

Assignment #1: Data Exploration & Static Visualization Design

1. Data Analysis and Story







From the dataset, one observation was fairly obvious. Novak Djokovic has never lost a match in the finals of any Australian Open. Naturally, I dove into the numbers behind his performance in every final match and found some interesting patterns in the rally point statistics on [Tennis Abstract](#).

Djokovic tries to keep rallies fairly short. Most of his rallies are comprised of 1-3 shots and he's won more than 55% of rallies played to date. There is also a very strong negative correlation between the length of rallies and ease of victory. Shorter the rally time, greater the gap in points.

I thought it'd be best to break down the number of points won in each rally category against the points lost in every AUS-open final match over the years. This design doesn't just highlight his strength in winning rallies but also brings to light a marked improvement over the years in winning them faster as well.

2. Design and Colors

While a traditional stacked bar chart may be the obvious route for categorical data, I used a [radial chart \(Highcharts\)](#) since it scales better for a larger dataset as polar coordinates don't increase in size as much as cartesian coordinates. It's also easier to compare as the bars circle around each other and you don't need to scroll along any axes as you would in a stacked bar chart (with a lot of data).

1. Each individual bar is a breakdown of points won and lost during rallies between Novak Djokovic and player X in the Aus Open final during year YYYY.
2. Every stack within the bars is the number of points won/lost by Djokovic in each category. The categories and color schemes are as follows:
 - a. Points won during rallies of 1-3 shots. 
 - b. Points won during rallies of 4-6 shots. 
 - c. Points won during rallies of 5-9 shots. 
 - d. Points won during rallies of 10+ shots. 
 - e. All rallies lost points. 
3. I chose a sequential color of green from dark(1-3 shots) to light(10+ shots) to signify the importance of winning a rally as fast as possible. The grey color is used as an indicator of points lost due to lost rallies.
4. Legends have been added for all categories. Data labels are added only for the lost points bar  for users to quickly compare total won and total lost points in a match.

3. Inferences

Looking at the chart, the following observations are quite clear.

1. A huge chunk of Djokovic's rally wins is in the 1-3 shots range. This means he gives the opponent little window to keep play going and settles it swiftly.
2. Rafael Nadal was Djokovic's toughest opponent during the AUS open 2012. With a long rally of 190 points won and 175 rally points lost, It was a close shave for Djokovic but most of his rally wins are still in the 1-3 shots range.
3. Looking at the match bars towards the center, it's evident that Djokovic's game has improved considerably year over year. He's not just significantly reducing rally time but also reducing the gap between total lost and won points.
4. With regard to the point above, The Murray bars keep shrinking drastically towards the center which is a clear indicator he's beating out Andy Murray with more ease over the years. Even Nadal who gave Djokovic a tough time in 2012 was defeated soundly in 2019.