

Nadal vs. Federer: Pace of Play

Kevin Havis

2024-08-29

Nadal vs. Federer: Pace of Play

Introduction

Who sets the pace when there's no clock? In tennis, it's the players who control the pace of the game, especially during their service games. With no time limit on matches, only score, players must be tactical when considering their pace of play.

This analysis will use data from a 538 article that focuses mostly on how the game has slowed over time, but today we'll be looking at the player data specifically.

Let's compare the time between serves and the following point (the “down time”) for a classic rival pair; Rafael Nadal and Roger Federer.

Setting up the environment

First we'll install and load the necessary packages. `tidyverse` conveniently has all of them.

```
install.packages("tidyverse")
```

```
library(dplyr)
library(ggplot2)
```

Loading the data

We'll load the `serve_times.csv` file from the Github repo, which contains data on how many seconds elapse between a given player's serve and the start of the next point.

```
# Load data from Github repo
serve_time <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/tennis-time/serve_times.csv")

# Take an initial look at the data
glimpse(serve_time)
```

```
## Rows: 120
## Columns: 7
## $ server      <chr> "Nicolas Almagro", "Nicolas Almagro", "Nicol~
## $ seconds_before_next_point <int> 22, 19, 23, 24, 19, 34, 21, 17, 17, 34, 29, ~
## $ day         <chr> "28-May-15", "28-May-15", "28-May-15", "28-M~
```

```
## $ opponent      <chr> "Rafael Nadal", "Rafael Nadal", "Rafael Nada~
## $ game_score    <chr> "15-30", "15-40", "30-40", "Deuce", "Ad-in", ~
## $ set           <int> 1, 1, 1, 1, 1, 5, 3, 3, 3, 1, 1, 1, 1, 1, 5, ~
## $ game          <chr> "0-0", "0-0", "0-0", "0-0", "0-0", "2-4", "0~
```

Sub-sectioning the data

Let's subset the data to get the player's we're interested in, Nadal and Federer, by filtering for their service games.

```
# Create list of players we're interested in
players_of_interest <- c('Rafael Nadal', 'Roger Federer')

poi_serve_times <- serve_time |>
  filter(server %in% players_of_interest) |>
  select(server, seconds_before_next_point) # Our variables of interest

# Take a quick look to confirm results are as expected
glimpse(poi_serve_times)
```

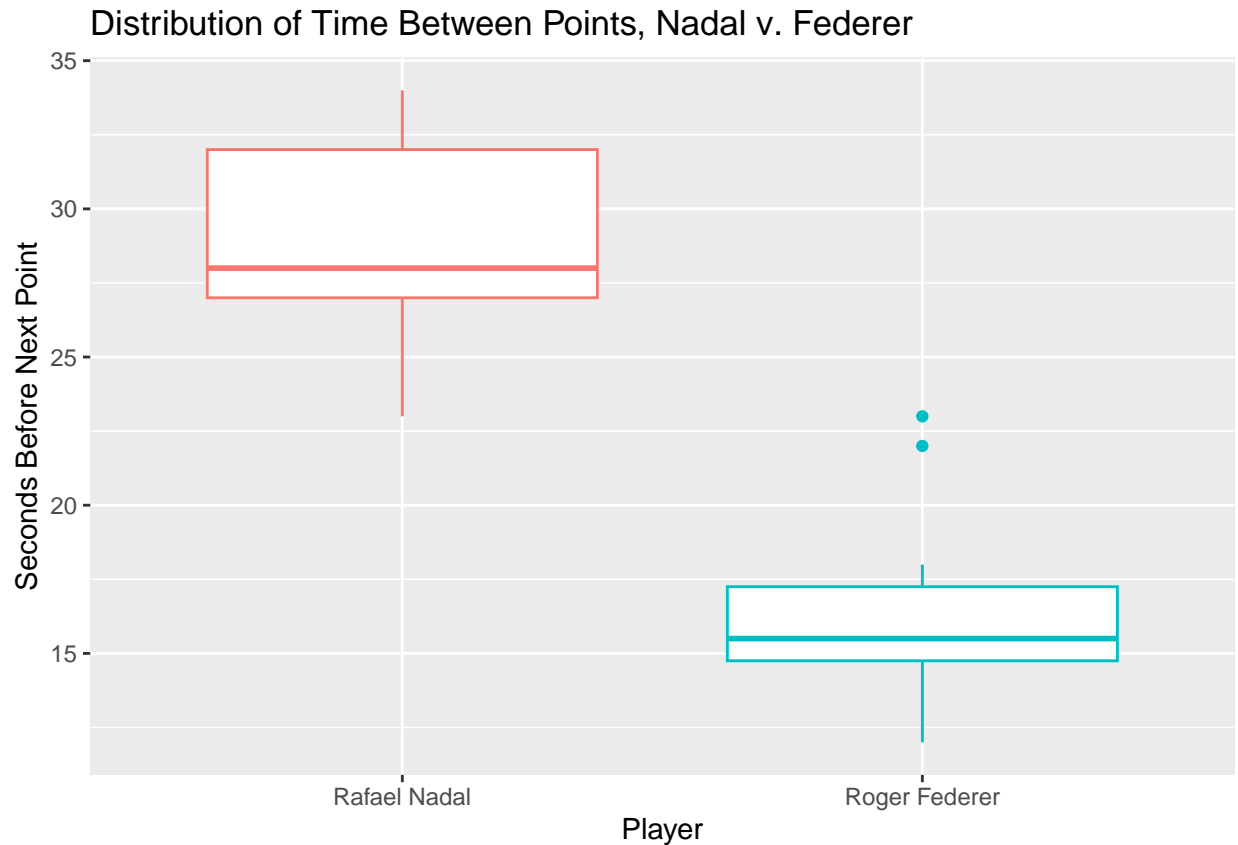
```
## Rows: 22
## Columns: 2
## $ server      <chr> "Rafael Nadal", "Rafael Nadal", "Rafael Nada~
## $ seconds_before_next_point <int> 34, 29, 33, 23, 27, 27, 15, 15, 14, 22, 17, ~
```

Visualize predictor vs independent variables

In this exercise, we consider `player` to be our predictor variable, and `seconds_before_next_point` to be the independent variable.

Let's visualize the distribution of the data for each player to see if there's anything interesting.

```
ggplot(poi_serve_times, aes(x = server, y = seconds_before_next_point, color = server)) +
  geom_boxplot() +
  labs(x = "Player",
       y = "Seconds Before Next Point",
       title = "Distribution of Time Between Points, Nadal v. Federer") +
  theme(legend.position = 'none')
```



We can see that for this sample, Nadal’s time between points is significantly longer than Federer’s, with a median `seconds_before_next_point` of approximately 28 compared to Federer’s 16.

Conclusion

According to the data available, Federer generally plays faster games than Nadal.

For tennis fans, this may make intuitive sense given Federer’s fluid and forward-moving play style. Nadal is known for playing topspin power shots and aggressive rallying from the baseline, so perhaps a longer time between serves is part of his strategy!

Admittedly, the sample size is small and limited to a few events. Additional data across a wider set of tournaments, surfaces, and players would make this a far more reasonable analysis and conclusion.