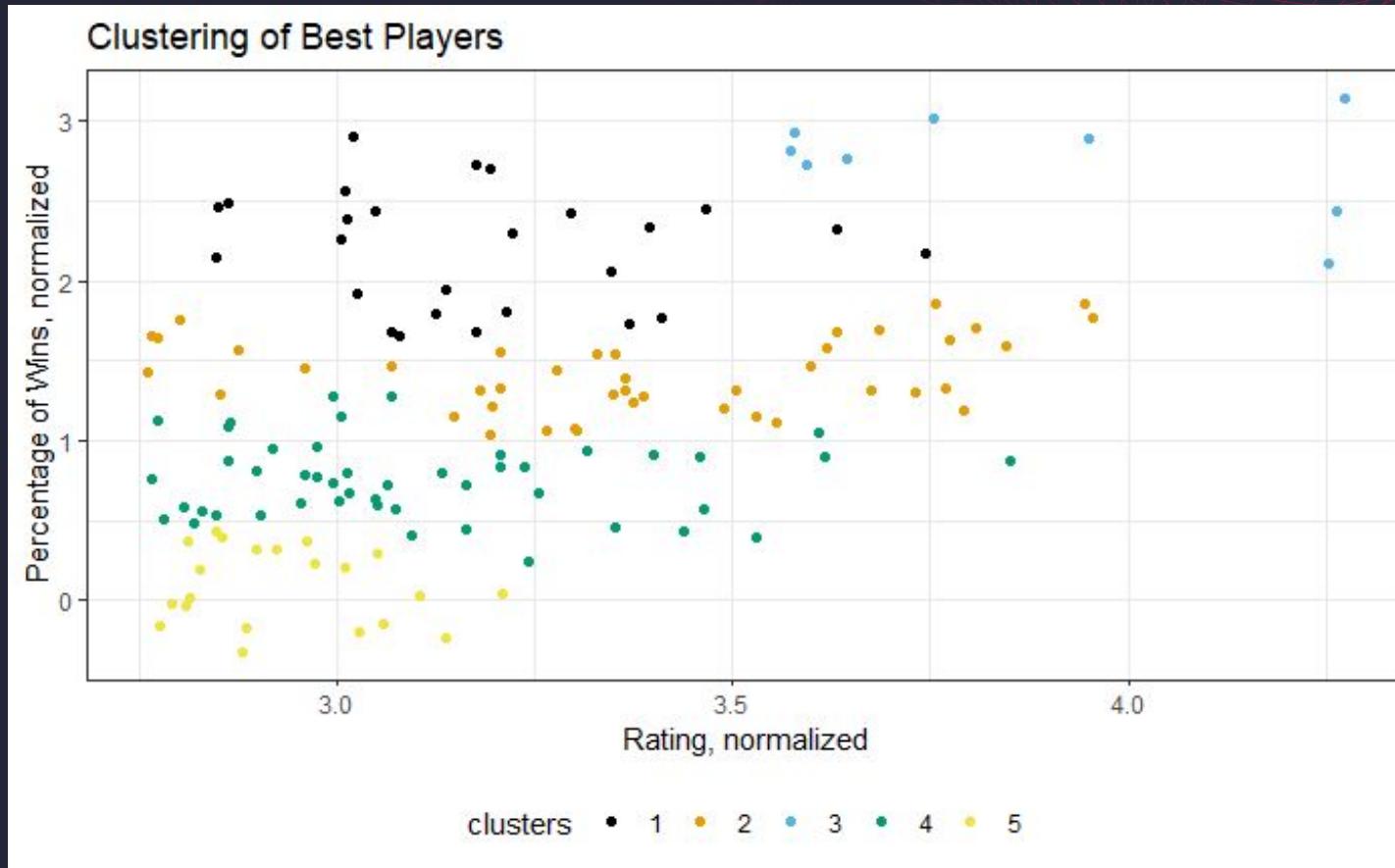
- Big range of games/rating
- 60-70% win perc

## Group 4: (Green) Above Average Players

- Lower rating
- 55-60% win perc

## Group 5: (Yellow) The Casuals

- Lowest rating
- 50%-55% win perc



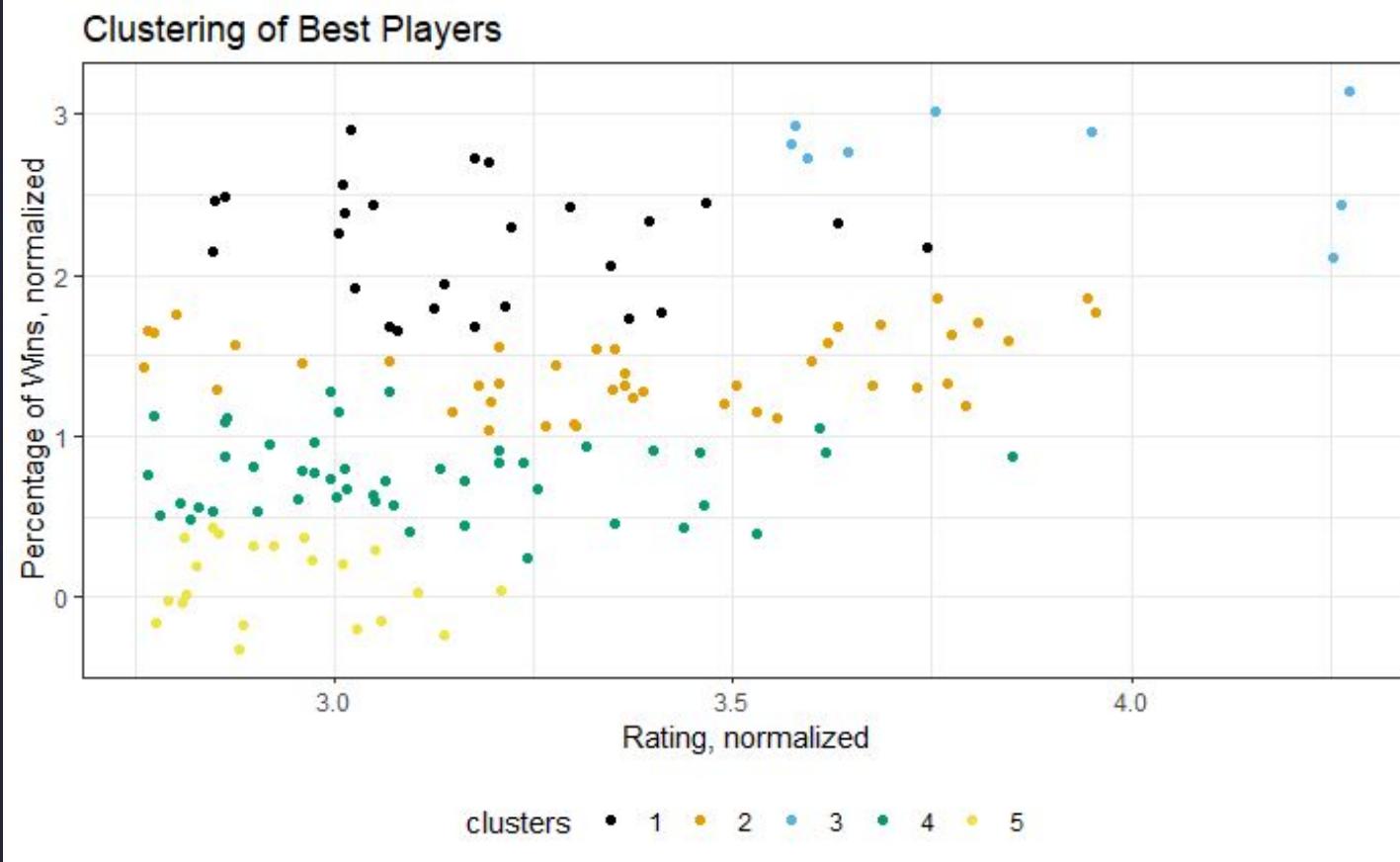
# Clustering

## Group 1: (Black) Underrated, Great Players

- Not many games played/lower rating
- 70-80% win perc

## Group 3: (Blue) 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
3: Greats	118.1	2590.2	276.33	230.4	45.9	83.1%
1: Underrated	710.8	2348.7	178.8	134.8	44.0	76.5%
2: Competitive	508.2	2417.5	284.9	192.6	92.4	67.4%
4: Above Avg	904.9	2321.2	377.8	223.1	154.6	59.6%
5: Casual	1176.3	2263.1	396.5	205.6	190.9	51.9%

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!**

---