WTA Grand Slam Statistics

Names: Maggie Byers, Lucca Ferraz, Amelia Yuan

Dataset Overview

- Women's Tennis Association Grand Slam Tournaments
  - January 2018 through August 2023
  - 23 Tournaments (Wimbledon 2020 canceled due to COVID)
- 2921 rows, 38 columns
  - 12 character, 26 numeric
- 6 columns about match information
  - tournament name, date, court type, minutes, score, round
- 18 columns about serving and responding to serves
  - No data about type of serve or left/right side of court
- 14 columns giving player statistics
  - Name, height, dominant hand, rank, etc

Case Study: Leylah Fernandez, 2021 U.S. Open

- Just 19 years old entering the tournament
- Ranked just 73rd in the world
- Went on a Cinderella run, reaching the finals of the tournament
- Defeated three of the top five seeds in the tournament on her way to the finals

Questions / Hypotheses to Explore

1. How do statistics such as length of match, player age and aces change as the tournament progresses into later rounds?

2. How does age interacting with match duration influence the likelihood of winning?

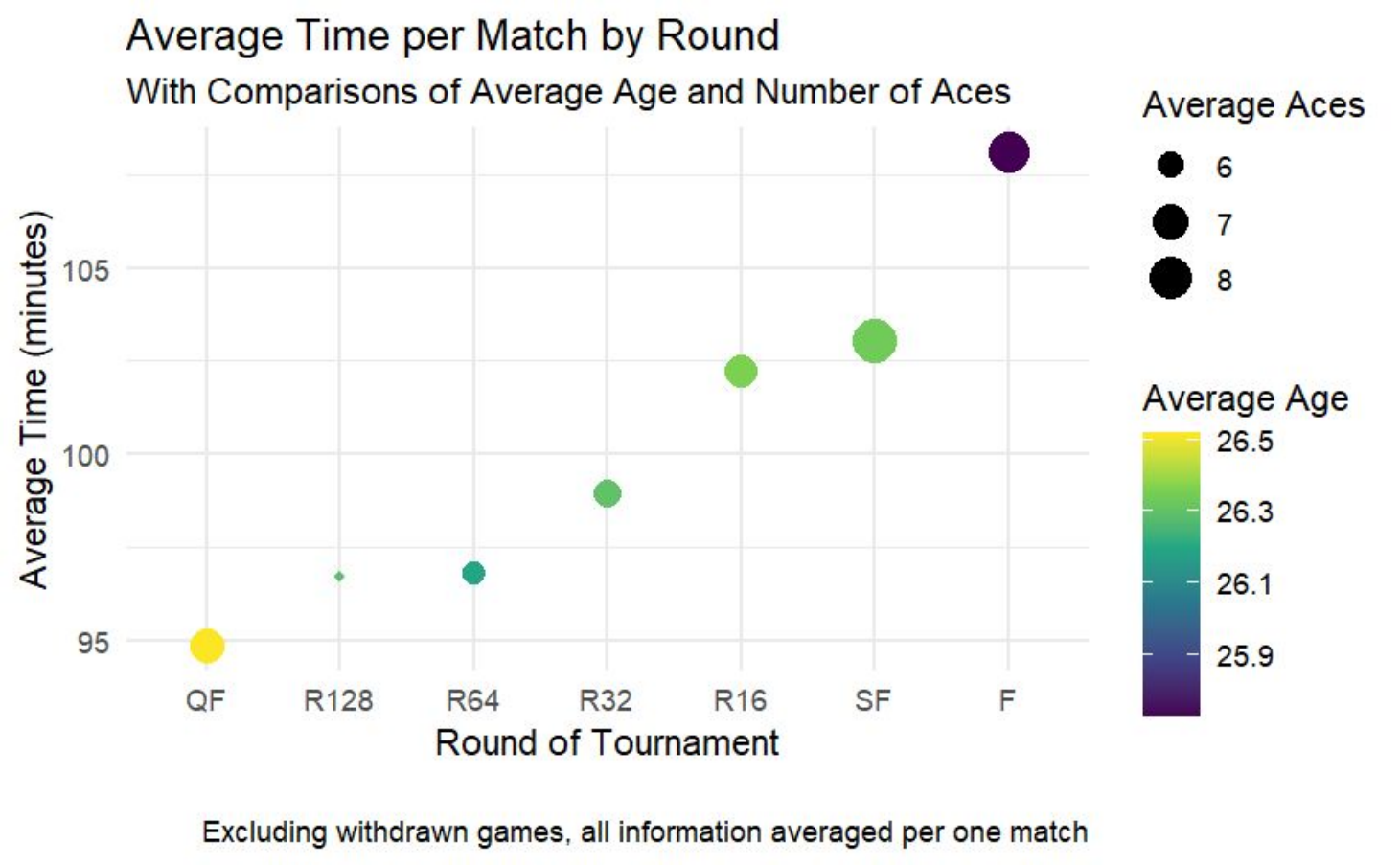
3. Which match statistics are most important to winning in general, and which are most important for pulling off a big upset?

Aces and Length of Match tend to Increase as Tournament Rounds Progress

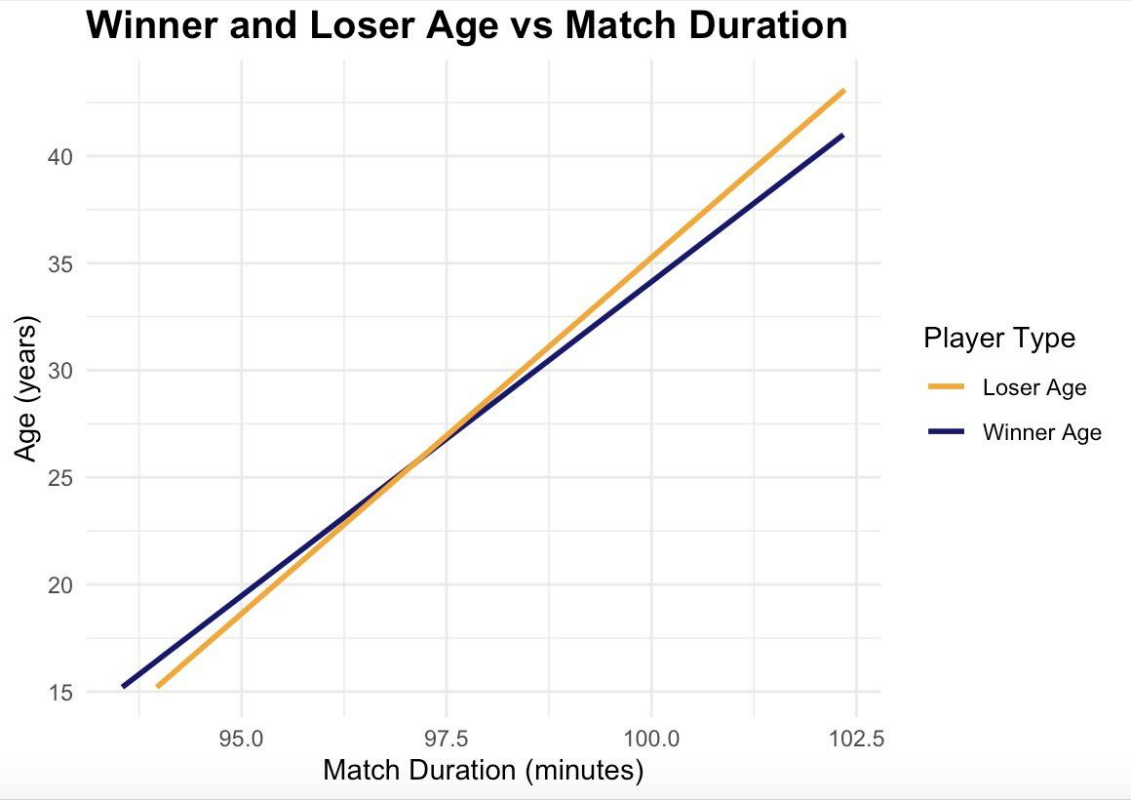
Observations:

Average player age tends to increase with the finals being an outlier

Quarter final round is an outlier for time but not age



Age of Winners and Losers in Relation to Match Duration

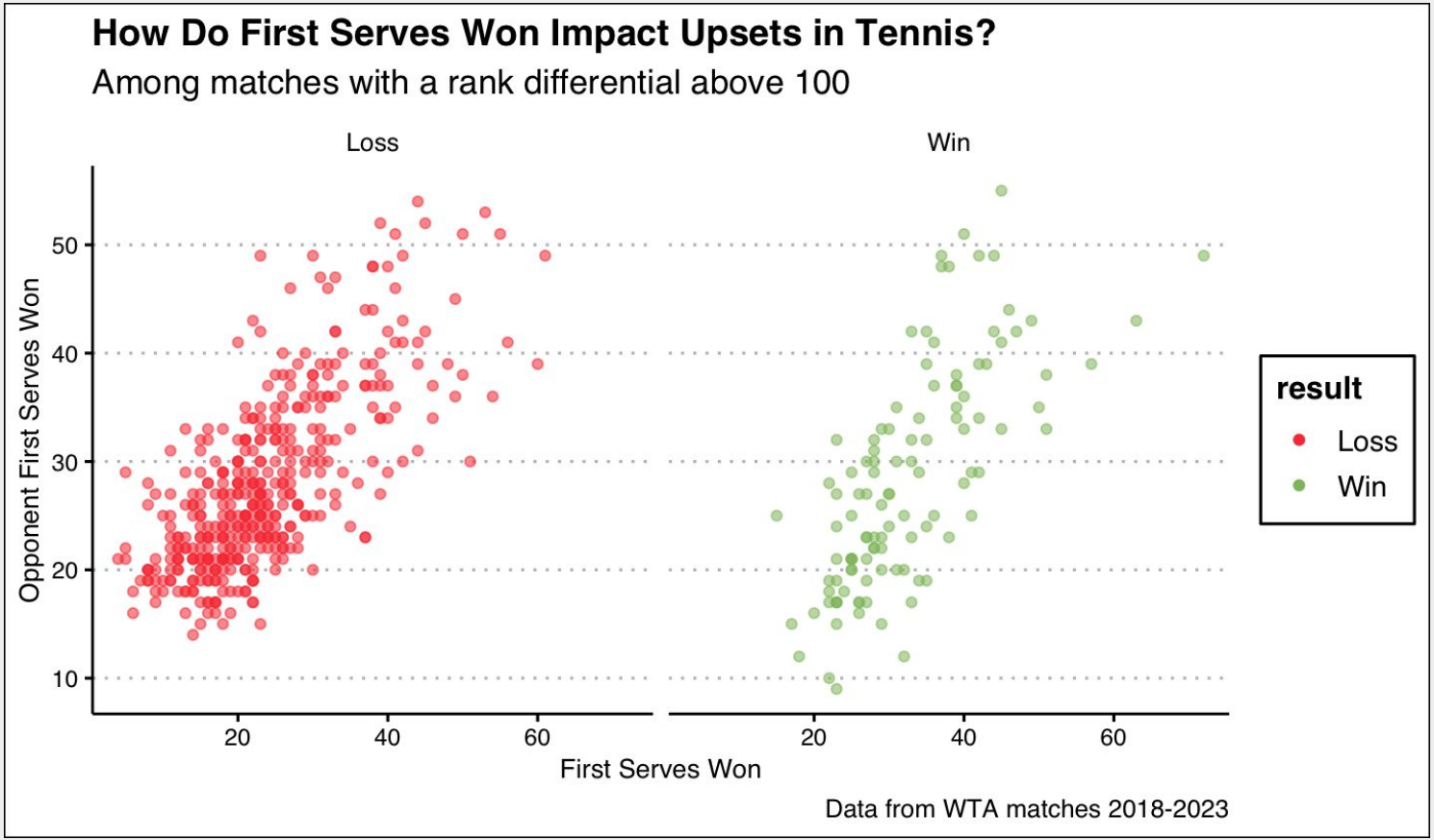


- **Intersection point**
- **Observations:** In the shorter games, the winners tend to be the older;

  However, in the longer games, the winner tend to be the younger one.

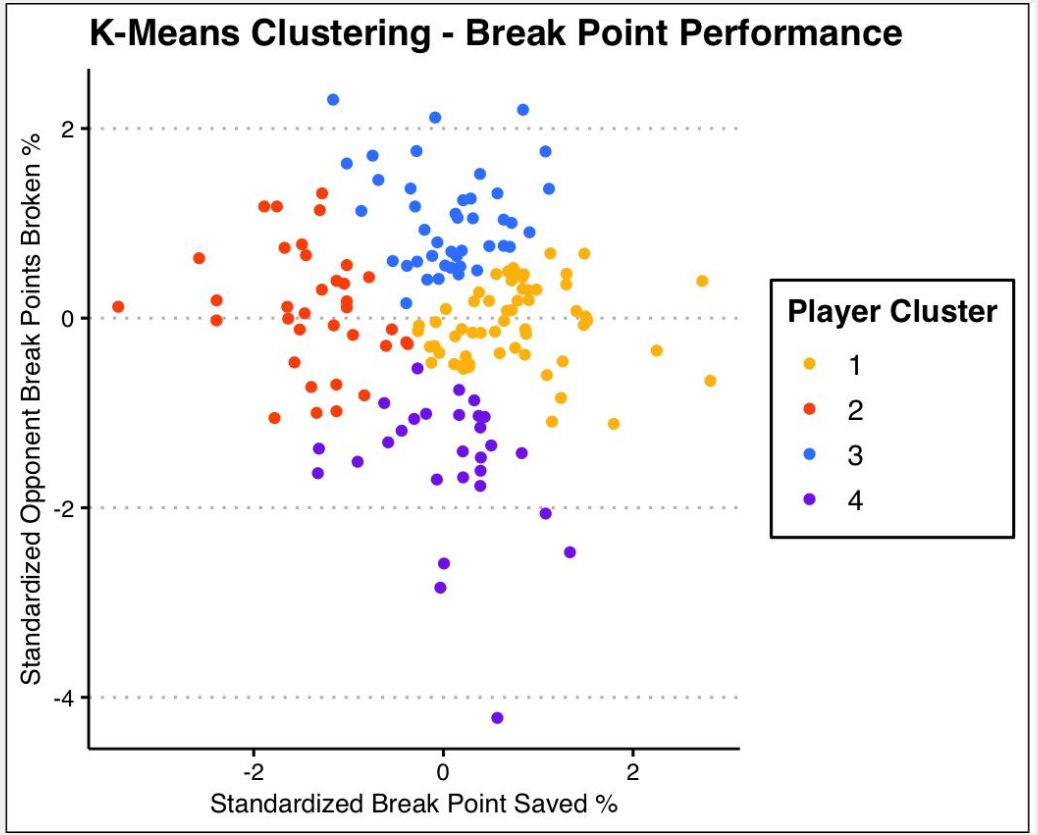
- **Insights:** physical fitness, experience

First Serves Are Crucial For Generating Big Upsets

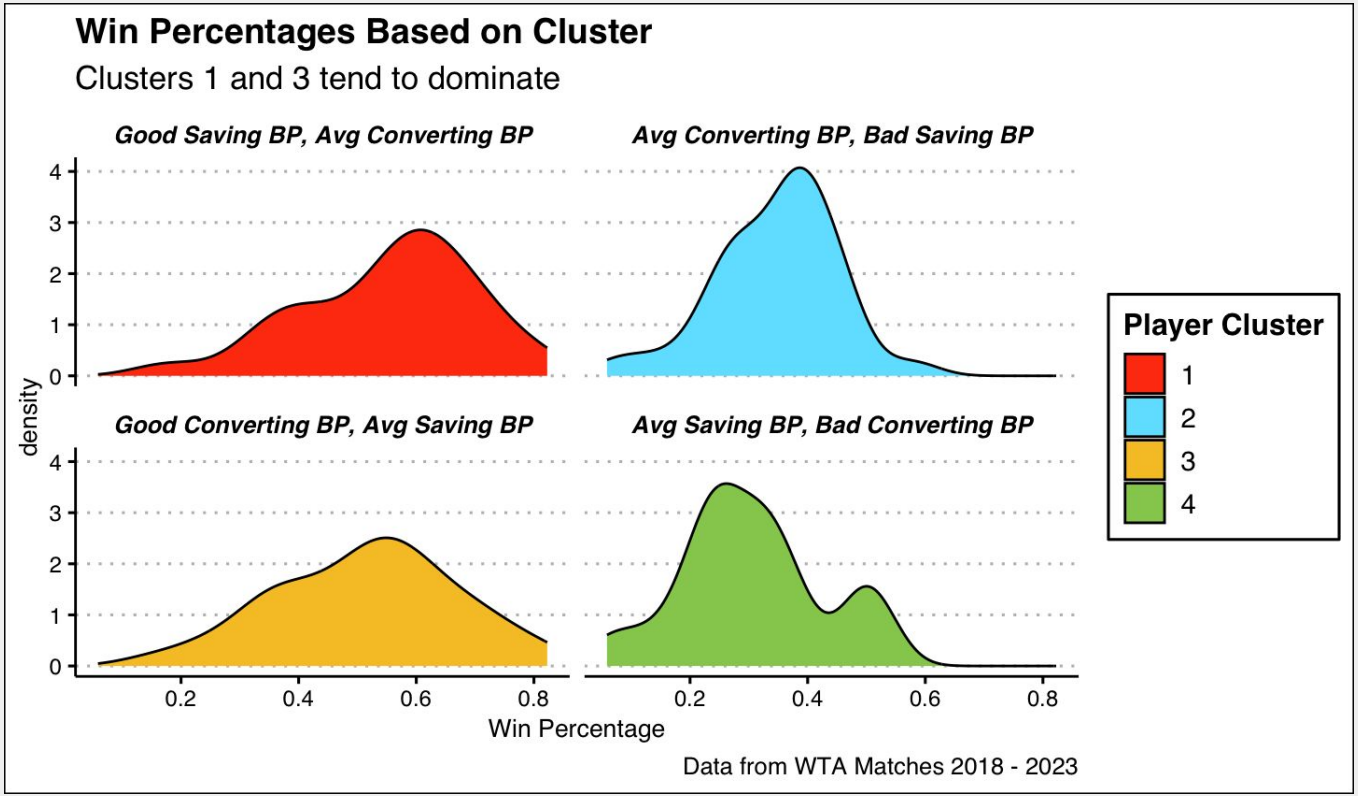


Clustering Analysis: Break Points

- Being able to “break” opponents serve is crucial in tennis
- Winning a breakpoint means you win a game on your opponents serve
- Top players tend to excel on breakpoints
- Chose number of clusters using nbClust package in R



Which Clusters Perform the Best?



- For the best players, saving breakpoints appears to be more important than converting them
- Worse players who are better at converting breakpoints perform better

Limitations

- Data includes only Grand Slam tournaments, not all WTA professional tournaments
    - 128 players per tournament of the 1650 players the WTA reports having
- No player tracking data
- No data on the location of serves or serve type
- No information on how a point was won
    - Into net, out of bounds, drop shot, smash

Discussion

- Later rounds tend to show more aces per match
- Match lengths increase as less talented players are eliminated
- Older players advance further into the tournament but younger players are more often in the finals
- Winning points on the first serve tends to be more important when attempting to pull off a big upset
- Breakpoints are a key differentiator between the best players and the rest