# The “Big Four” and the Andy Murray’s question

By: Jesus Fandino

## Summary

Male tennis has been living in a golden age for the last 14 years or so, thanks in a significant part to a generation of players including Roger Federer, Rafael Nadal, and Novak Djokovic. At some point after 2011, there was a push in the press for the “Big Four,” making room for Andy Murray to join the mix. Although those first three players have their places secured among the greatest in tennis history, a question often arises among tennis fans: Does Andy Murray belong there? We tried to answer that question using data.

## Design

We used colors in many ways, one of them was to highlight some binaries features we kept throughout the visualization:
- Orange/Beige: To make the distinction between active and retired players at the beginning, and Win / Lose after that.
- Red/Blue(#849DB1): Big 4 / Rest of the Tour (2 slides)
- Blue/Orage: Win / Lose

We also associated each one of the "Big Four" players with a single color all the way from start to end.

- Introductory Slides (1 – 4)

First of all, we understand that not all people have to be familiar with ATP tennis. For that reason, we spared four slides creating some context.
The first one is the Introduction, where we presented the question we were trying to answer with this visualization. Next, we displayed in a table the categories of the most important tournaments as well as their financial and ranking implications (That came up as a result of the feedbacks). Then we made the viewers familiar with the geographical and temporal context using an interactive map that shows all the cities that hosted the ATP World Tour since 2000. We used colors to highlight the playing surfaces, so the user will be able to track each of the tennis seasons throughout the year by checking/unchecking boxes. Finally, we provided a table to introduce the "Big Four" with pictures and some basic information.

- Justification of the “Big Four” using general statistics (5 – 7)

Four players stand out from the rest. In slide 5 we present the top ten players by the number of wins in the period 2000 – 2017 in a table where active and retired players are distinguishable. The timeline of their victories is plotted to the right for the top 5 players. We picked that amount for the graph to have a neat display. We used colors to encode each of the players and a line graph to be aware of the changes from one year to the next in a natural way.

We used the 6th slide to follow the average yearly winning percentage (that is the ratio of the number of wins to the total number of matches, times 100%) for the top ten players (2007 - 2017). The data was presented in a colored table, where we used a divergent combination of blue/orange to make the difference between mediocre and outstanding performances. Next, we introduced a summary of the winning percentage for the whole period of study (2000 - 2017) in the form of a horizontal bar plot, with the active status of the players encoded by color. Also, we created a box plot of the tournaments won, where each of the "Big Four" was identified using text labels. The color encoding had the same meaning that in the case of the bar plot next to it, that we previously addressed.

- The “Big Four” as a homogeneous group (8 and 9).

When we gather together the stats of these four players and compare against the rest of the tour, we obtain amazing graphs.
In slide 8, the number of tournaments won each year is displayed in the form of box plots. This is ideal to see how often in the last decade these players performed far away from the median of ATP Tour winners. In the 9th slide, we split the tournaments won by level, showing the Grand Slams as a horizontal bar plot because it allowed us to descend up to the individual tournament. Meanwhile, tournament levels M and A were represented as pie plots, since the number of tournaments is large enough to discourage any further splitting. Colors were used in the last two slides to identify the "Big Four" and the rest of the Tour as two different entities.

- How Andy Murray fits into the “Big” group (10 – 23).

So far, we have shown Andy Murray and the "Big Four" in the general context of the ATP World Tour. From now on, we started narrowing our comparison down up to the four of them.

We started by analyzing what happens at the Final stage of the Grand Slams (the most important tournaments in tennis). From that point, one graph leads to the next and we left the story to flow as it was emerging from them. We relied heavily on bar plots and line plots for the most part. This section of the visualization demanded a lot of changes as the work progressed, for that reason we are going to add more details to the discussion in the next section.

## Feedback

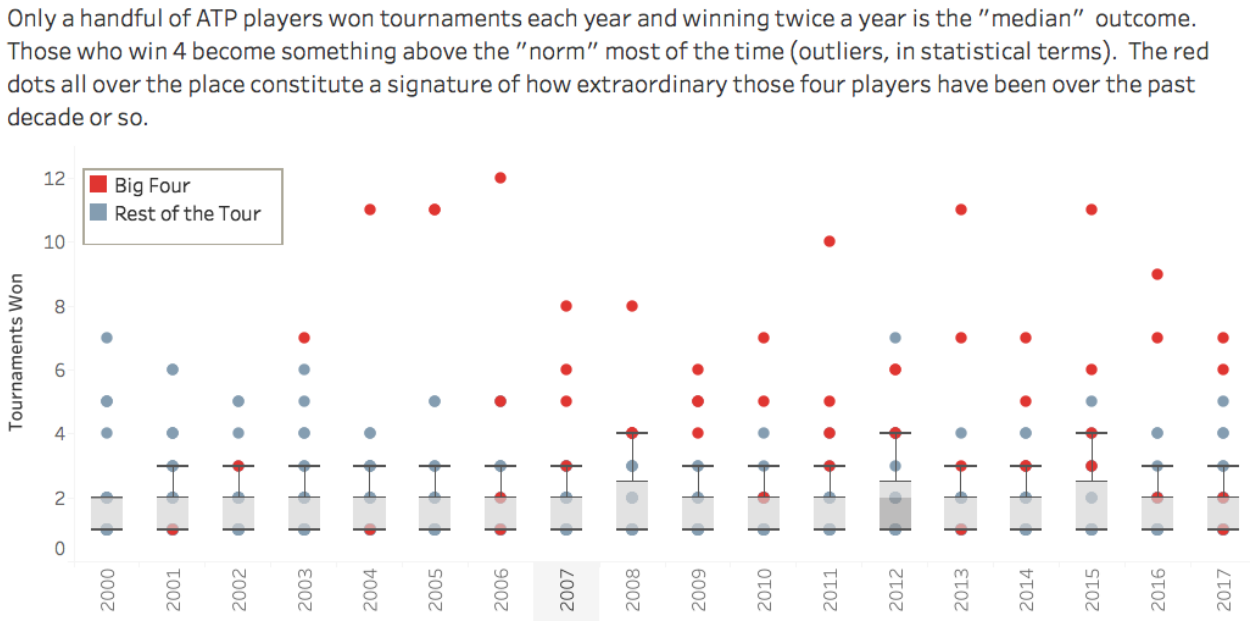
I have to say that as I was actively modifying the visualization as I was gaining more knowledge, and receiving feedbacks, not all the reviewers saw the first version the first time. I tried to organize the feedbacks as they were coming to me.

**(Reiner, via a phone call, his statements here are my recollection of them)**

*I believe that the combination of purple and gray makes difficult to see the numbers in the bar plots. In general, the choice of palettes could be improved for W/L cases.*
Me (Advice taken)

*Timeline of the Top 5 Winners*
*I would draw a line of the average of the 5 and another from the rest of the tour to see how they compare …*
That seems like a good idea.

*Big Four Stats (I)*
*It is difficult to spot the legend at first, I struggled to understand the meaning of the graph. Make sure you put the legend on a box or make it bigger and place in such a way it is the first thing we see in this graph.*
Understood. I will explain a little more and work on the legend.




*I'm not a tennis fan, and although I'm familiar with certain terms, such as Grand Slams, I don't know exactly what they mean and how are them different from other tournaments. You should clarify that a little bit.*
I understand. If I were to make a fair introduction to ATP tennis, that only will require a visualization. I'm short of space here, but I will see what I can do.

**The most relevant tennis tournaments in a given year.**

| Event category | Number | Total prize money (USD) | Winner's ranking points | Our Legend |
|---|---|---|---|---|
| Grand Slam | 4 | 167,690,000* | 2,000 | G |
| ATP Finals | 1 | 4,450,000 | 1,100–1,500 | F |
| ATP World Tour Masters 1000 | 9 | 2,450,000 to 3,645,000 | 1000 | M |
| ATP World Tour 500 series | 13 | 755,000 to 2,100,000 | 500 | A |
| ATP World Tour 250 series | 40 | 416,000 to 1,024,000 | 250 | A |

* 2018 total prizes according to http://tsmsportz.com/money/highest-prize-money-tennis-grand-slams/

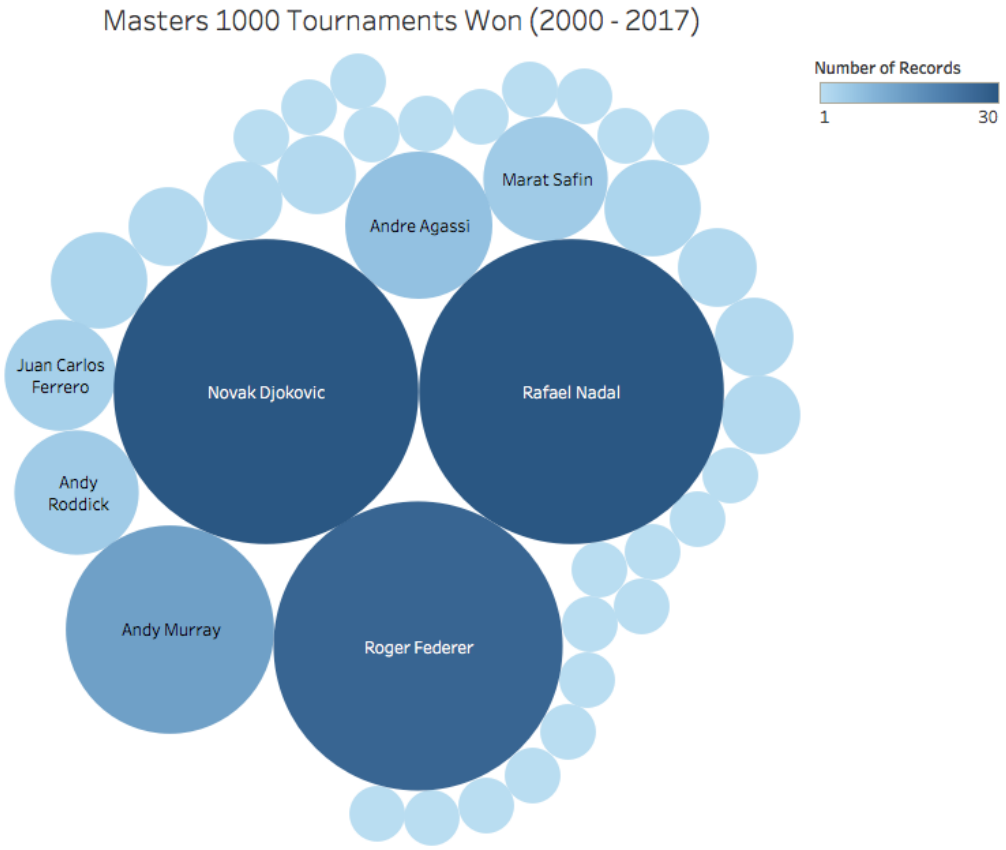
**(Daniel, via chat)**

*Hi Jesus, I actually review the tableau project so I can definitely give you some feedback! I would recommend that your captions be a little more in depth. Instead of viewing them as a title for each dashboard, think of them as what you would say in a powerpoint presentation. Basically, take the annotations in the later charts and make them your captions. For example, instead of having a caption 'Winning percentages over the years', change it to your annotation below.*
*Also, I would change any red/green color scheme to something else. 5% of all men are colorblind so this can be confusing.*

*Although I understand the appeal of bubble charts, it is intuitively difficult for most people to understand areas of a circle. I would switch to a bar plot*
Thanks! I will take care of all that.

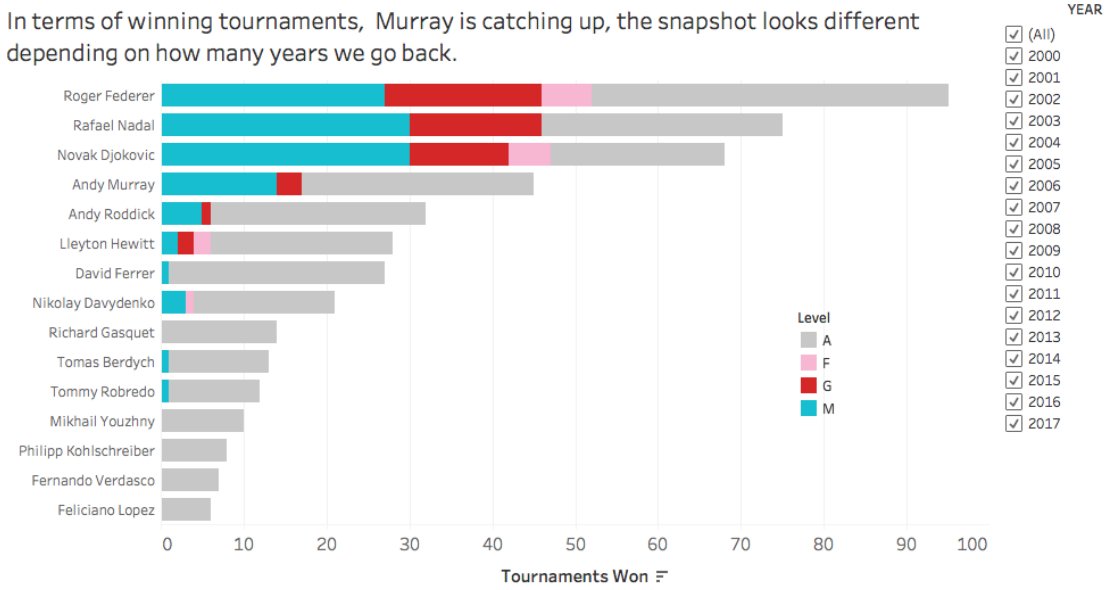
**Initially**

(This belongs to a version where I have already modified the original graph)




Name Player. Color shows sum of Number of Records. Size shows sum of Number of Records. The marks are labeled by Name Player. The data is filtered on Level, Round and Win/Lose. The Level filter keeps M. The Round filter keeps F. The Win/Lose filter keeps W.

After



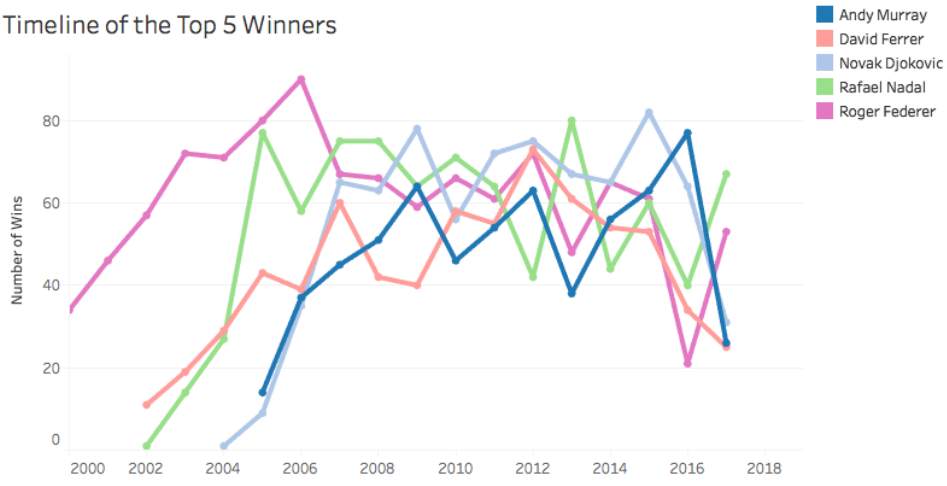
## Initially

Matches Won (2000-2017)

| | |
|---|---|
| Roger Federer | 1,089 |
| Rafael Nadal | 859 |
| Novak Djokovic | 763 |
| David Ferrer | 696 |
| Andy Murray | 634 |
| Tomas Berdych | 588 |
| Andy Roddick | 581 |
| Lleyton Hewitt | 526 |
| Tommy Robredo | 524 |
| Fernando Verdasco | 478 |
| Mikhail Youzhny | 477 |
| Richard Gasquet | 475 |

* Davis Cup matches not included

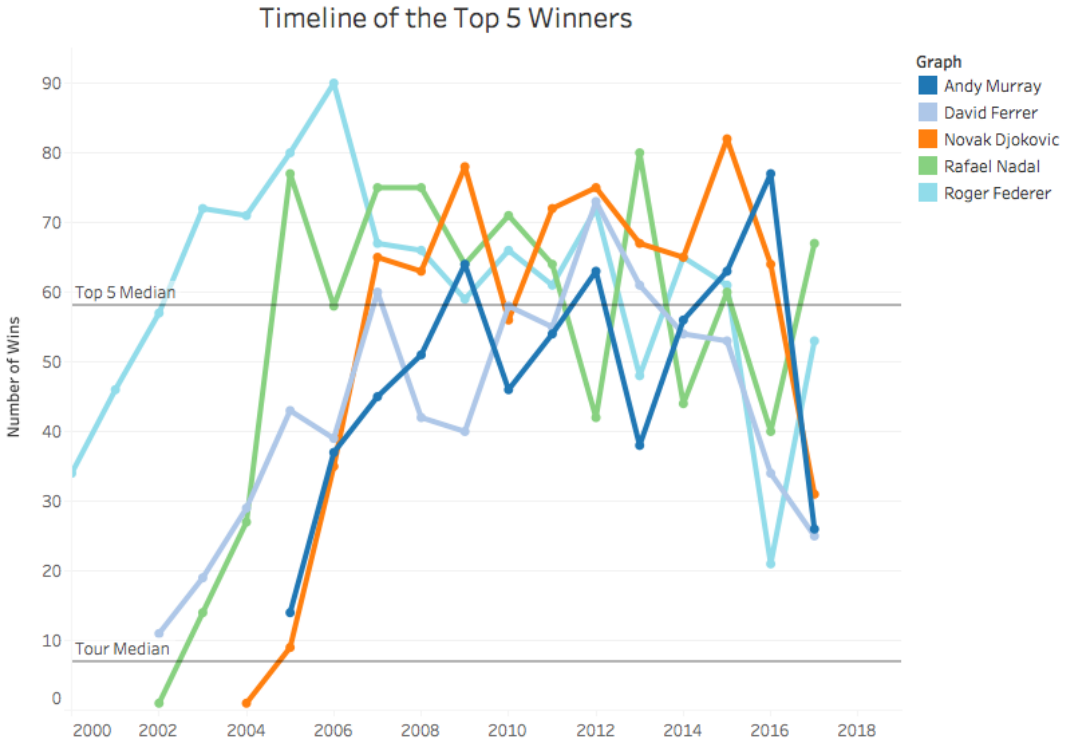
Active
Retired




The timeline graph shows the ups and downs of the best five players in wins over the years. It is clear that David Ferrer peaked in 2012 and has been in decline since. Andy Murray had to wait until 2016 to outperform the Big 3 for the first time. Unfortunately, he withdrew injured from the Tour after Wimbledon 2017. Djokovic also got injured in 2017. Both Federer and Nadal have enjoyed a renascence last year. Will it last?

## After taking into account the feedback from Reiner and Daniel

(Just the graph is shown at this moment)




The trend of sum of Number of Records for Tourney Date Year. Color shows details about Name Player. The data is filtered on Win/Lose and First 10. The Win/Lose filter keeps W. The First 10 filter keeps 5 members.

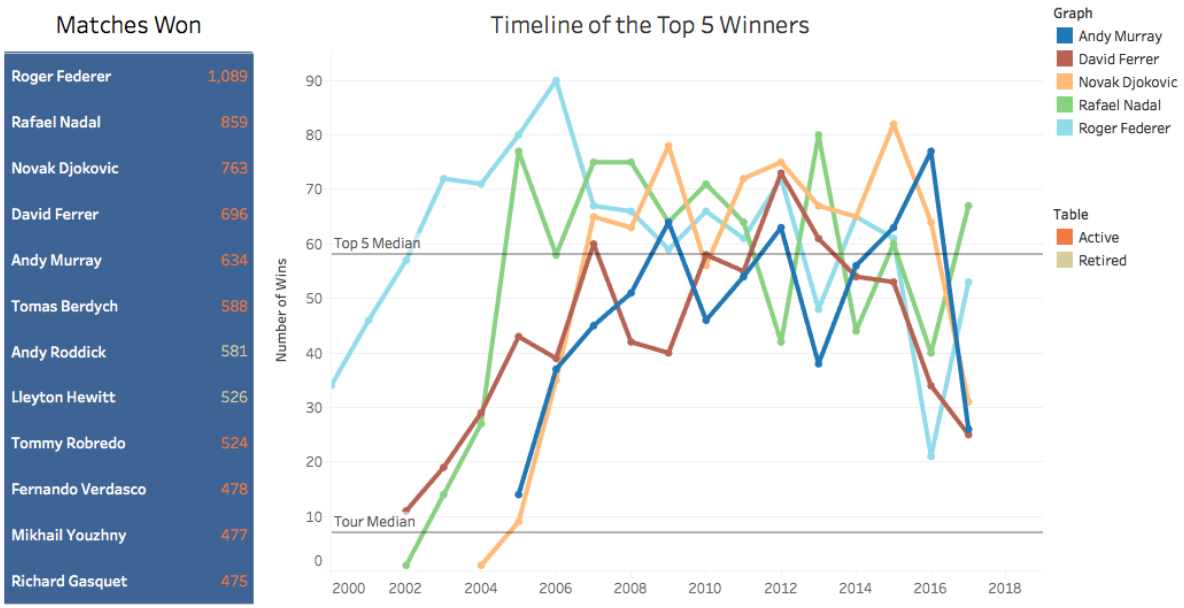
**(Manuel Ernesto, via email)**

*At the beginning, I would try to justify the importance World Tour Map on the story you are telling.*

I introduced a text with examples of the kind of questions that map could help to answer, I hope that constitutes a good enough reason to keep it.

*On "The best performances of the recent past"...*
*Can you move the legend for the table next to it?*
*If not, change the color of Novak Djokovic in the graph for it to be different than the one for active players on the table.*

You are absolutely right, I'm changing that right away:
I also took care of a change suggested by Daniel, regarding David Ferrer's color, that I will show at some point after Manuel's comments.




*On "Winning Percentages of the Years" I would invert the color scale. At first impression can be confusing due to traditional use in maps, where blue is associated with "below", and orange with "above"...*

I believe it is well documented in the legend, I will stick with my choice on this one.

*On "Tournaments Won and Winning Percentage"... Can you make the box on the boxplot more visible at first sight? Also, I would make more explicit the meaning of the graph in the title.*

I do not understand what "more visible" means, to me it is fine as it is now, however I wrote a few sentences to clarify the meaning of the box plot graph in that context.

*On "The Big Four Stats (I)" I would make each the Big Four in a different color or shade of color to highlight explicitly each one.*

That would have defeated the purpose of the graph, which is showing them as a compact group of players and see how extraordinary they are as such. I used this approach in just two slides, to show the viewer how persuasive the “Big Four” nickname could be.

*On "The Big Four Stats (II)"... I don't totally understand the meaning of the bar and pie graphs.*
I introduced a title explaining what those graphs are about.

*On "Performance in Grand Slams Finals"... I would use the words "Won", "Lost" in the legend instead of simply W/L (This apply for some other charts). I would also say on the title that results are ranked in decreasing order of percentage of matches won. Anyway, the graph tends to be deceptive: it is not "fair" to compare percentages for sample sizes that different, maybe it should be good to introduce a standard deviation or another measure of dispersion or error.*
I agree with many of your remarks. I believe W/L is unequivocally clear in sport contexts, and this way I could make some of my graphs more compact. I followed your recommendation with respect to alert the reader on how this graph was crafted. Even though it is generally “not fair” to compare samples of such differences, in the case of tennis, that also has a meaning. It tells the reader how often a given player reached high stages.

*On "Overcoming a Formidable Barrier Requires..." L/W on the legend stands for matches won/lost BY MURRAY? I would say that explicitly.*
I changed the title of that tab to “Murray’s problems? A formidable barrier of three” and introduced some text to explain that.

*On "Player's performance within a year..." Is that a specific year? Or the average over the years?*
It I implicit that it contains all the data available, if otherwise, I should have made it clear.

Hi Daniel. Me again, thanks for all your tips and recommendations on my Tableau project. Would you mind taking a second look at it for the purposes of follow up? I tried to make all the changes you suggested and also, I made some new slides. Thank you.

*Hi Jesus, I'll write comments slide by slide as I go through it:*
*1) Looks great*
*2) Good chart, I would just flush out the caption a bit more to guide the reader. Something along the lines of 'As you can see, Grand Slam wins are critical in tennis because of x,y,z about ranking points and prize money'*
Done!

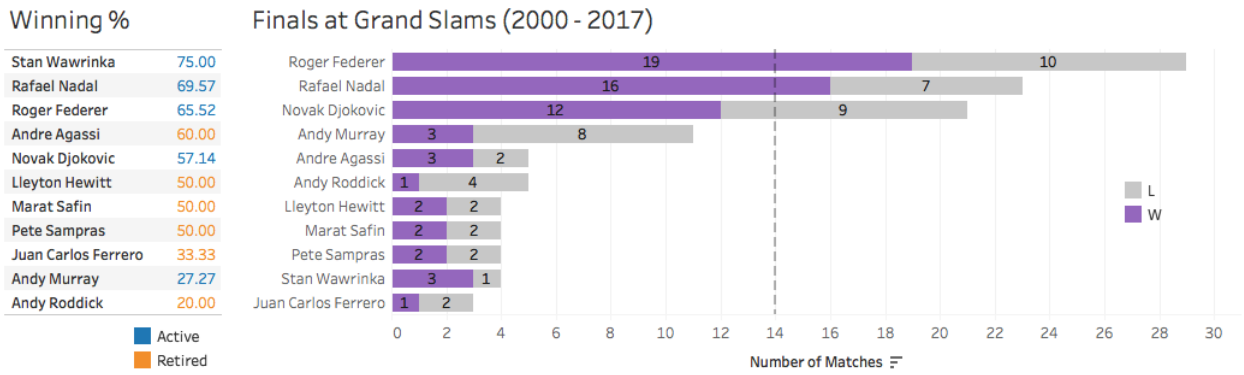
*3) Good, just include in the legend what the colors mean*
Done!

*4) Just add a little bit more to the caption - 'The Big Four: the current top performers in the last x number of years'*
Done

*5) David Ferrer and Roger Federer's colors are nearly identical*
I will take care of that (that picture is shown above)

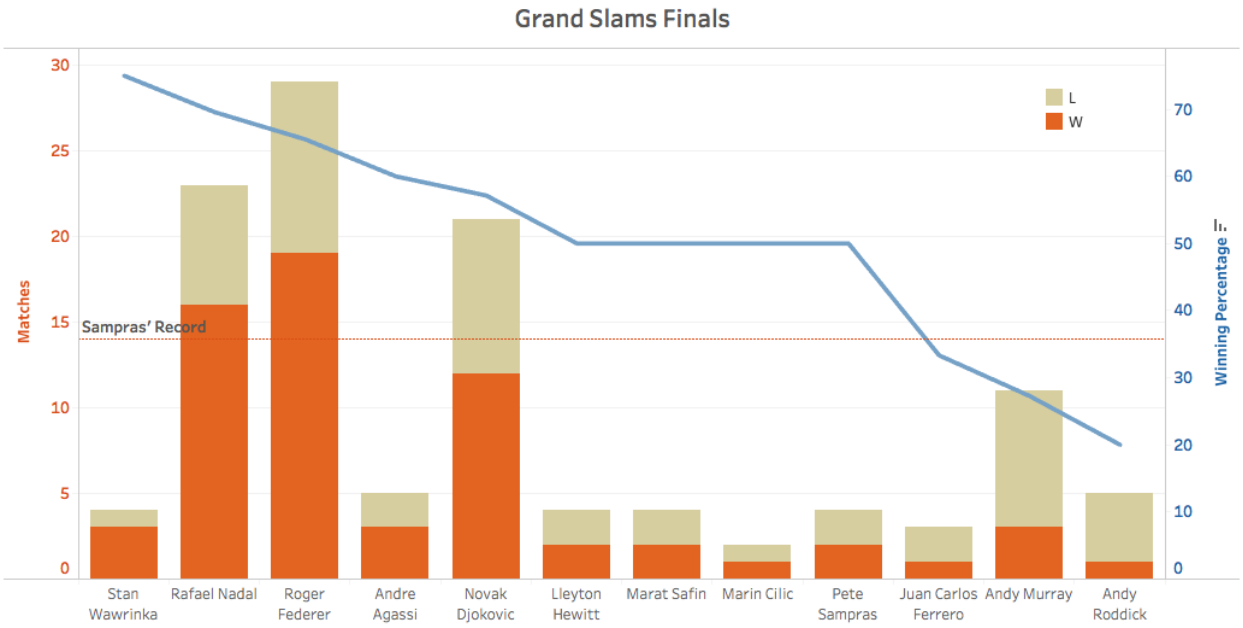
*6) Great*
*7) Great*
*8) Great*
*9) I think you should remove the names of the 'rest of the tour' players that won. it was a bit confusing*
Done

*10) I am not really sure how to interpret this graph - maybe a more flushed out caption would help. How is Pete Sampras's record 15 when he has a small bar at 2 in the graph?*
Fair point. This is one of the toughest decisions I made. My first attempt to address this was as this:



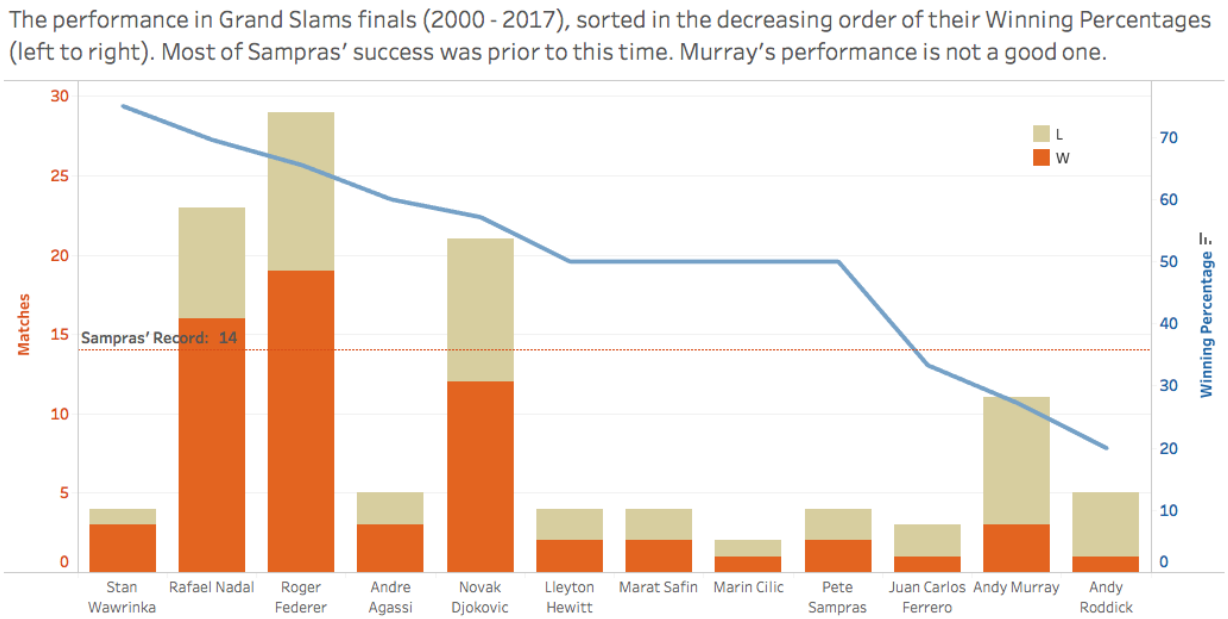

A table and a graph, that could have been two graphs. But the information they contain is somehow redundant to be presented in separate ways. This design has some flaws as I saw it. For instance, the players are ordered in a different manner and that alone complicates the readability. If I sort the order of the graph to match that of the table, then the graph will not have a visual meaning on its own. That is where the idea of merging the two came to my head.

After thinking a little, I ended up doing this:




That was the version that both Daniel and Manuel saw. It was (I thought), a good solution to that problem, it made all the sense to me, but will this graph transmit the information I intended it to?

After feedbacks from Manuel and Daniel, it was clear that some work has to be done explaining this graph:

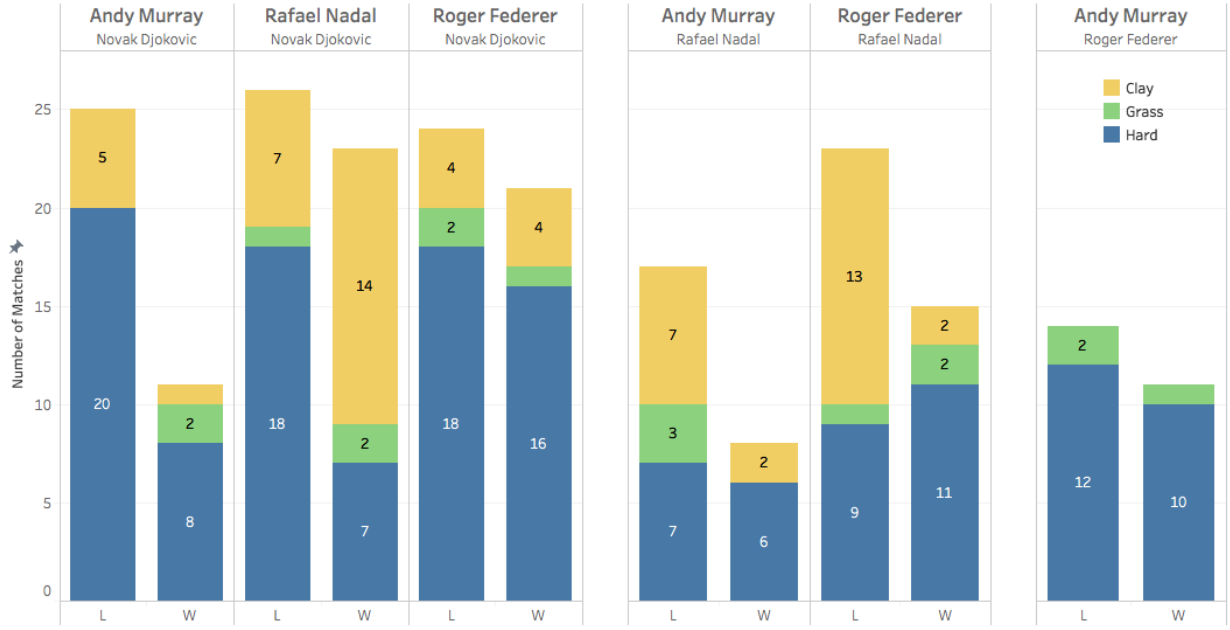


*11) Great slide. Maybe try adding win % next to each bar for each player. That's essentially the info you are trying to convey.*
I did it with the help of the tooltip, when the user hovers the mouse over the data, that information will show up.

*12) I would just flush out the caption a bit more - we can see that Andrew Murray has a formidable barrier (the other Big Three) for winning a title. We can see that when plays in the semi-finals or finals he often loses to xyz, otherwise he often wins. Something like that*
Done

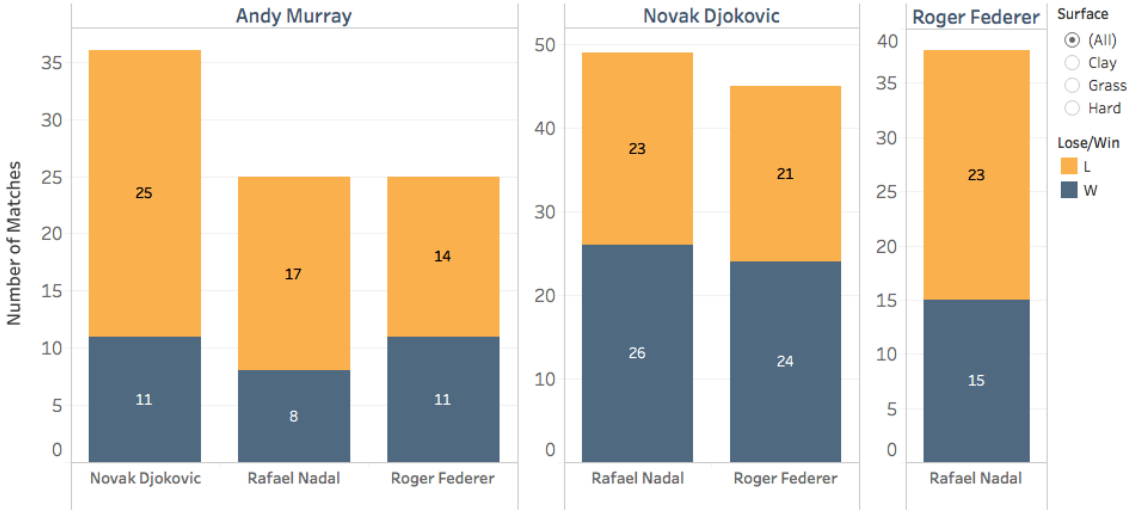
*13) This is great conceptually but a bit difficult to understand. I would change the charts to win % instead of number of wins/losses in order to reduce the number of bars*
Fair comment.
This is how the graph looked like (after taking care of color palettes):

Murray has a losing record against each one of the other three. It is one-sided in the case of Djokovic and Nadal (especially on clay) and not so against Federer.




How to address the issue of reducing the complexity of this graph without losing the information I believe was relevant? That was a good piece of a puzzle for me. The solution involves a little digging by the user, and at the end I cut in half the number of columns and simplified the color encoding.
As the results of the head to head for every surface will be displayed independently, the user will reach a good understanding of the graph in a short period of time. I believe the statements made on top of the graphs invite them to play around with the check boxes.
This is how the graph looks now:

Murray has only one favorable head to head versus the big three and that is against Novak Djokovic on grass. The clay court stats give a hint of why defeating Nadal on that surface is the ultimate challenge in tennis. Djokovic holds a comfortable overall lead over Murray and narrow ones versus the other big two. Federer and Murray have never met in a clay court. How strange is that?




*14) Good*
*15) Great*
*16) Great*
*17) Great*
*18) Great*
*19) I'm not sure what the color is supposed to signify. If it is just number of wins the height is enough. I would color it by player here instead*
In my first version of the visualization it was as Daniel suggested, but I decided to explore with color encoding before he had time to see it. It turned to be not a good idea. I went back to the original color encoding.

*20) I would explain this one a bit more - as you can see x,y,z won this many tournaments this year but xyz has been dominating lately...*
Done

*21) Good*
*22) Great*
*23) I don't really think this slide is needed*
Ok. That was intended as a summary but I was doubting about it anyway. Deleted

Deleted picture.

| | | | | |
|---|---|---|---|---|
| **Grand Slams** | 20* | 16 | 12 | 3 |
| **Masters 1000** | 27 | 31* | 30 | 14 |
| **Tour Finals** | 6 | 0 | 5 | 1 |
| **Olympics** | 0 | 1 | 0 | 2 |
| **Career Slam** | YES | YES | YES | NO |
| **Weeks at No. 1** | 308 | 171 | 223 | 41 |
| **Best Slam Tournament** | 8 (Wimbledon) | 10 (Roland Garros) | 6 (Australian Open) | 2 (Wimbledon) |
| **Best M 1000 Tournament** | 7 (Cincinnati) | 11* (Monte Carlo) | 6 (Miami) | 3 (Shanghai) |
| **Most Titles in a Year** | 12 (2006) | 11 (2005, 2013) | 11 (2015) | 9 (2016) |

** Results updated as April 2018*

*24) Excellent*
*Hope this helps!*
*4:41 pm*

*MAY 04, 2018*
*Awesome! Thank you very much! That helps me a lot*
*2:33 pm*
*My pleasure Jesus!*
*5:52 pm*

## Resources

- https://github.com/jedfarm/City-and-Country-Code-for-ATP-Matches-Data.git
- http://www.atpworldtour.com/en/rankings/former-no-1s
- https://github.com/JeffSackmann/tennis_atp.git (dataset)
- Wikipedia
- Tableau Forums