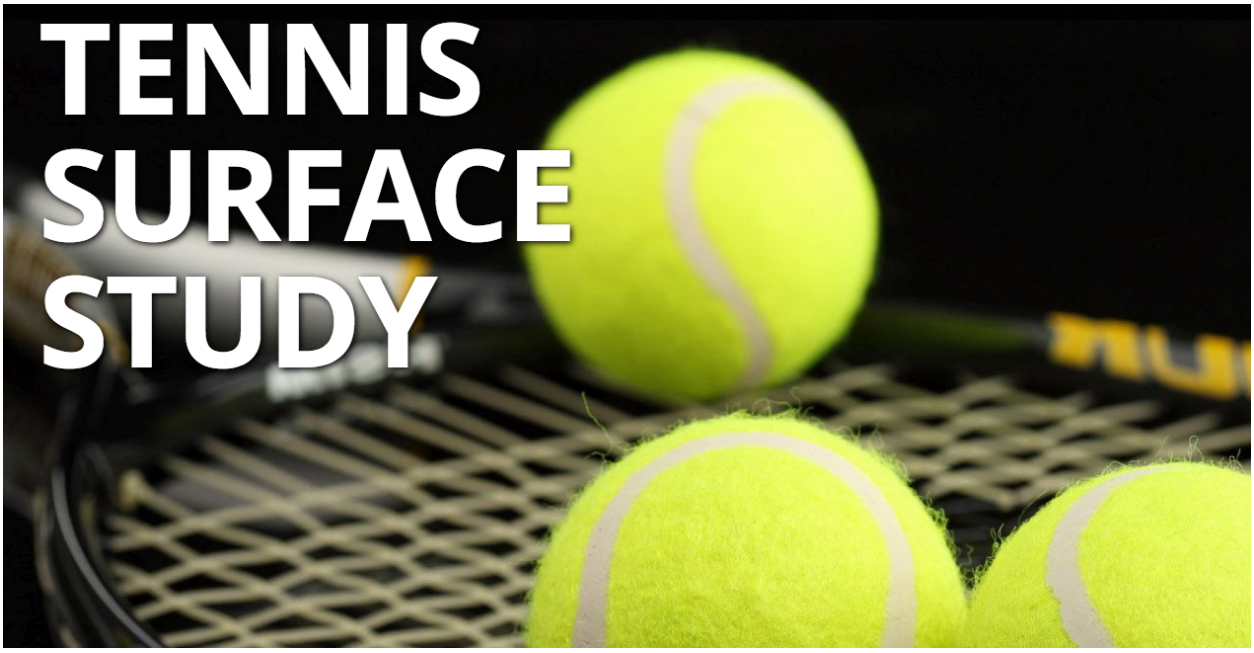


Information Visualization and Presentation

Spring 2016

Kinshuk, Proxima DasMohapatra, Puneet Sharma

Project website: http://people.ischool.berkeley.edu/~kinshuk/tennis/

Code repository: https://github.com/kinshuk-16/tennis

Overview

Tennis sport across the world revolves around four major grand slam tournaments that are held throughout the year. Each tournament is played over two weeks starting with Australian open (mid January), followed by French open (May - June), the Wimbledon (June - July) and ending with the US Open (August - September). The best players in the world take on each other in their respective categories like men's singles, women's singles, men's, women's and mixed doubles. One of the most interesting aspects of the game of tennis is the court surface itself and how different players adopt different playing styles in order to increase their chances of winning. So for this project we decided to present a study of these three main surfaces, Grass (at Wimbledon), Clay (at the French Open) and Hardcourts (US Open & Australian Open) and various metrics that highlight a certain aspect of the game.

Goals/Introduction

High level goals :

The main goals we aimed to achieve through the project are

1. Inform the viewer about the difference the surface affords on the playing style and strategies of the players.
2. Provide concrete metrics to measure such differences.
3. For someone new to the sport, give an introduction of the major tournament and surfaces.
4. Test out our hypothesis about the effect of surfaces on different aspects of the game.

How we went about achieving them:

Since the target audience of this project are tennis followers, enthusiasts as well as general sports news readers, we broke the main tasks that the interface containing the visualization supports into four main sections:

1. Overview:

The overview section provides a brief introduction to the game of tennis. It allows the user to quickly glimpse through the main four grand slams and grasp the concept of the three different surfaces on which the game is played.

2. Know the Surfaces:

Using the pre-attentive property of color, we wanted to make sure that an association is well established between the surface type and a corresponding color. This would allow the user to easily process the quantitative as well as qualitative information with respect to the court type. We decided orange for clay, azure for hardcourts and green for wimbledon, since these colors are used in sports literature to depict the surfaces.

3. Shift towards hardcourts:

Since hardcourts are most easy to maintain at a very low cost compared to the other surfaces, it's the most prefered surface of many tennis associations across the world. An increasing number of national and international tournaments are now held on hardcourts. So we implemented a cartograph using D3 to showcase all the major tennis tournaments color coded with the type of surface that is being used.

4. Exploratory Data Analysis with Tableau:

The change in playing strategy is more of a qualitative aspect because there aren't any changes to the rules or guidelines between these tournaments. Therefore, in order to illustrate the changing tactics we relied on different quantitative metrics such as number of Aces, Length of Rallies, Serve Holds and number of Net Attacks. We wanted the user to not just be able to view but also interact with the charts and explore information on each aspect.

5. Linking and Brushing:

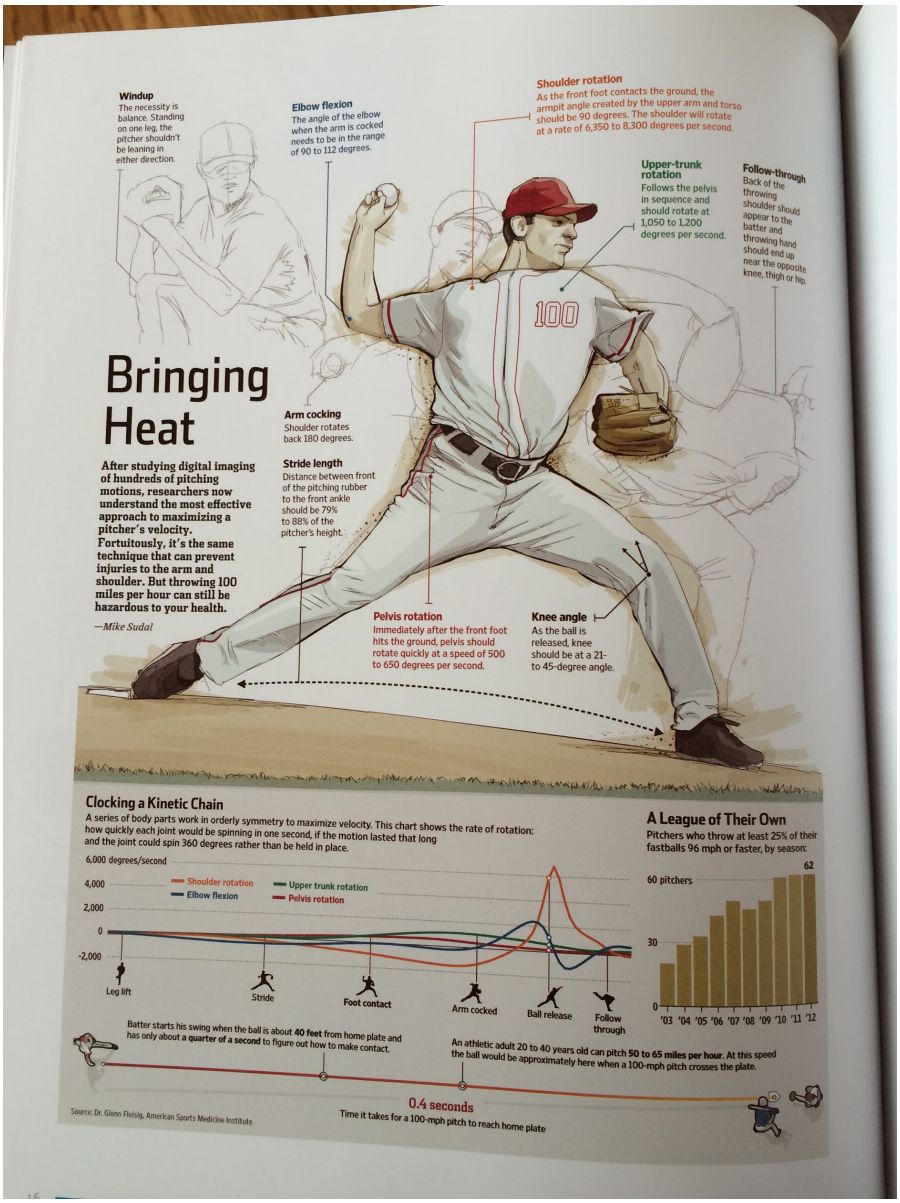
Each of the four main sections, aces, rallies, serve holds and net points were at an aggregated as well as match by match level. Tableau charts allowed the users to brush and link between the two views and discover information.

Related Work

In order to come up with the focus of the project we researched different types of visualization that already existed in field of sports and tennis in particular. After we decided to focus on surface factor of tennis we referred to articles, research papers and visualization specific to the surface factor. We also referred to different D3 visualizations in order adopt in our project.

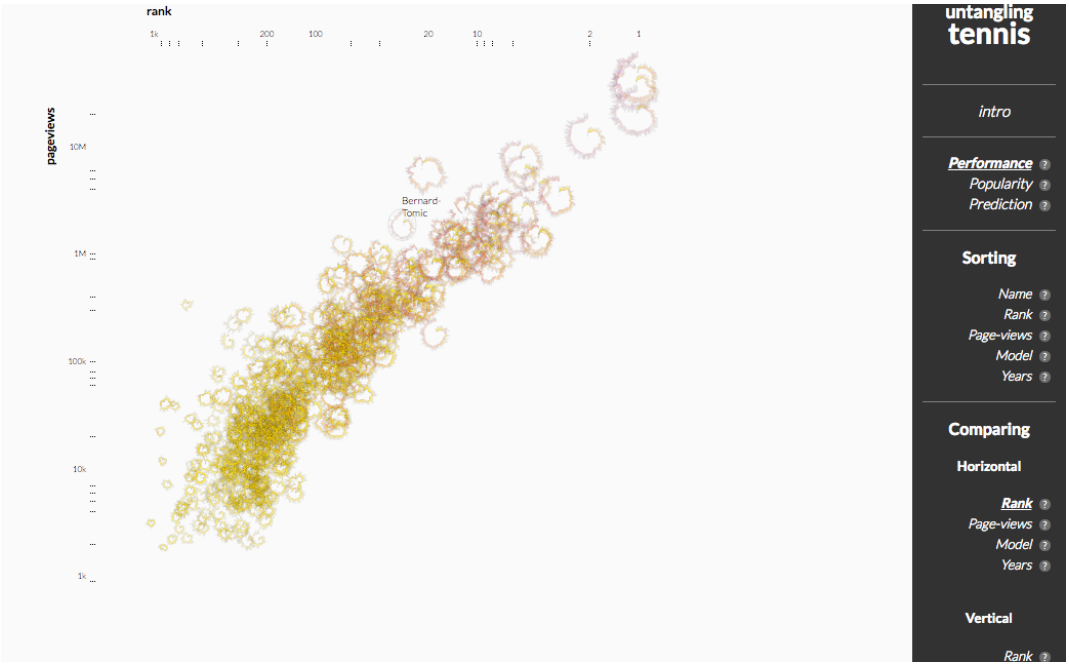
General research on sports and tennis visualization

- **Baseball infographic from book "The best American Infographics 2014"**



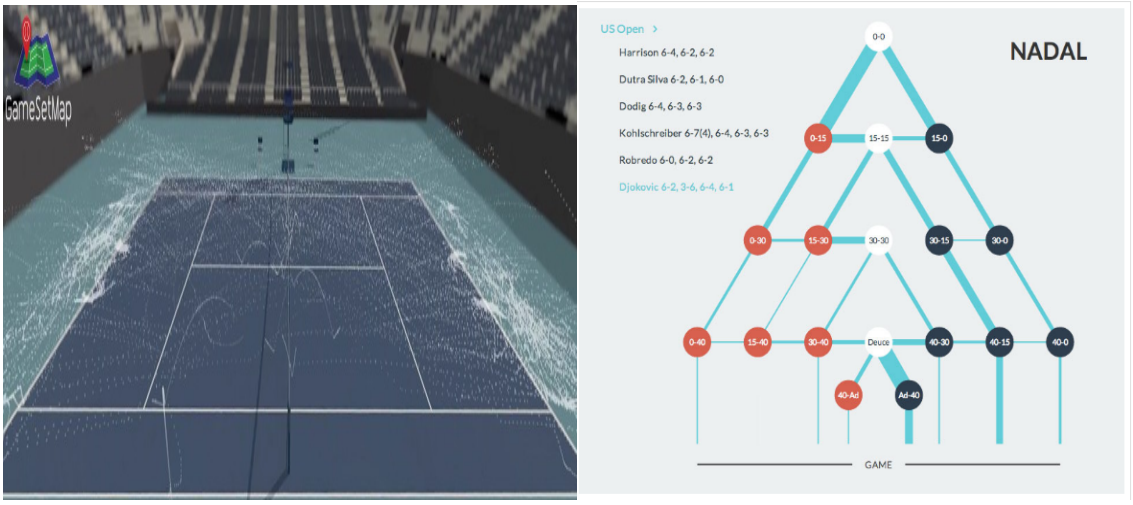
This gave us the idea to employ an infographic to clearly explain the concepts to viewers in the beginning. We were also inspired by the good use of player silhouettes.

- **Untangling Tennis**



The visualization tool allows us to explore the relationship between tennis player's popularity and performance. We were exploring this visualization prior to deciding the focus for inspiration on ideas to focus on. We had also initially planned to structure our site similarly where the visualization appears in the left and list of metrics would appear on right and viewers could toggle through the metrics. Though in the final design we differed significantly from this design.

- **Hawk eye analysis and game tree**



The page shows two main visualizations. One is hawk eye data of player movement and other is tree map of points scored in a game. In the early stages of project idea development, we explored

the possibility of doing these visualizations. The hawk eye analysis would've been a great to strengthen our explanation of how the quantitative metrics such as net attacks and the number of aces are an indicative of the a certain playing styles. But we didn't have all the data we needed, especially for all the three types of courts (or the four tournaments) to be able to show that information.

Specific work on surface

- **The effect of surface change in tennis**

  The article talks about differences in performance in tennis due to change in surfaces. It goes on to explain how change of playing surface from one match to next affects the performance of players. Although the analysis is aimed at betting market, it does a good job of setting out a few points about the relationship of surface and performance.

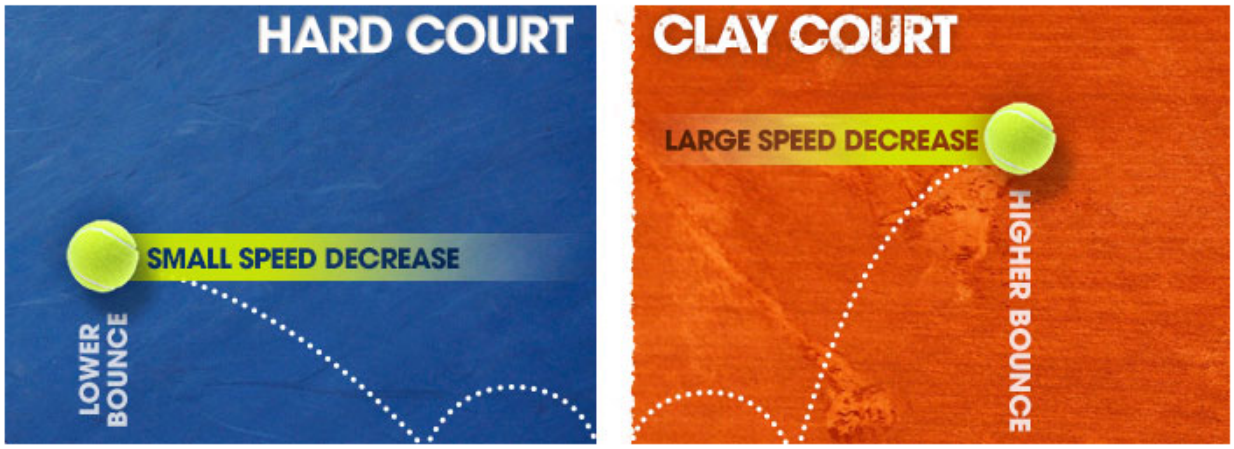
- **Surface Speed**

  The blog uses data to show the speed difference in different tennis surfaces. It has many articles about the speed convergence in different surfaces, speed report with data of every surface. It also provides raw dataset it used to come up with the hypothesis and analysis of all the point it makes.

- **How tennis court surface effects player performance and injuries**

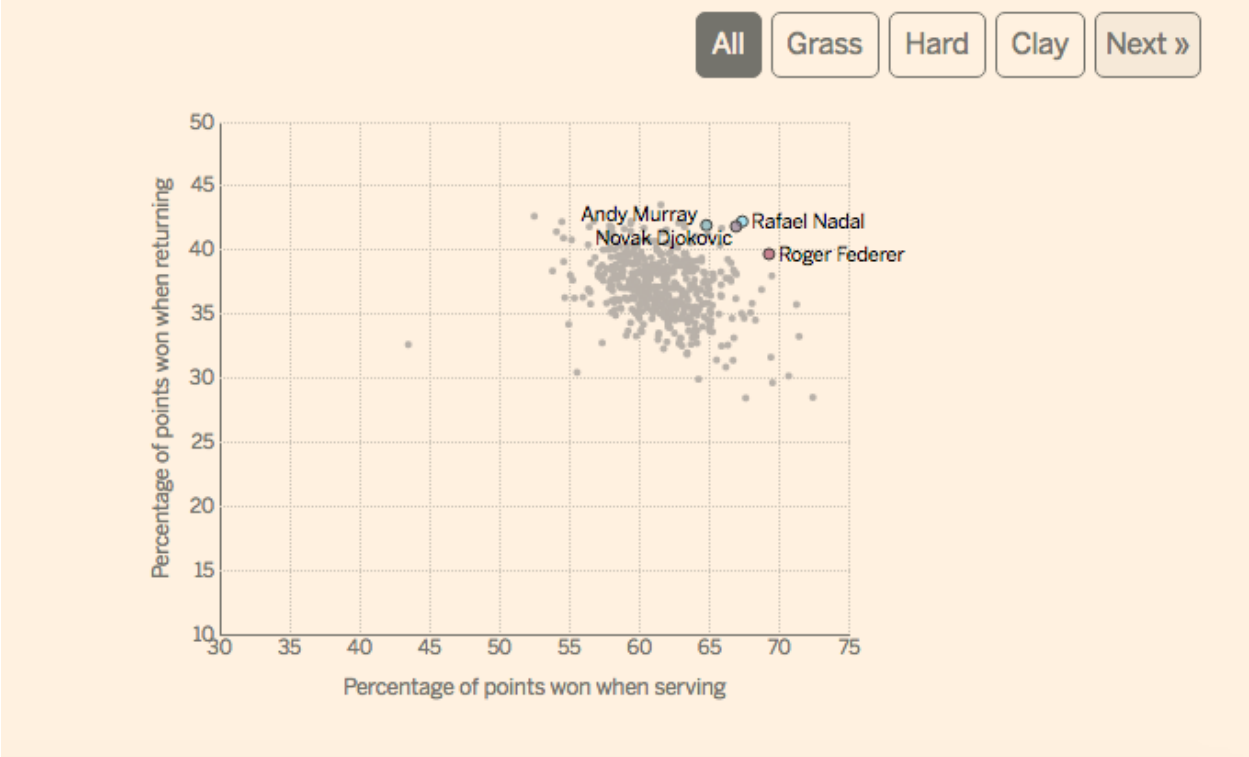
  This paper published by Tristan Barnett and Graham Pollard talks about player performance in different surfaces. It uses different matrices to rate surfaces. These matrices are matches won on a surface, first serve won, surve point won, aces , break point conversion ,net approaches and injuries. We used this paper as basis for selecting the matrices that we would highlight to distinguish the effect of different surfaces in our project.

- **How do clay court affect players at French Open?**



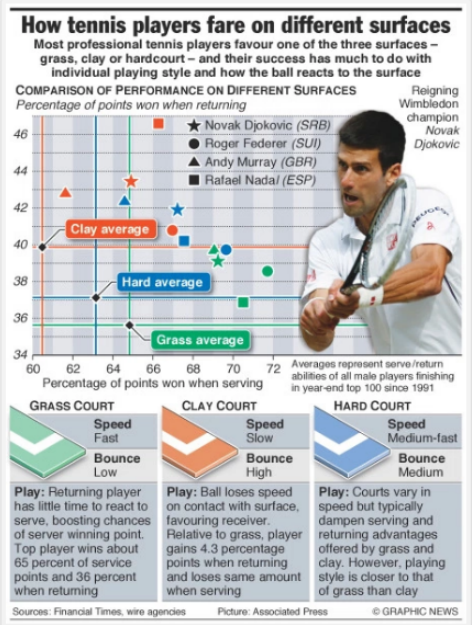
This article deals with the topic of surface effects from a betting perspective but manages to give a very clear explanation of the ball behaviour on different tennis surfaces. It highlights how aces, serve points and return points are affected by characteristics of surfaces. It inspired us to show the speed and bounce via similar ball bounce infographic.

- **Surface Tension in Tennis**



The visualization from Financial Times was one of our early inspirations for the project. It uses D3 to show the relation between points won while serving and points won while returning on

different surfaces. Our initial designs revolved around this design but ultimately we ended up changing our final designs. We did however decide to show serve points as one of the metrics but could not include break point as our dataset did not have the information.



- **How tennis players fare on different surfaces**

This infographic greatly inspired our own infographic in the project. The graph in the above infographic illustrates how the average serve holds vs returning points vary by each type of surface. Just like in this infographic, we decided to show different cards for the three major surfaces, accounting for their dominant properties.

D3 work referred

- **Parallel coordinates**

This visualization was used in the end to tie all the concept shown in visualization together.

- **Map**

This visualization was used to show tournaments played on different surfaces across the world.

- **Other D3 visualizations**

While finalizing the focus of the project we tried out different D3 visualizations to show tennis data. Initially we planned on showing tournament draw data. We created a proof of concept viz using radial tree and collapsing tree.

Data

Data was used from Jeff Sackmann's repository of scraped tennis data from 2011 onwards.

Initially we began developing the idea of doing a tennis visualization with only with 2013 grandslam data. As we explored further we found a similar dataset for all grandslam tournaments beginning from 2011 to 2016. We decided to exclude 2016 as only one of the four grand slams have taken place so far. The data was web scraped, and therefore had various inconsistencies and missing values. The data was also very detailed and had various levels of abstraction. There was the most detailed game level data, to set level, to match level and finally tournament level and year level. For most of the visualizations we aggregated the data to bring it to match level detail as any more detail was not necessary for the purpose of our project. We also took averages in our visualization instead of sum to account for some missing match information interspersed in the dataset. The data was also present separately for both players in the match, we decided to aggregate this as we were more concerned with match level information and not player specific. The dataset for every tournament in each year was divided into two files, one which had a codified the matches and named the players, other which had actual statistics. For all visualization we had to combine these files.

Summary of the Data:

No. of Tournament Per Year: 4 (AusOpen, FrenchOpen, Wimbledon, USOpen)

No. of Years included in the dataset: 5 (2011, 2012, 2013, 2014, 2015)

Total No. of Tournaments: 20

Average No. of Matches Per Tournament: 160

Average No. of Points Per Tournament: ~27,000

Tools Used:

I. Data Cleaning

Each of the team members used different tools for data cleaning, ranging from Python tools such as Pandas and regular expression libraries, to spreadsheets (MS Excel), and also MySQL database so that SQL queries can be fired to extract information. While some of the information was relatively straightforward such as the number of aces for each of the players, but for information like serve holds, net approaches and net point won, we ran SQL join queries between two tables to extract data in the format we wanted.

II. Exploratory Data Analysis

Most visualizations in the webpage have an exploratory aspect it, in which the user has the flexibility to change filters/views on the data. Most of the visualizations are rendered using Tableau, however the world map that shows the spread of surfaces is rendered using D3.

III. Infographic

We use adobe illustrator and photoshop to create the visualizations for the introductory section on our web page. Initial the visualization was designed to be pasted as a large composite image directly on the webpage (see appendix). However, while implementing the design we realized it was easier to assemble multiple images of specific dimension using CSS on the decided layout compared to one large image that would also increase the load time for the web page. We exported all the illustrations individually to be placed at their specific positions.

IV. Website

The website was created using HTML and CSS. It was developed using a framework for parallax scrolling and responsive design.

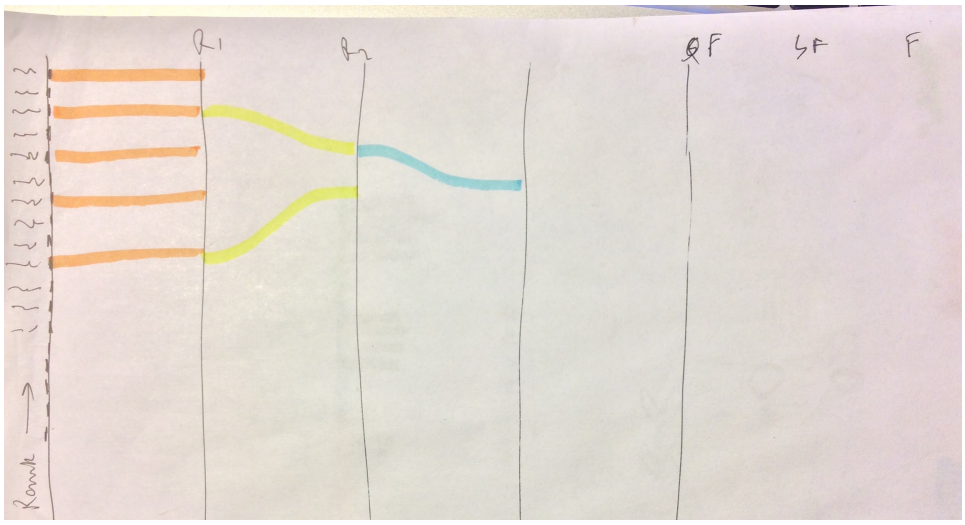
Steps taken to accomplish goals

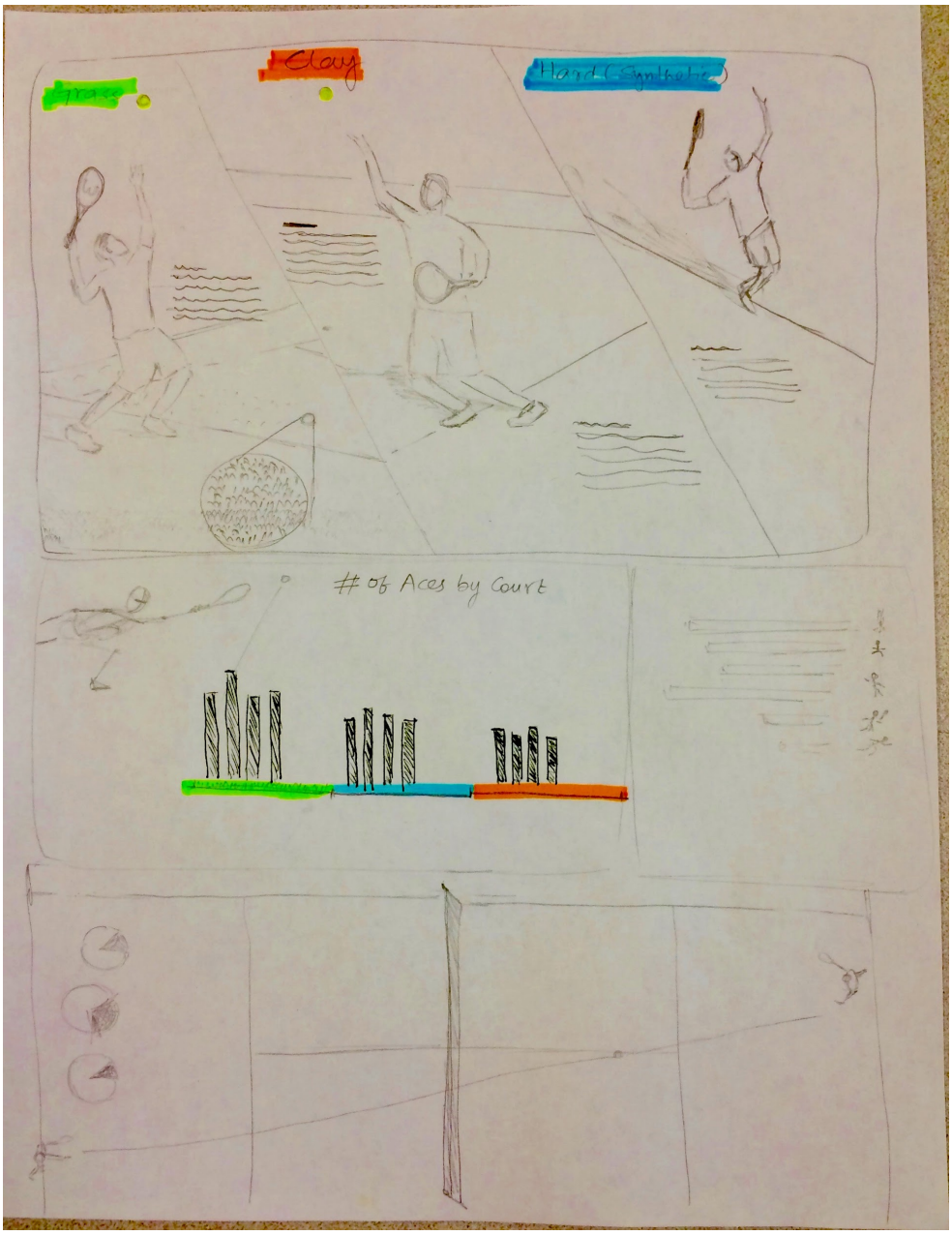
I. Introductory narrative infographic:

1. Initial concept sketches:

   To quickly test out our ideas for the introductory visualization we came up with rudimentary sketches. This not only helped us towards our final concept but also informed our subsequent EDA and D3 sections. Below are some of the initial sketches at the time of Mid-Final-Project Presentations:


Murray
Djokovic
Murray
final
Semi- final
Quart- final


R1
R2
QF
SF
F
Rank

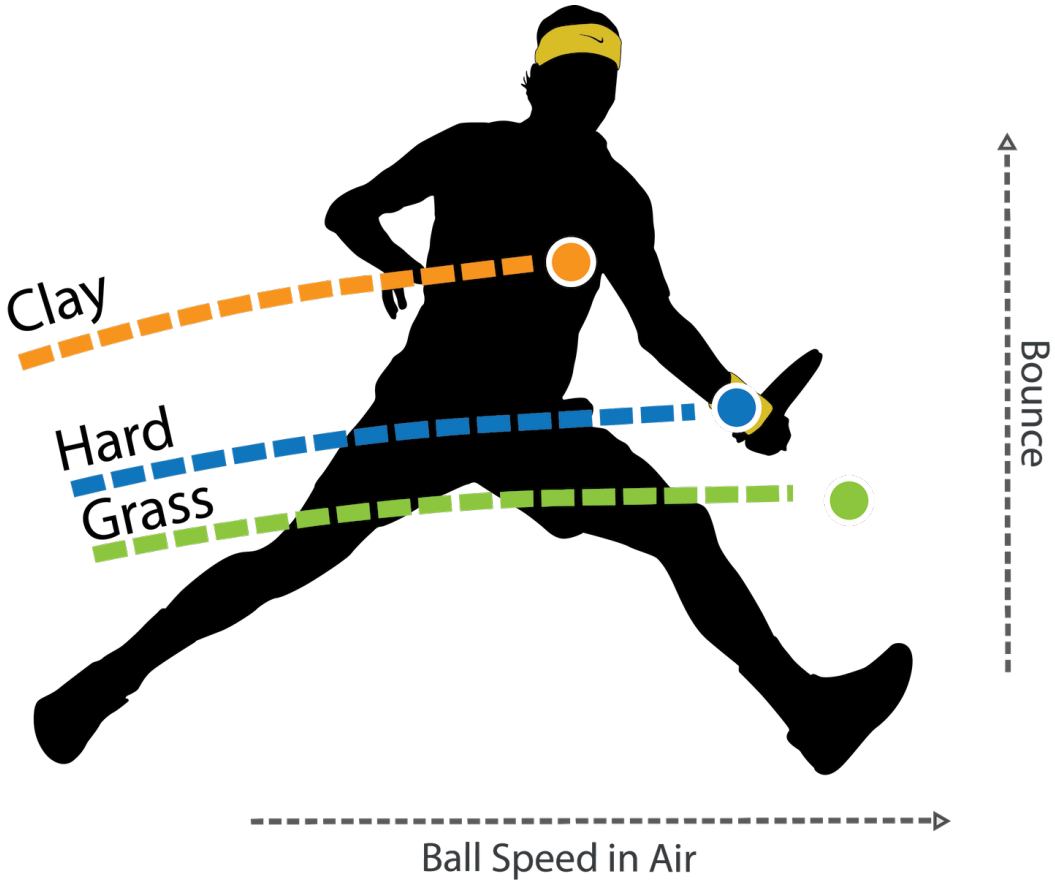


2. Development of the overview, surfaces, and bounce differences:

   Based on the feedback we got from our partner team, we changed the layout of the introductory section so that even if the user is not a tennis follower, they may still be able to grasp information about the four grand slams and the surfaces without much effort.

We carefully selected the color tone corresponding to each of the court surfaces using the color picker tool in Illustrator and then making some modifications to it. Adding player silhouettes in the background help convey the element of action in the illustration.

The main illustration in the introductory section was the bounce differences. After going through many variations of side by side court comparison, we decide to instead depict a player in his stride about to reach for a forehand. And plot the relative trajectories of a ball according to the court type. Important thing to consider while looking at this infographic is that the magnitude of bounce for the three surfaces is relative and the representation is more qualitative in nature. Though, the actual bounce is proportional to the coefficient of friction for the surface and angle at which the ball meets the surface, but the illustration captures the overall nature of the three courts.



Though initially the above illustration was oriented in the opposite direction (balls were flying from right to left). However, swapped the direction after getting a very good feedback during user testing with one of our classmates (Ganesh). He pointed out that since most of our audience are going to be left to right readers, so reversing the direction of the balls may give them a better reading of that frame.

3. Final Modifications: In the end all of the illustrations were individually extracted and were displayed on the webpage. The layout was decided keeping in mind the color associations of each surfaces as well as principle of proximity so that all related information is consumed together. See Appendix to view the final outline.

II. Surface distribution of tournament

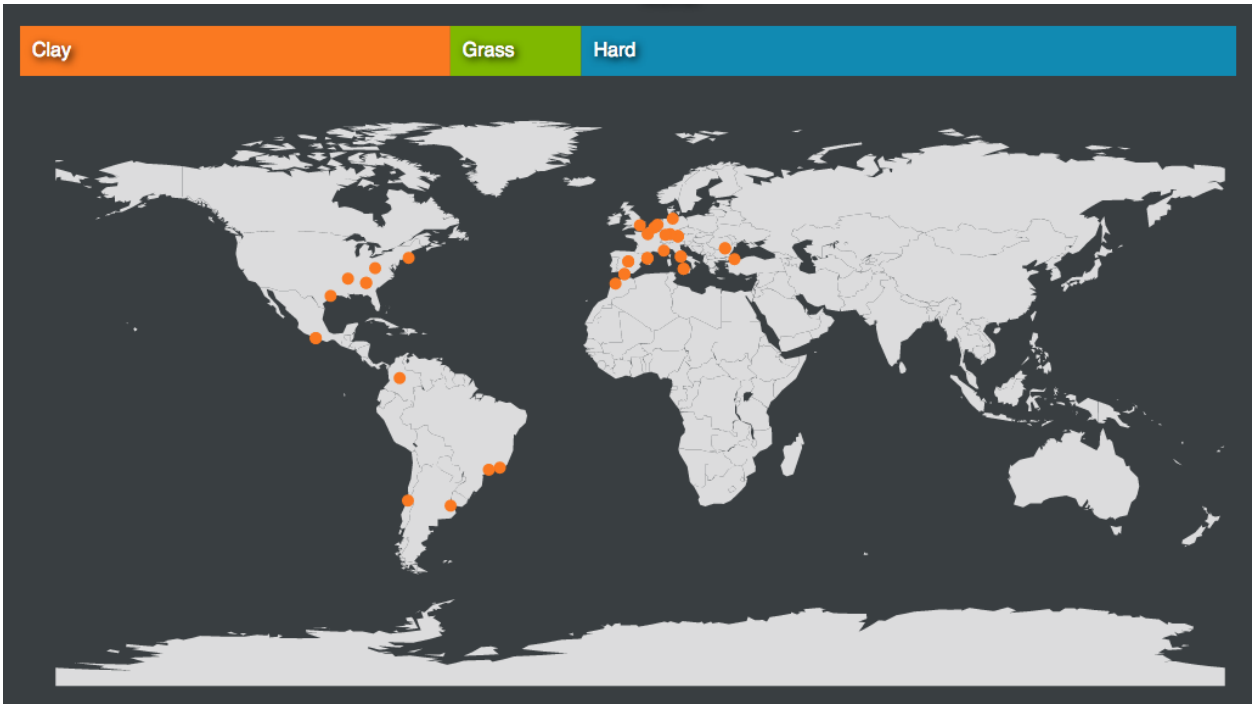
Through this visualization we wanted to give user an over all view of tournaments played on different surfaces across the world. We wanted to plot the major tournaments, cities that they are played in and surfaces that they are played on.

Data:

To get this data, we scrapped wikipedia for a list of all the major tournaments played in a year across the world. We used python for scrapping the page. We recorded, the name of the tournament, town, country and continent it is played in.

Visualization:

The aim of this visualization was to show the disparity of surfaces over all tennis tournaments played over the world. Grass courts are considerably less in number, owing to higher maintenance costs as compared to Clay and Hard courts, which have relatively less maintenance costs.



We used D3 to create the visualization, which was interesting since D3 renders the world map using a source data file that has geometric coordinate information for each country, which is treated as a polygon. We started out by downloading map geography data from Natural Earth. The shape file downloaded was converted into GeoJSON using GDAL. The GeoJSON file was then converted into TopoJSON file using topojson. The file size of TopoJSON tends to be really large (~2.5MB) due to the large number of attributes in the data file. Hence the size was cut down by including attributes of interest to us, i.e. country & city names.

The resulting file had still a lot of detail, and was close to 1 MB in size. Since this would have affected the load time of webpage, the polygons were further simplified using mapshaper.org. This website allows to upload a GeoJSON file, and reduce the detail of the map. The converted GeoJSON file was then converted into TopoJSON, which D3 can use to render the map. D3 uses a plugin called topojson, which is a part of D3 framework.

The data from tennis data files was connected to the D3 map by city names, which helped render the city locations as points on the map. To provide an overall view of the number and type of courts played in tennis tournaments, a bar graph was added. This graph was linked to the points displayed on the map, making the places for a particular surface visible. Each point on the map also supports hovering, providing more details to the user, such as the name of the tournament, city it was played in, and the surface type.

As per our hypothesis, the number of grass court tournaments is very few, as compared to clay and grass courts. Also, hard courts are marginally more in number than clay courts, since hard courts are synthetic in nature, and require less maintenance than clay surfaces.

III. Aces

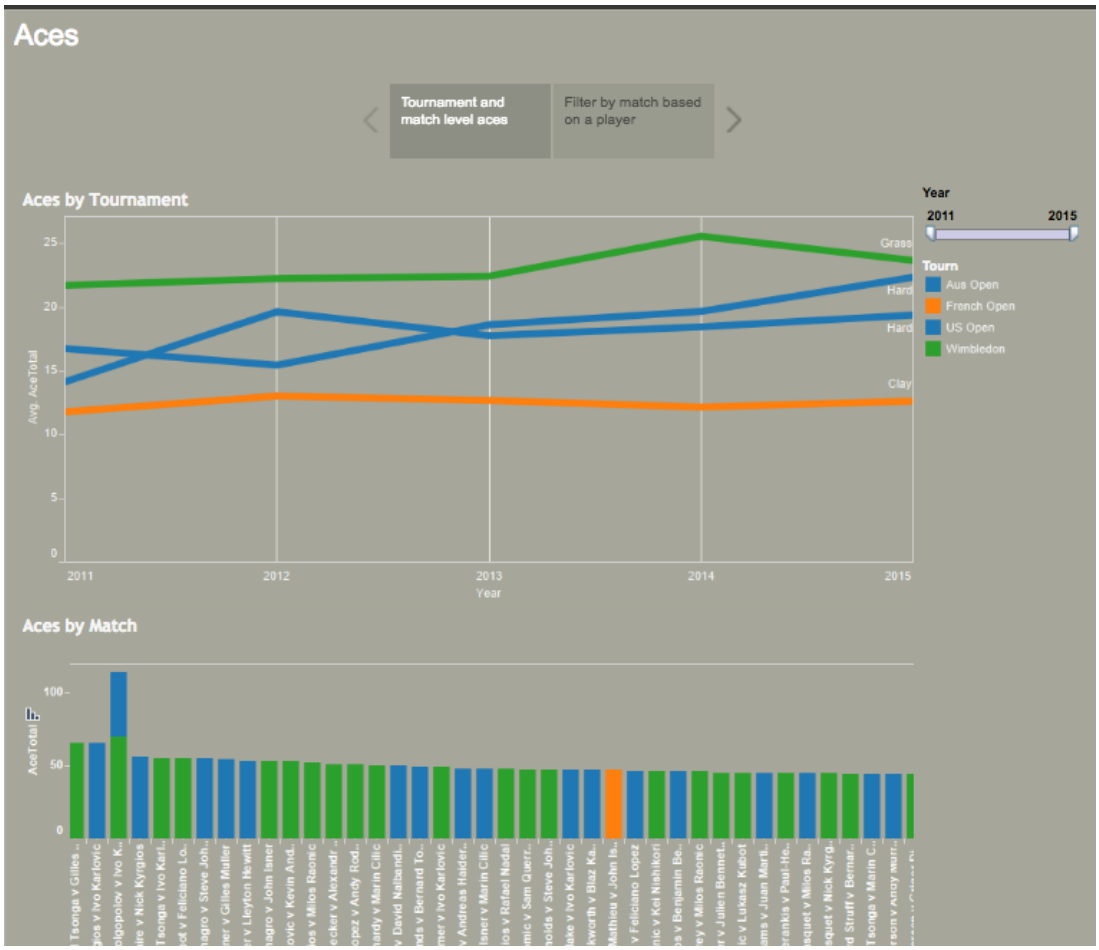
In tennis, an ace is a legal serve that is not touched by the receiver, winning the point. In professional tennis, aces are generally seen on a player's first serve, where the server can strike the ball with maximum force and take more chances with ball placement, such as the far corners of the service box. This makes returning the shot almost impossible and hence wins the server a point.

Data:

The data consisted of aces scored per game in a match divided by the tournament and year. For this visualization, the data was first aggregated by calculating the total number of aces played by each player in the match. The aces by both the players were used separately and also added up to create total number of ace per match. A match is represented by combining player 1 field with player 2 field with a "v" key word. We decided to aggregate data in this way because we were concerned with match level data and not player level. A lot of other factors come into consideration when we see player level data, like the player's performance, capability, etc. Aggregating data in this way does not differentiate between players, while maintaining the level of detail for user.Since almost the same set of players play across the tournament in a particular year, the comparison between tournament in a year makes sense as a comparison of surface rather than a comparison of players themselves. Then the data of all the tournaments in every year was combined to form one comprehensive dataset. Most of the aggregation was done using scripts written in python.

Visualization:

We wanted to explore how aces are affected by surface in tennis games. We had hypothesised that aces would be more in grass than clay. The reason behind this hypothesis is that since the speed of the ball is higher in grass, it is easier to serve a shot that is not returned. In clay, the probability of throwing off your opponent while serving is low as clay slows down the ball and gives more time for player to position himself to return the shot.

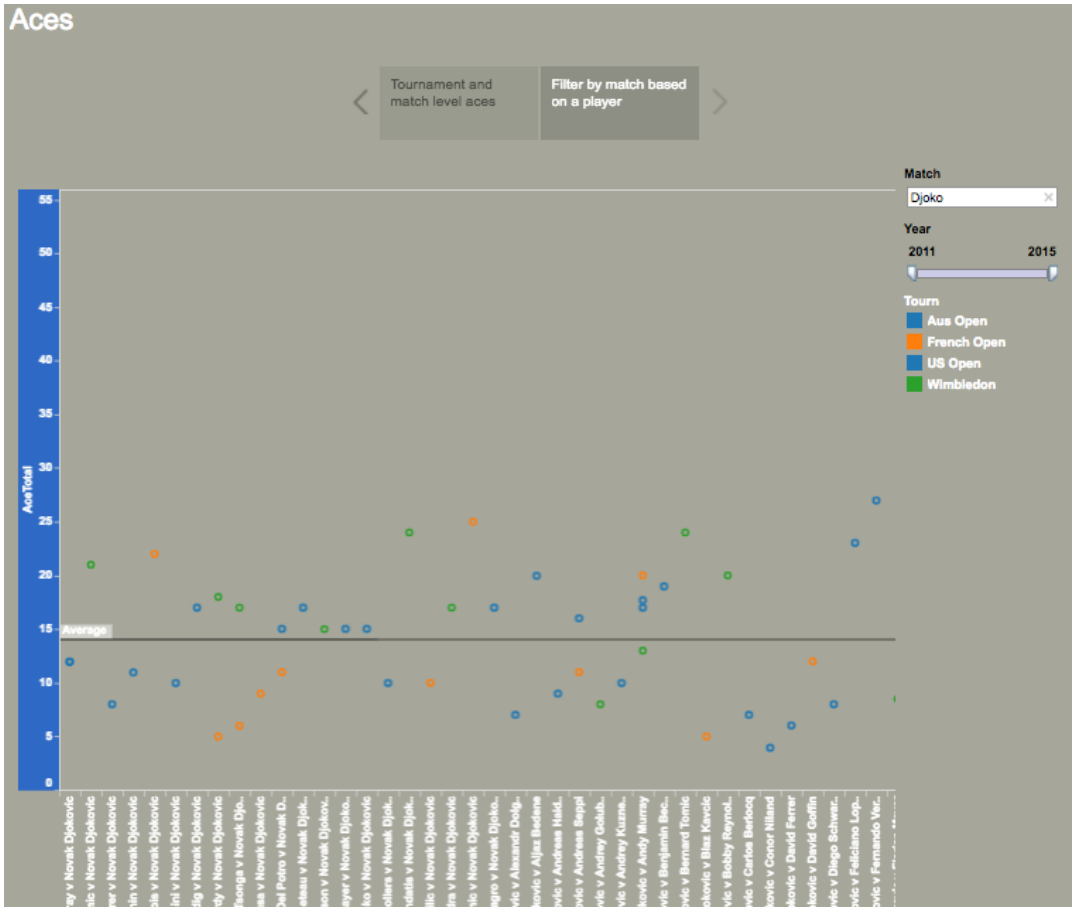


As hypothesised, we see that the average ace in each game of wimbledon, played in grass, is consistently more than that of French Open, played in clay. This is true for all the five years of data that we explored for aces, proving that surface is definitely playing a role here as players might change from year to year. There is a vast difference in number of aces between grass and clay. The tournament played on hard surfaces have more inconsistent trend. The surface of hard court are artificially created and tweaked with every year. The speed, bounce and other factors can be better controlled and designed for. The aces on the two tournament on hard surface, Australian Open and US Open, fall in between grass surface and clay surface.

The visualization was created using Tableau software. It consists of a line chart that shows the average number of aces per match in the four different tournaments. We decided to take average as in some of the year data of some of the matches were missing. In such a case sum would have been a misrepresentation of data as it would not account for the number of matches in tournament. In earlier design we had a bar chart in place of the line chart that showed the data for one year at a time and a time slider which could be adjusted to show the data for different year. We changed the design to a line chart based on feedback we got in the showcase. A line chart is better here because we can see the data for all the years in one view. It is easier for user as he

does not have to adjust the slider, also it conveys our point better as it shows that without doubt the trend is same across years. The year slider is still there as it adjusts the data for the second graph. The second graph is a bar chart which is detailed match level information. We decided to include this graph because we wanted to make a point that all match level data is not the same. It was used to account for match level differences in aces and player level difference. Every player will have varying performances based on his own capabilities ,playing style and also the opponent he is playing against. Viewer can hover on a tournament to highlight matches of that particular tournament and he could select a tournament to filter. Viewer can also arrange in increasing and decreasing order based on number of aces. Viewer can hover over a match to know the aces scored by both the players.

The color palette of the dashboard is chosen to blend in with the color scheme of the page on the site. It does not match exactly with the background so that it stands out on the page. The color palette used for graphs in visualization themselves maintain the color used to depict grass (green), clay (orange) and hard (blue). To begin with we had different colors for Australian Open and US Open to emphasise the difference but on user feedback we decided to make both of them blue because the focus here is the surface and not the tournaments themselves.



The visualization is arranged as a storyboard where the second page gives more exploratory control to the viewer. In this visualization viewer can filter out matches based on player that he is interested in exploring more about. In the above figure we have selected matches by Djokovic. His performances in different matches, in different tournament against different opponents and in different can be seen, compared and controlled by the viewers. There also an average line that shows how the number aces differ in different circumstances. Information about aces by player 1 and player 2 can be seen on over over a point.This was an experimental view, implemented to see if user would be interested in this level of exploration.

IV. Match rally

A rally in tennis is a sequence of shots within a point. A rally starts with the serve and the return of the serve, followed by a sequence of continuous shots until the point is won by either player or team.

Data:

We used a dataset that provided us with details of tennis grand slam matches on a per-point basis. Hence we had multiple rows for each match, each row representing a point in the game. The column 'Rally' contained the length of the rally for each point played in the match.

In order to generate a file with Rally statistics, we used Pandas to filter out relevant data. The resulting file had player information, tournament, year, minimum length of rally, maximum length of rally and average length of rally. The resulting data frame was as follows:

	match_id	min_rally	max_rally	avg_rally	tournament	year	players
0	2011-frenchopen-1101	0	23	3.290000	French Open	2011	Rafael Nadal v John Isner
1	2011-frenchopen-1104	0	19	3.670270	French Open	2011	Denis Gremelmayr v Nikolay Davydenko
2	2011-frenchopen-1105	0	14	3.048544	French Open	2011	Sam Querrey v Philipp Kohlschreiber
3	2011-frenchopen-1106	0	28	3.859551	French Open	2011	Ivan Ljubicic v Somdev Devvarman
4	2011-frenchopen-1108	0	16	3.490637	French Open	2011	Juan Monaco v Fernando Verdasco

The data frame was then exported as a .csv file to be used in Tableau.

Limitation: Our dataset contained incomplete rally information for years other than 2011. However, since it was an important aspect to our story, we decided to portray the information along with others.

Visualization:

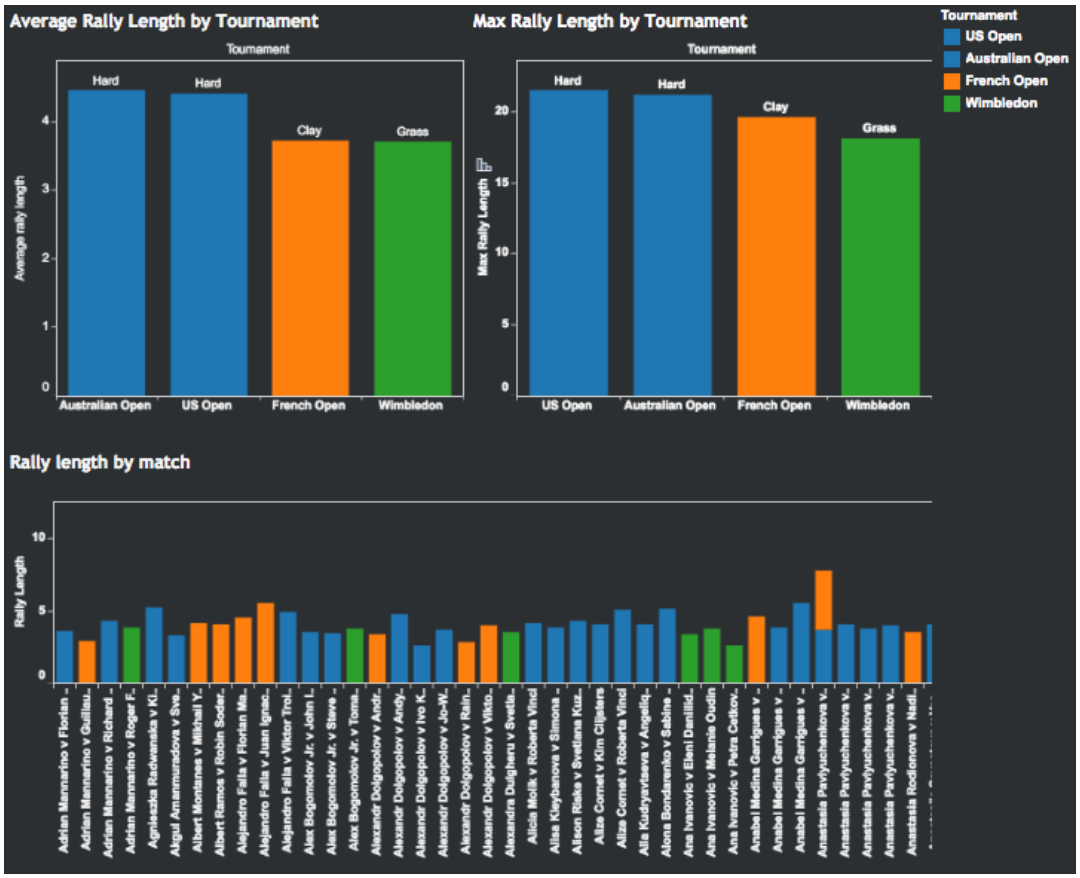
In this visualization, we wanted to explore how tennis rally length varies with surface. Basing on ball speeds from our study of tennis court surfaces, our hypothesis was that the length of rally would be shortest on grass surfaces, owing to the higher speed of the ball. Clay courts would have

the longest rally, since the lower speed of the ball would provide the returning player with ample time to return the shot. Hard courts would have rally lengths that would be in between that of clay and grass courts.

We used Tableau to create this visualization, splitting up the aggregated values to show the average and maximum rally lengths. This was done to emphasize the fact that rally lengths are also highly dependent on the player - evident from the fact that average lengths of rallies are close to each other. However, the potential of surfaces to hold rallies longer can be seen by the maximum values of rally lengths across surfaces. A detailed view for rally lengths per match was provided to a user who would be more interested in looking at rally lengths of the individual matches played across tournaments.

One of the most interesting findings we came across after plotting the values on the graph, was that therally duration distribution across the three courts wasn't the way we expected. Because clay surfaces are slower, we assumed that it would give the returning player extra time to position themselves and thereby resulting in longer rallies compared to other surfaces. After doing a bit more research we figured that as against to clay and grass, hard courts offer a very even and predictable bounce. So once a player has adjusted during the warm up, they don't get any surprises, thereby resulting in longer rallies. The 2012 Australian Open Final match between Rafael Nadal and Novak Djokovic illustrates this point very well as it lasted for 5 hours and 53 minutes making it the longest grand slam final in the history of the game.

The background color of the dashboard was changed to be harmonious with the color of the website. Since our story talks about three surface types, US Open and Australian Open tournaments were color coded by blue, which was used earlier to indicate hard surface. Colors for Wimbledon and French Open tournaments were synchronous with colors used for Clay (orange) and Green (grass) surfaces.



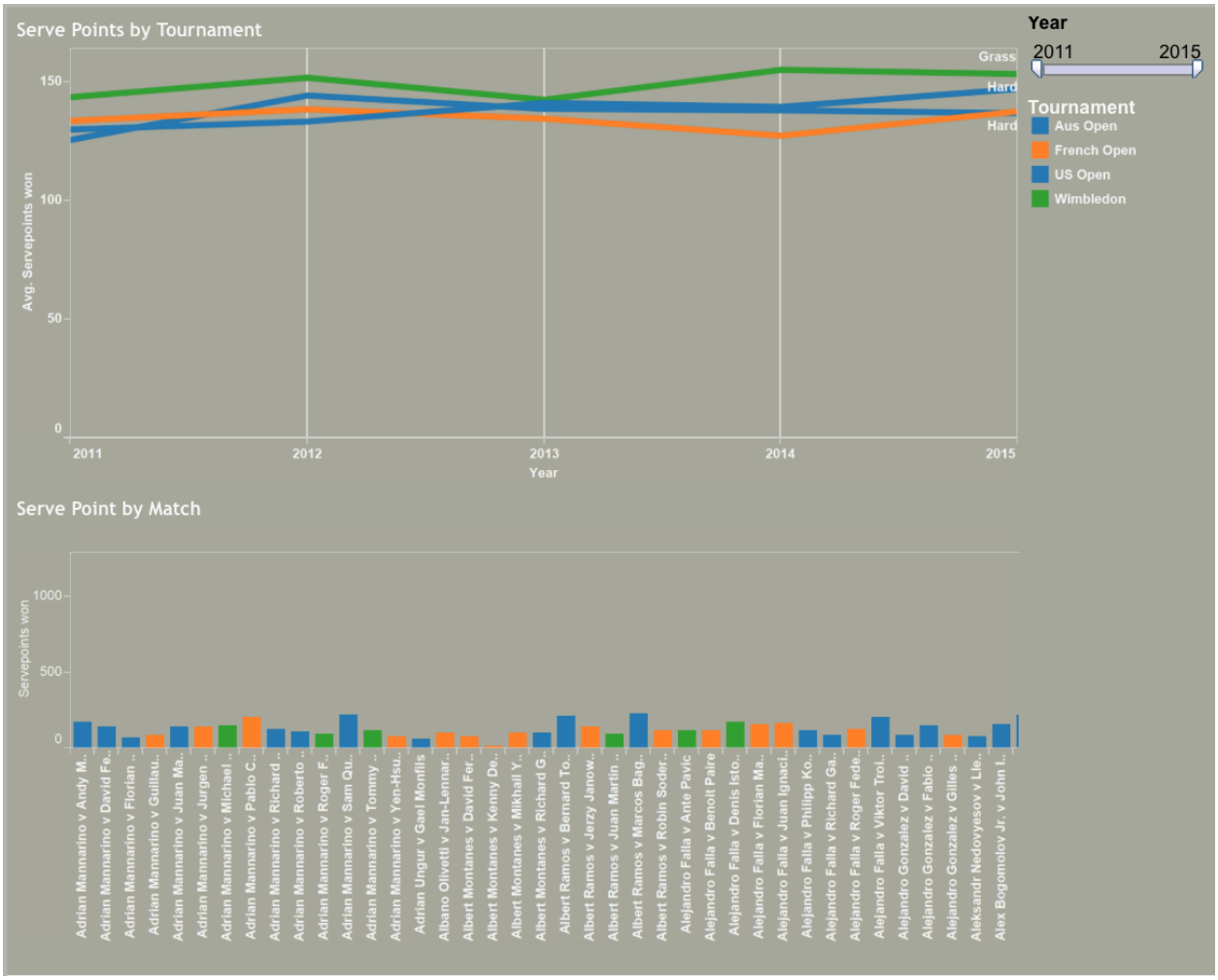
V. Serve Points Won

Serve points are perhaps one of the most important aspect of the match. At the grand slam level, the match winner is decided based on best of 3 sets (women's singles and doubles, men's and mixed doubles) and best of 5 sets (men's singles). Each set consists of at least 6 games per player. Player alternate serves by every game until the set is complete. And the set winner is decided if one of the player breaks the serve of their opponent i.e. win the game in which the opponent was serving.

Data:

In order to extract this information, we looked at two data files for each of the tournament. These files were imported as tables into a database. Table one was the match information table, which we used to extract the names of player 1 and player 2 and the match_id. We used the match_id to extract the point winner information from the second table which contained all the point-by-point information. The most import thing for us was to make sure all the data points for point winner are only counted when the point winner was also the player who was serving.

Visualization:



In this visualization we are we trying to highlight how winning points while serving happens more on which kind of court surface. The graph showcases that generally grass courts has the highest average serve points won followed by hard courts and then the clay courts. This also supported our initial hypothesis that since grass courts are the fastest, the opposite player has relatively less time to react and prepare for a return. This also leads to players using a tactic called serve and volley in which the server quickly approaches the net after serving and hits a volley in an opposite direct where the serve return came from leaving almost no chance for the opposite player to get to the ball.

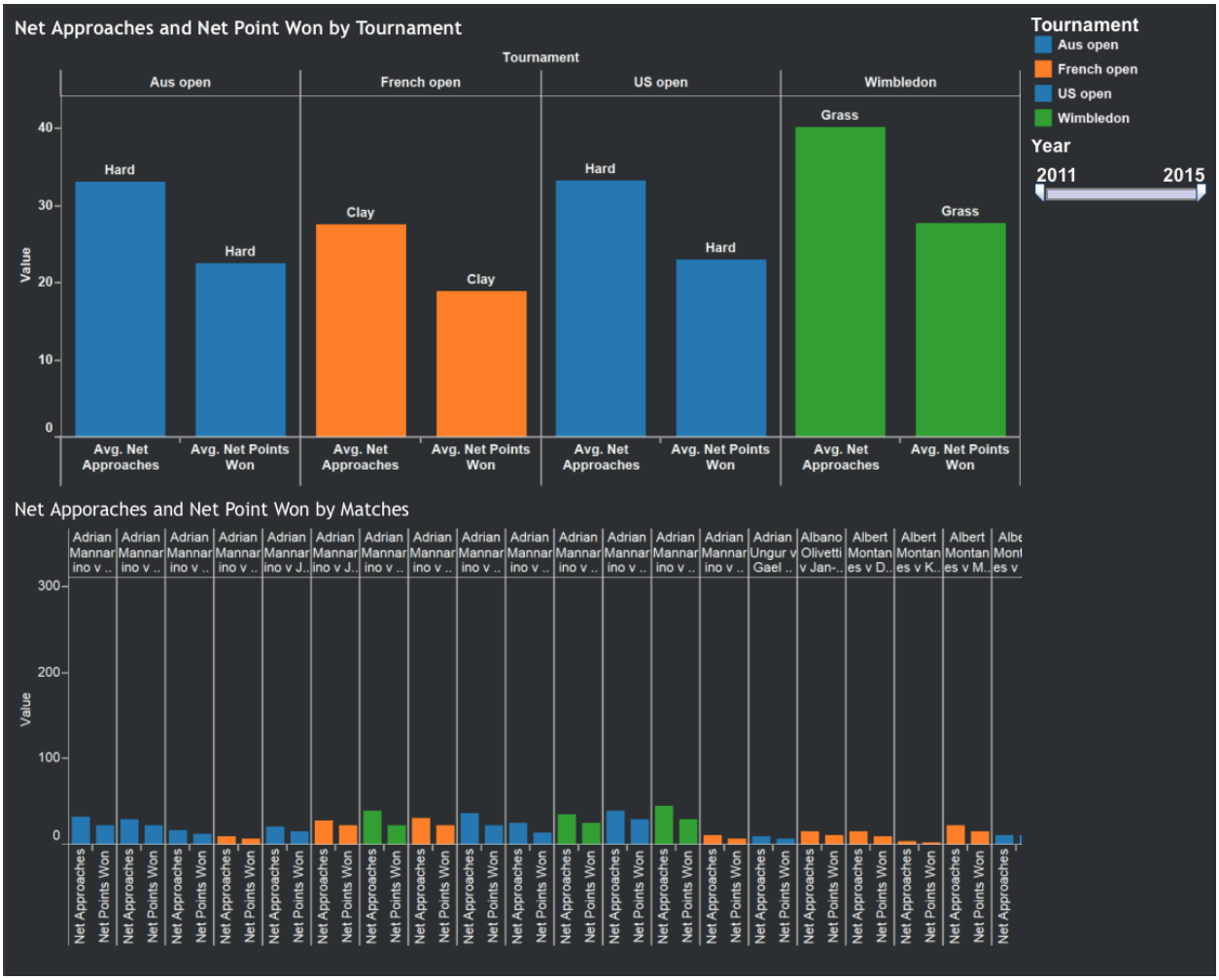
VI. Net Approaches and point

Data:

The point by point table also had information on all times a player made attempts to get to the net and also how many times they won that point.

Visualization:

In this visualization we wanted to highlight that the slow of court surfaces and deflation of balls promotes longer rallies. The good part of this is that the longer rallies means more enjoyment for the spectators, but this has also put a damper on the serve and volley style. This is because in general slower the surface, lesser the benefit of approaching net. Therefore net approaches and net points are least in Clay (French Open) and most in grass (Wimbledon).



However, an interesting aspect about the effectivity of net approaches across all these surfaces was highlighted by showing the total net approaches versus the total net points won. The data shows that in Wimbledon, players tend to attack more at the net, perhaps thinking that they definitely have more

chances of winning the net points compared to other surfaces. But looking at the results or total count of net points won, it seems as if the ratio of net approaches vs net point won may not be the highest for wimbledon among the four tournaments. And the fact it may just be equal to that of French Open, completely changes the perspective of net approaches as well as the serve and volley tactic.

Going forward, we feel that we may drill in further on this particular aspect. As it raises really interesting questions about the playing tactics that are so well established so far in the game.

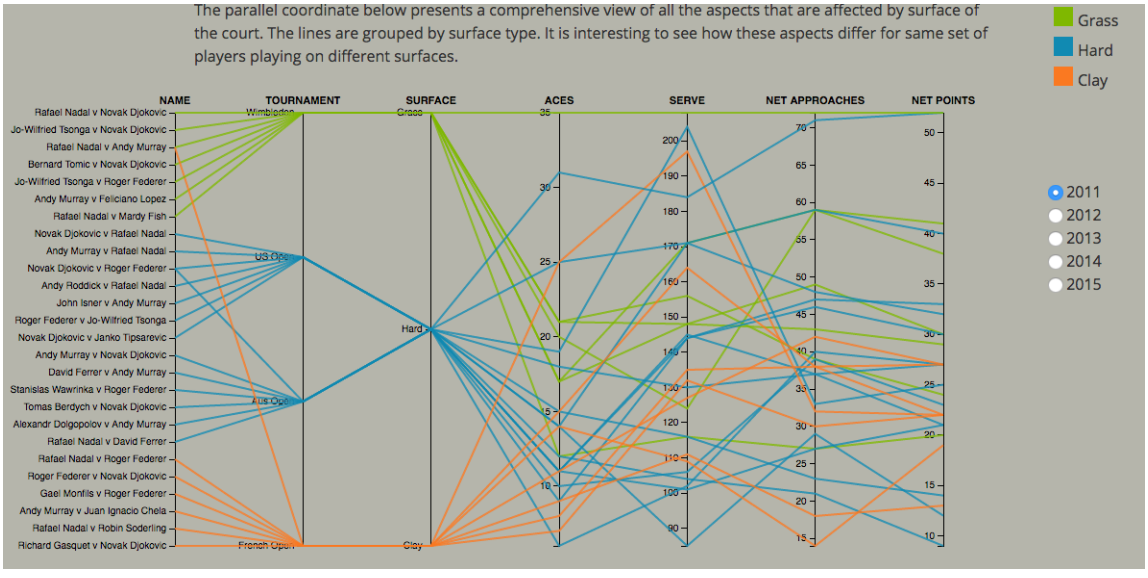
VII. Parallel coordinates

After discussing all the matrices that highlight how difference in surfaces affect the game, we bring all of it together in one frame using the parallel coordinates. The coordinates does not present any new information but serves to reinforce and recapitulate the information already presented in one view for the user to see and explore.

Data:

The dataset used for parallel coordinate is derived by compiling data from previous dashboards. The parallel coordinate has top seven matches from each tournament and each year. The reason only top seven matches were chosen was because there are four tournaments and seven matches per tournament would mean twenty eight matches to look at at once. This we concluded was the right number as any more matches would have crowded the display and made the selection on match unuseful to the viewer.

Visualization:



The visualization has seven axises, Name of match, Tournament, Surface, Aces, Serve (point won), Net Approaches and Net Point (won). It shows top 7 matches from each tournament. Consequently there are 14 blue lines, 7 orange lines and 7 green lines. The color is used to clearly divide surfaces. User can use the key to populate data from different years. The visualization provides a comprehensive view of how matches played in one surface differs from another. User can hover on a match to highlight its line path. It is also possible to brush on quantitative coordinates. We have maintained the color palette of green, blue and orange to specify, grass, hard and clay.

It is interesting to note the matches between same set of player played on different surfaces and how they differ. Form a quick view we see that most of the green lines are concentrated on top and the orange lines are more towards the bottom. The blue lines, representing hard court are more distributed across the scale of an axis. This is consistent with the trends observed in all the above visualization.

VIII. Web site and content development

The focus of the project was on the surface of tennis courts. This is not the easiest concept to present to a wide audience, many of whom may be unaware of the basics of tennis. Therefore the flow of the information and content of the site was of principal importance. We decided to begin the flow with narrative infographic which gives an introduction to a not so savvy user. It starts with describing tennis as a game of different surfaces, explaining the significance of surfaces, tournaments related to them and explaining how surface affects the ball in the game. Next we show another broad picture to situate major tournament around the world. This also gives viewer a broader view of the topic so that he gets comfortable with the idea of surfaces before diving into more detailed analysis. Then we move on the the four metrics explored in the project. The content on the top explains the what and why of the metric . The what part covers the definition and explanation of the tennis terminology used and the why part ties the concept introduced in narrative infographic to the metric in order explain how surface effects that metric. We had this part without any labels, but we decided to add these labels ,“what is it?” and “Why the difference?” based on user feedback to give more structure to the text. After showing all the metrics to compare surface we close the project with a parallel coordinate to bring together all the points discussed so far and show consolidated trends. All the visualizations are interactive and user gets to explore all the visualization.

The dominant color scheme used in site is subdued greys, interlaced with black. We decided to go with a more subdued palette because the data itself was visualized with bright colors, such as orange, light green and bright blue. The reason for selection of green, orange and blue was because they are generally associated and used with the surfaces we were focusing on. Also, it matches the actual color of the court. It would be easy for viewer to form a mental image of the courts when used these colors. We also kept the color constant across all visualization so that

viewer does not have to look at key and tournament surface relation to know which surface is which color. Also these visualizations are labeled with names of surface to reinforce the message.

It is a one page scrollable site where a alternate background colors are used to create page like structure to mark moving to next topic area. The menu bar on top allows user to skip to a particular visualization in order to explore freely if he wants. The menu bar also highlights the section of page user is on currently. Menu bar also persists on top to even while scrolling to make navigation easier.

The project website was developed using a parallax scrolling framework called cool kitten by Jalxob. Apart from the framework, HTML and CSS were used to create the website.

User testing results

User testing was done on users to understand what in the visualization work and what does not. We got several feedback with regard to the content and the visualization. Some of the suggestion for improvements and response to them are listed below.

- Lot of text in infographic part.
    - In response to this comment we tried to reduce the text in each card of the viz as much as possible. We almost eliminated all text from the tournament cards but we decided to keep some information in surface card as it was important to explain about the surface as it is our prime concept.
- Per match graph not very effective.
    - We decided to keep per match graph because we wanted to make a point that the statistics are different in different matches owing to the difference in individual player style.
    - We added an average line in the ace dashboard that allowed user to filter for a player in match and see the average between different surfaces. This makes it more relevant to the point.
- Legend for surface color with the map.
    - We added the name of surface at top of the stacked bar chart. Since the color of points on map is same as the color of the bar chart , it serves both as a stacked bar chart and a

legend for the surface. This is also enforced by linking the bars with the respective points on map.

- Ball bounce infographic direction.
    - The ball bounce infographic was initially right to left. One of the user pointed out that majority of our audience would be left to right readers, so it maybe a better idea to flip the direction of the infographic to make it consistent with the general direction of reading.
- Grand slam cards in infographic should be aligned.
    - In the initial design the grand slam card followed the surface card and were not arranged in one line leading to back and forth scrolling to look at all the grand slams. In response to this we aligned all the grand slam cards in one line at the top so that they all can be viewed in one frame.
- The color was of US open was purple that did not correspond to the color we had been using for hard
    - We had decided to keep it different to differentiate between US Open and Australian Open both of which are hard, however we decided to do away with the distinction as the main focus was showing the difference in surface and not the tournament.
- Visualization (Tableau) did not have a legend for surfaces
    - In response to this we made a calculated field for surface based on tournament and added labels with each line/bar with the name of surface.
    - We still maintain the legend with name of tournament as it is needed to link the color with individual tournament. With the labels and key together it is possible to get information about both tournament and surface.
- Have a hierarchy or structure of text with each page pointing viewer to what part is important and what should he read
    - Divided the text in two part. First talking about the definition (the “what” section) and then the affect of surface (the “why” section). Introduced sub headings to differentiate between the section. Kept the structure consistent across pages for ease.
- Changing between years does not make sense.
    - We were trying to show the consistency of trend across the year. Changing added extra burden on the viewer to change between years to view this pattern so we changed to visualization to line graph to show pattern over time in a single view.

Positive feedback

- People liked the consistency of colors for surface type across the project.
- People liked the overall view provided by map.
- Users who were aware of tennis concepts found the liberty to explore data points in detail very helpful.

Work distribution table

Task	Assigned to
Research on tennis strategies & surfaces	Kinshuk/Proxima/Puneet
Narrative infographic - research	Kinshuk/Proxima/Puneet
Narrative infographic - data collection	Kinshuk/Proxima/Puneet
Narrative infographic - overview	Puneet
Narrative infographic - Clay	Kinshuk
Narrative infographic - Grass	Proxima
Narrative infographic - Hard	Puneet
Narrative Infographic - Bounce Differences	Puneet
Narrative infographic - consolidation	Puneet
Surface distribution of tournament - research	Proxima
Surface distribution of tournament - D3 development	Proxima
Aces - research	Kinshuk
Aces - Tableau dashboard	Kinshuk
Match rally - research	Proxima
Match rally - Tableau dashboard	Proxima
Serve points - research	Puneet
Serve points - Tableau dashboard	Kinshuk
Net points - research	Puneet
Net points - Tableau dashboard	Kinshuk
Bringing it all together - research	Kinshuk
Bringing it all together D3	Kinshuk
Website development	Kinshuk

User testing	Kinshuk/Puneet/Proxima
Report	Kinshuk/Puneet/Proxima

Appendix:

Initial Infographic

Final Outline

Grass Courts

Speed: Slow, Bounce: High

Returning player has little time to react to the serve, boosting chances for the server winning point. On an average a top player wins about 65% of service points and 36% of returning points.

Grass courts reward net attacks more than any other surface. "Every point starts with a serve. On grass, many points end with a serve."

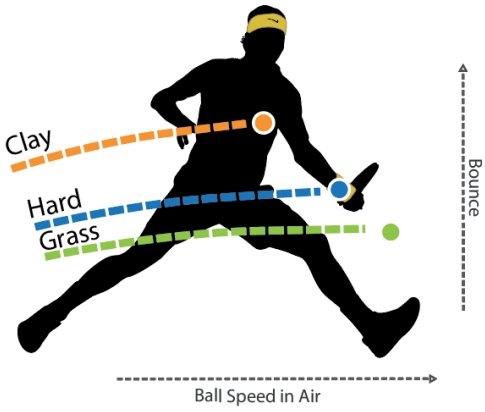
Serve-and-volleyers benefit from playing on fast courts, such as grass or fast concrete. The quick bounce and faster pace of play give them an advantage because opponents have less time to set up for a passing shot.

Hardcourt

Speed: Medium-Fast, Bounce: Medium

Courts vary in speed, but typically dampens the serving and return advantages offered by grass and clay surfaces. This is due to the sand that is mixed with the top paint surface. However, playing style is closer to grass than it is to clay.

Hard courts are generally considered to be the best surface for an offensive baseliner who often hit higher risk shots. Novak Djokovic, though known to be an all-rounder, is also a notable Aggressive Baseliner



Clay Court

Speed: Slow, Bounce: High

The high bounce and slow court speeds encourages player to hang behind the baseline and play defense, trying to force or goad the opponent into a mistake.

Offensive baseliners with height especially have an advantage on clay courts because the high bounces land in their hitting zones, allowing them to strike the ball cleanly and more powerfully. One great example for this is Maria Sharapova, who is 6'2", and the high bounce of the ball really helps her produce winners and force errors from her opponents.