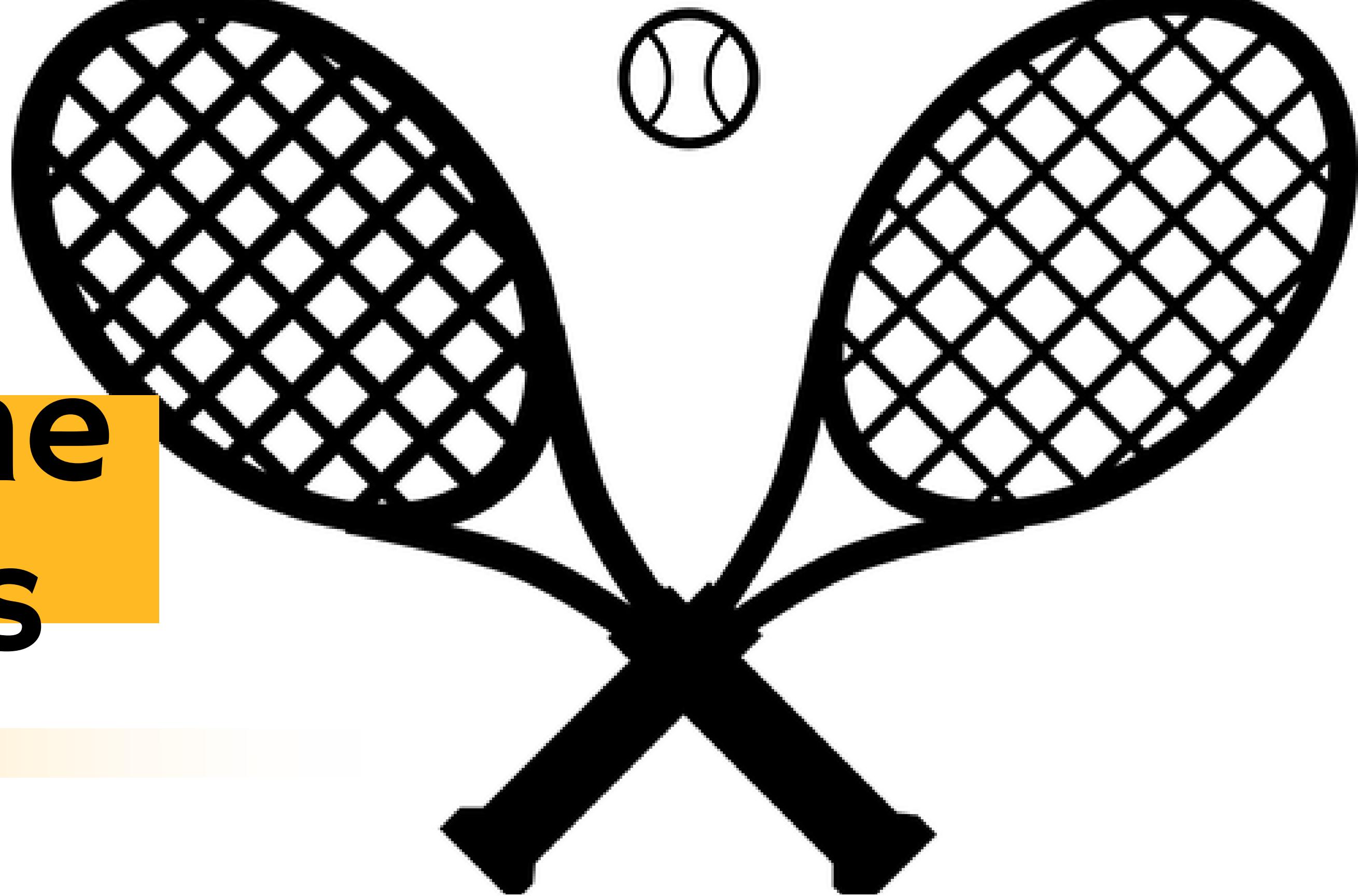




# ATP Game Analytics



Presented by: Group -2

# MEET THE TEAM MEMBERS



ANDREW MEYER



NAVEEN KUMAR



DIKSHANT JOSHI



PEGAH KARIMI



RAJASHEKAR  
REDDY VEMULA

# AGENDA

Overview of today's presentation session



- ✓ Guiding Motivation
- ✓ Project Overview
- ✓ Data Source Review
- ✓ Data Base Design
- ✓ Analytical Approach
- ✓ Analysis Results

# Guiding Motivation

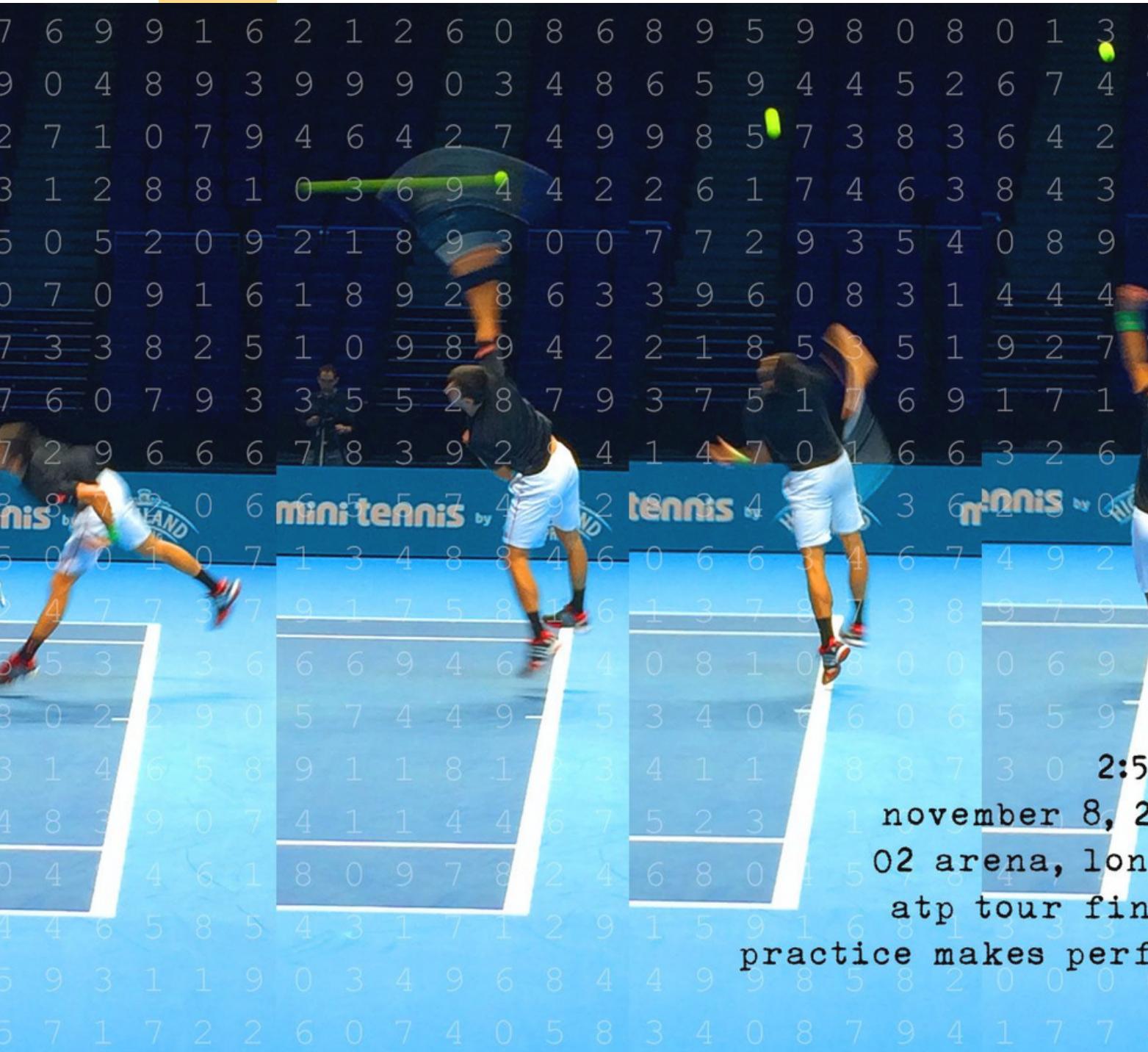
Leverage personal experience playing competitive high school tennis to understand what variables are important to a strong player.

## High School Tennis:

- Struggled against big servers
- Nerves when facing breakpoints led to more errors

## Takeaways:

- The best players always had a strong first serve allowing them to control the point
- Tournament winners always seemed to find a way to save breakpoints



# DATA OVERVIEW



## Goal:

- Create a database capable of analyzing player characteristics to help better inform betting decisions and improve betting agencies' monetary outcomes by providing statistics that will allow for more accurate odds.

## Data:

- ATP Tour match-level results from 2000-2016
- ATP Tour match-level results with betting odds from 2010-2016

## Analytics:

- Player height effect on Serve Measures
- Breakpoints saved vs. win percentage
- Breakdown of David Ferrer vs. Rafael Nadal betting odds vs outcome



# Project Tools

- ✓ AWS
- ✓ MYSQL Workbench
- ✓ Tableau
- ✓ Excel
- ✓ R studio

# DATA



- ATP match level data was found on Datahub while ATP Betting data was from Kaggle
- Years 2000-2016 were used for the ATP match table and 2010-2016 for ATP Betting data to test the database design
- 103 unique variables
- Many variables are the same statistic measured for the winner and loser
  - Ex: Winner\_firstserve\_pts & Loser\_firstserve\_points

# ATP data set with players info

1	tourney_i	tourney_r	surface	draw_size	tourney_l	Date	match_nu	winner_id	WID_DATE	winner_se	winner_e	winner_n	winner_h	winner_h	winner_ic	winner_ag	winner_ra	winner_ra	loser_id	loser_see	loser
2	2000-451	Doha	Hard	32	A	1/3/2000	12	101962	962-36528	4		Younes El R		193 MAR	28.29021				102179		
3	2000-451	Doha	Hard	32	A	1/3/2000	22	101962	962-36528	4		Younes El R		193 MAR	28.29021				103292		
4	2000-451	Doha	Hard	32	A	1/3/2000	27	101962	962-36528	4		Younes El R		193 MAR	28.29021				102025		
5	2000-891	Chennai	Hard	32	A	1/3/2000	12	102538	538-3652812	LL		Tuomas Ki R		185 FIN	24.846				102087		
6	2000-891	Chennai	Hard	32	A	1/3/2000	22	102538	538-3652822	LL		Tuomas Ki R		185 FIN	24.846				103181	LL	
7	2000-891	Chennai	Hard	32	A	1/3/2000	10	102494	494-3652810			Tomas Bel R		193 GER	25.04038				102880		
8	2000-339	Adelaide	Hard	32	A	1/3/2000	16	102450	450-36528	2		Tim Henm R		185 GBR	25.30595				102607	Q	
9	2000-339	Adelaide	Hard	32	A	1/3/2000	24	102450	450-36528	2		Tim Henm R		185 GBR	25.30595				103444		
10	2000-339	Adelaide	Hard	32	A	1/3/2000	1	102358	358-36528	1		Thomas Er R		190 SWE	25.79055				103096		
11	2000-339	Adelaide	Hard	32	A	1/3/2000	17	102358	358-36528	1		Thomas Er R		190 SWE	25.79055				103819		
12	2000-339	Adelaide	Hard	32	A	1/3/2000	25	102358	358-36528	1		Thomas Er R		190 SWE	25.79055				103206	7	
13	2000-339	Adelaide	Hard	32	A	1/3/2000	29	102358	358-36528	1		Thomas Er R		190 SWE	25.79055				102796	3	
14	2000-339	Adelaide	Hard	32	A	1/3/2000	14	101792	792-3652814			Slava Dose R		183 CZE	29.36893				103344		
15	2000-451	Doha	Hard	32	A	1/3/2000	4	102854	854-36528	8		Sjeng Sch R		193 NED	23.29911				102831		
16	2000-451	Doha	Hard	32	A	1/3/2000	18	102854	854-36528	8		Sjeng Sch R		193 NED	23.29911				102644		
17	2000-339	Adelaide	Hard	32	A	1/3/2000	4	103206	206-36528	7		Sebastien R		175 FRA	21.57974				102776		
18	2000-339	Adelaide	Hard	32	A	1/3/2000	18	103206	206-36528	7		Sebastien R		175 FRA	21.57974				102998		
19	2000-891	Chennai	Hard	32	A	1/3/2000	4	101086	086-36528	7		Ronald Ag R		180 USA	35.11841				104156	WC	
20	2000-339	Adelaide	Hard	32	A	1/3/2000	2	103819	819-365282			Roger Fed R		185 SUI	18.38467				102533		

## Betting data set

1	ATP	Locatio	Tournament	Date	Series	Court	Surface	Round	Best of	Winner	Loser	WRank	LRank	W1	L1	W2	L2	W3	L3	W4	L4
46597		52	New York US Open		9/9/2016	Grand Slam	Outdoor	Hard	Semifinal	5 Djokovic N	Monfils G.	1	12	6	3						
46598		52	New York US Open		9/9/2016	Grand Slam	Outdoor	Hard	Semifinal	5 Wawrinka	Nishikori I	3	7	4	6						
46599		52	New York US Open		9/11/2016	Grand Slam	Outdoor	Hard	The Final	5 Wawrinka	Djokovic N	3	1	6	7						
46600		53	Metz Open de Moselle		9/19/2016	ATP250	Indoor	Hard	1st Round	3 Bennetea	Klizan M.	238	31	6	2						
46601		53	Metz Open de Moselle		9/19/2016	ATP250	Indoor	Hard	1st Round	3 Muller G.	Zeballos F	37	66	7	6						
46602		53	Metz Open de Moselle		9/20/2016	ATP250	Indoor	Hard	1st Round	3 Gojowczy Almagro N	209	48	6	4							
46603		53	Metz Open de Moselle		9/20/2016	ATP250	Indoor	Hard	1st Round	3 Basilashvi	Barrere G.	106	202	5	7						
46604		53	Metz Open de Moselle		9/20/2016	ATP250	Indoor	Hard	1st Round	3 Garcia-Lo Dzumhur I	57	74	5	7							
46605		53	Metz Open de Moselle		9/20/2016	ATP250	Indoor	Hard	1st Round	3 Brown D.	Mathieu P	68	69	6	4						
46606		53	Metz Open de Moselle		9/20/2016	ATP250	Indoor	Hard	1st Round	3 Millot V.	Paire B.	162	38	6	4						
46607		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	1st Round	3 Marchenk	Seppi A.	50	92	6	7						
46608		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	1st Round	3 Struff J.L.	Robredo T	87	251	7	6						
46609		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	1st Round	3 Herbert P.	Cervantes	108	77	7	6						
46610		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	1st Round	3 Mahut N.	Halys Q.	40	154	6	3						
46611		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	1st Round	3 Jaziri M.	Dodig I.	55	89	6	4						
46612		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	2nd Round	3 Bennetea	Brown D.	238	68	6	3						
46613		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	2nd Round	3 Thiem D.	Gojowczy	10	209	7	5						
46614		53	Metz Open de Moselle		9/21/2016	ATP250	Indoor	Hard	2nd Round	3 Muller G.	Basilashvi	37	106	6	4						
46615		53	Metz Open de Moselle		9/22/2016	ATP250	Indoor	Hard	2nd Round	3 Jaziri M.	Millot V.	55	162	4	6						
46616		53	Metz Open de Moselle		9/22/2016	ATP250	Indoor	Hard	2nd Round	3 Simon G.	Garcia-Lo	28	57	6	7						

AM	AN	AO	AP	AQ	AR
EXW	EXL	PSW	PSL	Wpts	Lpts



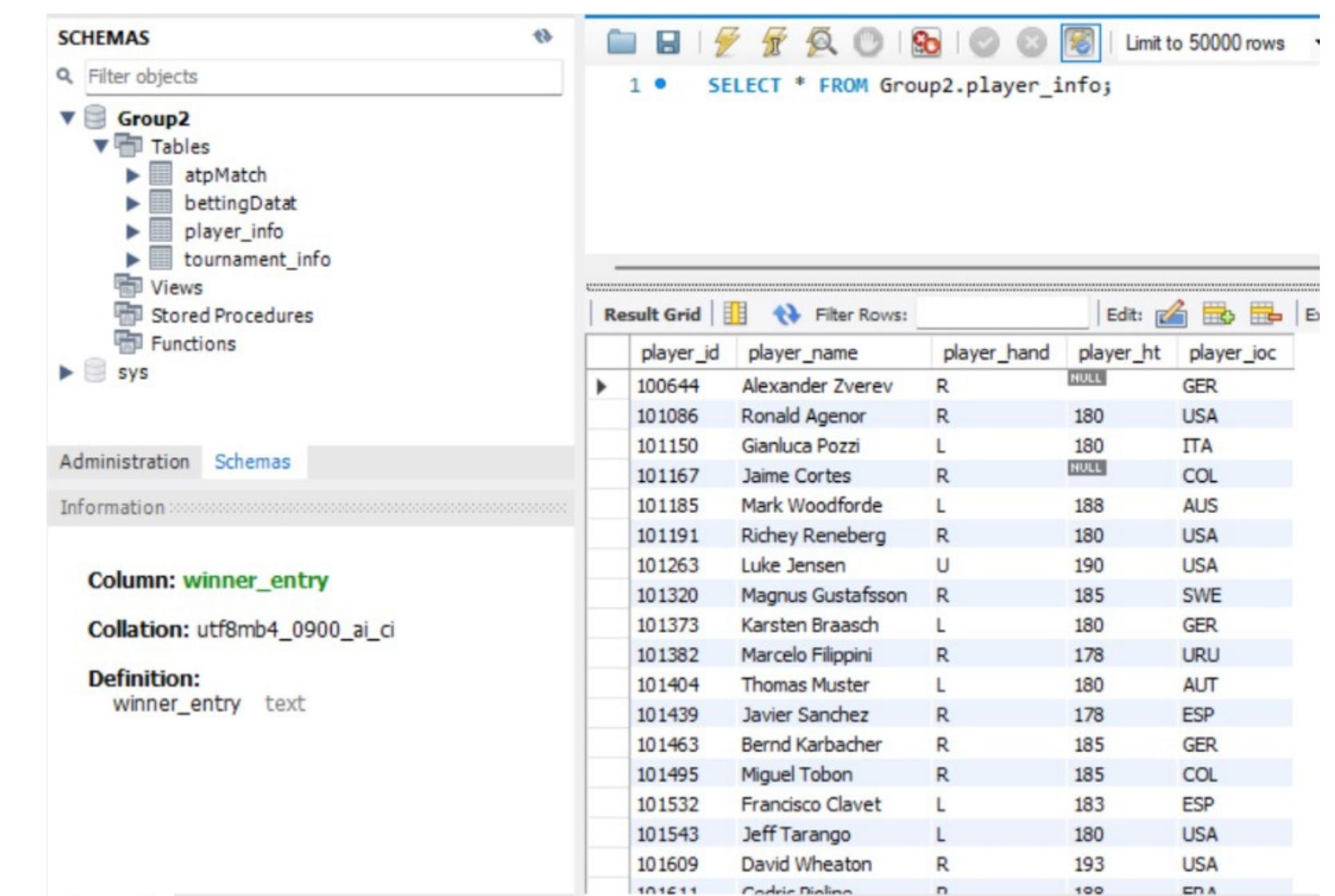


</tbl

# Normalization

## Normalization

- Originally we had ATP data set with tournament information, player information, and tennis insights.
- Using this we created a composite primary key TID\_MID (combination of tournament Id and, match Id) after creating primary key we did normalization of ATP data set into three new tables including one parent table (ATP)
- We removed the redundant rows from betting data, ATP tables which can be referred from a normalized tables



SCHEMAS

Filter objects

Group2

Tables

- atpMatch
- bettingData
- player\_info
- tournament\_info

Views

Stored Procedures

Functions

sys

Administration Schemas

Information

Column: winner\_entry

Collation: utf8mb4\_0900\_ai\_ci

Definition: winner\_entry text

Result Grid | Filter Rows: Edit: |

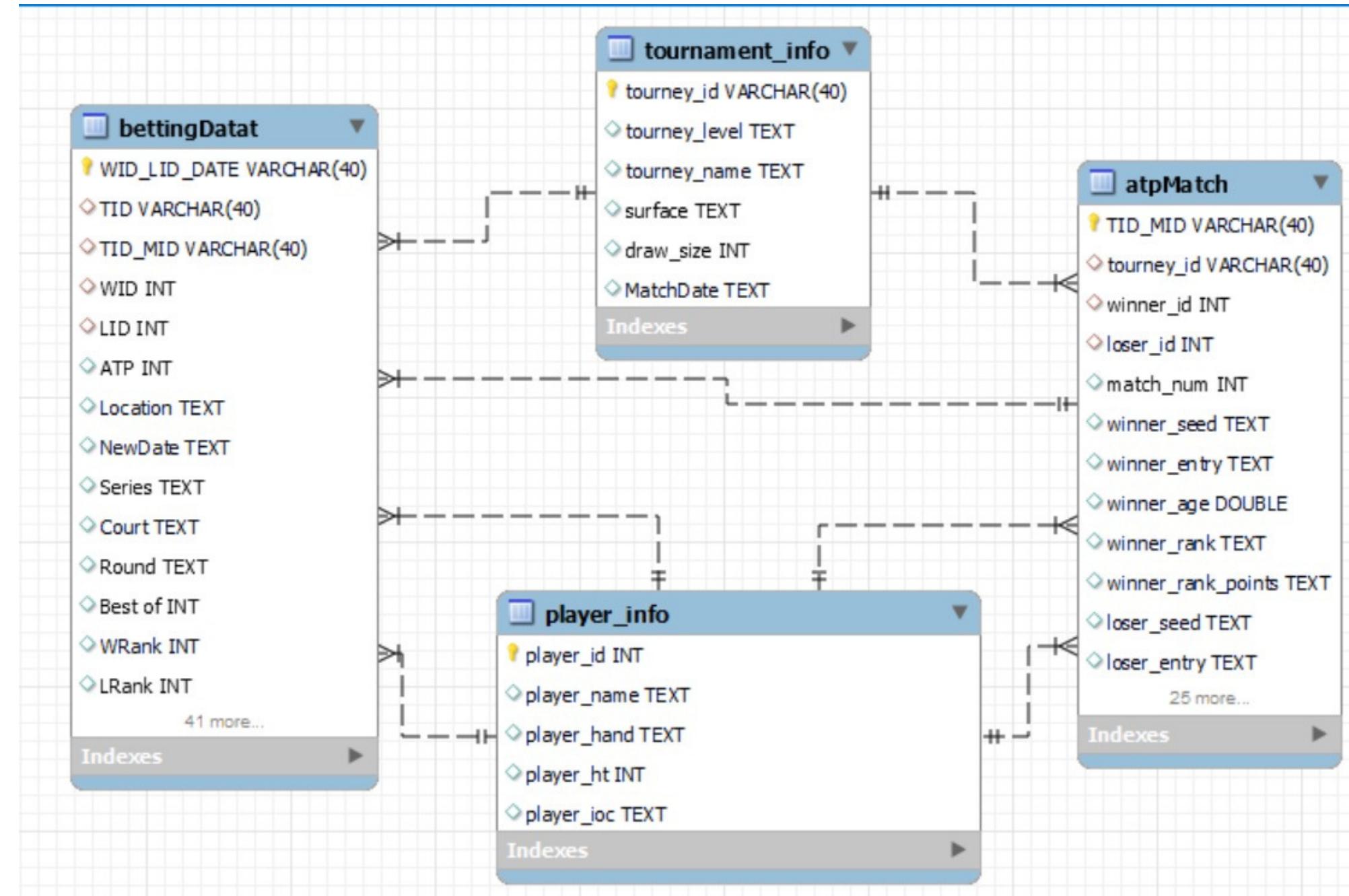
	player_id	player_name	player_hand	player_ht	player_joc
▶	100644	Alexander Zverev	R	NULL	GER
	101086	Ronald Agenor	R	180	USA
	101150	Gianluca Pozzi	L	180	ITA
	101167	Jaime Cortes	R	NULL	COL
	101185	Mark Woodforde	L	188	AUS
	101191	Richey Reneberg	R	180	USA
	101263	Luke Jensen	U	190	USA
	101320	Magnus Gustafsson	R	185	SWE
	101373	Karsten Braasch	L	180	GER
	101382	Marcelo Filippini	R	178	URU
	101404	Thomas Muster	L	180	AUT
	101439	Javier Sanchez	R	178	ESP
	101463	Bernd Karbacher	R	185	GER
	101495	Miguel Tobon	R	185	COL
	101532	Francisco Clavet	L	183	ESP
	101543	Jeff Tarango	L	180	USA
	101609	David Wheaton	R	193	USA
	101611	Cedric Pioline	N	190	FR A

# ERD

- We created ERD for the normalized tables in MYSQL Workbench.
- The ERD clearly explains the primary key, foreign key and the relation of the tables.

◆ Primary key

◆ Foreign key



# Ref images for data dictionary

## Data Dictionary

- We created a data dictionary for the normalised tables

table_nm	obj_typ	ord	is_key	column_nm	data_typ	nullable	column_descr
atpMatch	TBL		1 PK	TID_MID	varchar(40)	NOT NULL	composite primary key/(Tournament Id + matchId)
atpMatch	TBL		2 FK	tourney_id	varchar(40)	NULL	Tournament ID
atpMatch	TBL		3 FK	winner_id	int(10)	NULL	winner_id
atpMatch	TBL		4 FK	loser_id	int(10)	NULL	loser_id
atpMatch	TBL		5	match_num	int(10)	NULL	match_num
atpMatch	TBL		6	winner_seed	text(65535)	NULL	Winner_seed position
atpMatch	TBL		7	winner_entry	text(65535)	NULL	
atpMatch	TBL		8	winner_age	double(22)	NULL	winner_age
atpMatch	TBL		9	winner_rank	text(65535)	NULL	winner_rank
atpMatch	TBL		10	winner_rank_points	text(65535)	NULL	winner_rank_points
atpMatch	TBL		11	loser_seed	text(65535)	NULL	Loser seed position
atpMatch	TBL		12	loser_entry	text(65535)	NULL	
atpMatch	TBL		13	loser_age	double(22)	NULL	loser_age
atpMatch	TBL		14	loser_rank	text(65535)	NULL	loser_rank
atpMatch	TBL		15	loser_rank_points	text(65535)	NULL	loser_rank_points
atpMatch	TBL		16	score	text(65535)	NULL	score
atpMatch	TBL		17	best_of	int(10)	NULL	best_of
atpMatch	TBL		18	round	text(65535)	NULL	round
atpMatch	TBL		19	minutes	int(10)	NULL	minutes
atpMatch	TBL		20	w_ace	int(10)	NULL	winner_Ace Point
atpMatch	TBL		21	w_df	int(10)	NULL	Winner_#DoubleFault
atpMatch	TBL		22	w_svpt	int(10)	NULL	Winner_Total Serve Points
atpMatch	TBL		23	w_1stIn	int(10)	NULL	winner # 1st serves in
atpMatch	TBL		24	w_1stWon	int(10)	NULL	Winner # points won on 1st serve
atpMatch	TBL		25	w_2ndWon	int(10)	NULL	Winner # points won on 2nd serve
atpMatch	TBL		26	w_SvGms	int(10)	NULL	Winner # Serve games
atpMatch	TBL		27	w_bpSaved	int(10)	NULL	Winner break point saved
atpMatch	TBL		28	w_bpFaced	int(10)	NULL	Winner break point faced
atpMatch	TBL		29	l_ace	int(10)	NULL	Loser_Ace Point
atpMatch	TBL		30	l_df	int(10)	NULL	Loser_#DoubleFault
atpMatch	TBL		31	l_svpt	int(10)	NULL	Loser Total Serve Points
atpMatch	TBL		32	l_1stIn	int(10)	NULL	Loser # 1st serves in
atpMatch	TBL		33	l_1stWon	int(10)	NULL	Loser # points won on 1st serve
atpMatch	TBL		34	l_2ndWon	int(10)	NULL	Loser # points won on 2nd serve
atpMatch	TBL		35	l_SvGms	int(10)	NULL	Loser # serve games
atpMatch	TBL		36	l_bpSaved	int(10)	NULL	Loserbreak point saved
atpMatch	TBL		37	l_bpFaced	int(10)	NULL	Loser break point faced



# Ref images for data dictionary

schema_nm	table_nm	obj_typ	ord	is_key	column_nm	data_typ	nullable	column_descr
Group2	bettingDatat	TBL	1	PK	WID_LID_DATE	varchar(40)	NOT NULL	composite primary key/(Winner Id + Loser Id + date)
Group2	bettingDatat	TBL	2	FK	TID	varchar(40)	NULL	Tournament ID
Group2	bettingDatat	TBL	4	FK	WID	int(10)	NULL	Winner ID
Group2	bettingDatat	TBL	5	FK	LID	int(10)	NULL	Loser ID
Group2	bettingDatat	TBL	6		ATP	int(10)	NULL	Tournament number (men)
Group2	bettingDatat	TBL	7		Location	text(65535)	NULL	Venue of tournament
Group2	bettingDatat	TBL	8		NewDate	text(65535)	NULL	Match Date
Group2	bettingDatat	TBL	9		Series	text(65535)	NULL	Name of ATP tennis series (Grand Slam, Masters, International or International Gold)
Group2	bettingDatat	TBL	10		Court	text(65535)	NULL	Type of court (outdoors or indoors)
Group2	bettingDatat	TBL	11		Round	text(65535)	NULL	Round of match
Group2	bettingDatat	TBL	12		Best of	int(10)	NULL	Maximum number of sets playable in match
Group2	bettingDatat	TBL	13		Wrank	int(10)	NULL	ATP Entry ranking of the match winner as of the start of the tournament
Group2	bettingDatat	TBL	14		Lrank	int(10)	NULL	ATP Entry ranking of the match loser as of the start of the tournament
Group2	bettingDatat	TBL	15		W1	int(10)	NULL	Number of games won in 1st set by match winner
Group2	bettingDatat	TBL	16		L1	int(10)	NULL	Number of games won in 1st set by match loser
Group2	bettingDatat	TBL	17		W2	int(10)	NULL	Number of games won in 2nd set by match winner
Group2	bettingDatat	TBL	18		L2	int(10)	NULL	Number of games won in 2nd set by match loser
Group2	bettingDatat	TBL	19		W3	text(65535)	NULL	Number of games won in 3rd set by match winner
Group2	bettingDatat	TBL	20		L3	text(65535)	NULL	Number of games won in 3rd set by match loser
Group2	bettingDatat	TBL	21		W4	text(65535)	NULL	Number of games won in 4th set by match winner
Group2	bettingDatat	TBL	22		L4	text(65535)	NULL	Number of games won in 4th set by match loser
Group2	bettingDatat	TBL	23		W5	text(65535)	NULL	Number of games won in 5th set by match winner
Group2	bettingDatat	TBL	24		L5	text(65535)	NULL	Number of games won in 5th set by match loser
Group2	bettingDatat	TBL	25		Wsets	int(10)	NULL	Number of sets won by match winner
Group2	bettingDatat	TBL	26		Lsets	int(10)	NULL	Number of sets won by match loser
Group2	bettingDatat	TBL	27		Comment	text(65535)	NULL	Comment on the match (Completed, won through retirement of loser, or via Walkover)
Group2	bettingDatat	TBL	28		CBW	text(65535)	NULL	Centrebet odds of match winner
Group2	bettingDatat	TBL	29		CBL	text(65535)	NULL	Centrebet odds of match loser
Group2	bettingDatat	TBL	30		GBW	text(65535)	NULL	Gamebookers odds of match winner
Group2	bettingDatat	TBL	31		GBL	text(65535)	NULL	Gamebookers odds of match loser
Group2	bettingDatat	TBL	32		IWW	text(65535)	NULL	Interwetten odds of match winner
Group2	bettingDatat	TBL	33		IWL	text(65535)	NULL	Interwetten odds of match loser

# Ref images for data dictionary

Player Table :

schema_nm	table_nm	obj_typ	ord	is_key	column_nm	data_typ	nullable	column_descr
Group2	player_info	TBL		1 PK	player_id	int(10)	NOT NULL	Player ID
Group2	player_info	TBL		2	player_name	text(65535)	NULL	Player Name
Group2	player_info	TBL		3	player_hand	text(65535)	NULL	Player Hand
Group2	player_info	TBL		4	player_ht	int(10)	NULL	Player Height
Group2	player_info	TBL		5	player_ioc	text(65535)	NULL	Country of origin

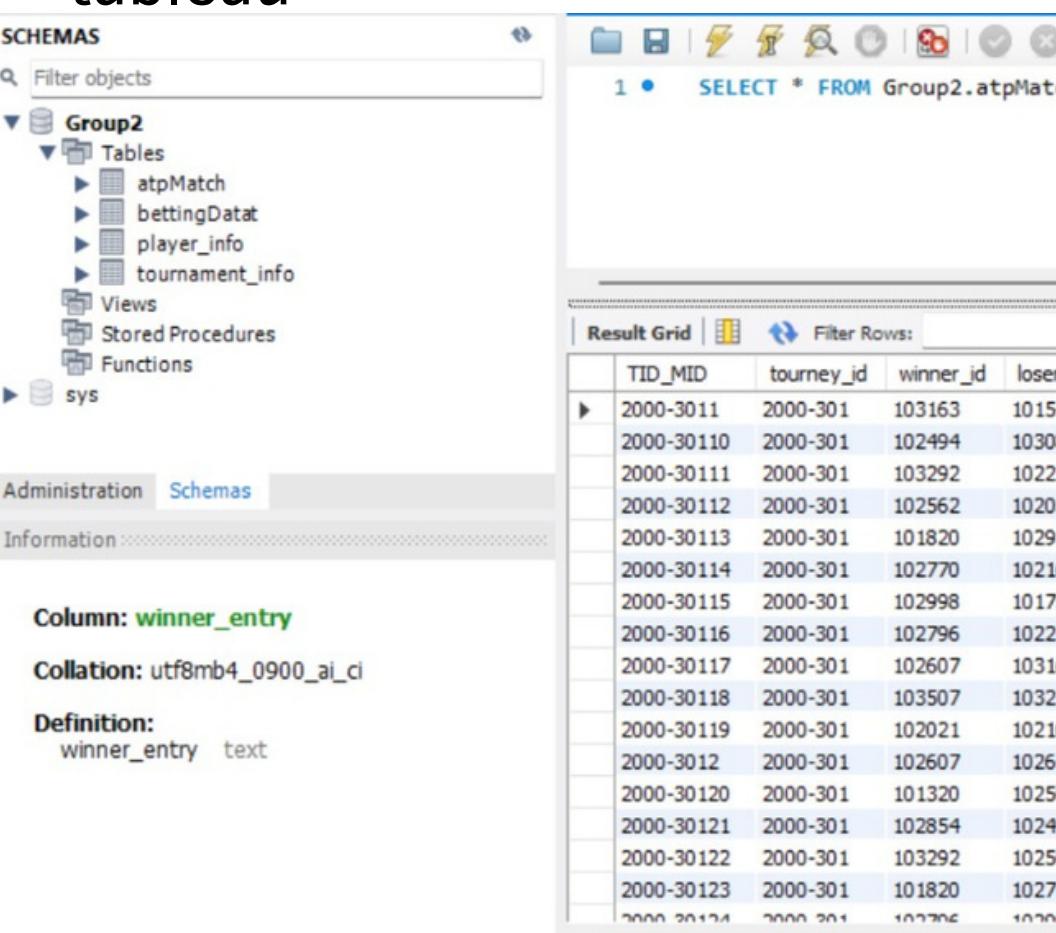
Tournament Table :

schema_nm	table_nm	obj_typ	ord	is_key	column_nm	data_typ	nullable	column_descr
Group2	tournamen	TBL		1 PK	tourney_id	varchar(40)	NOT NULL	Tournament ID
Group2	tournamen	TBL		2	tourney_level	text(65535)	NULL	Level of Tournament
Group2	tournamen	TBL		3	tourney_name	text(65535)	NULL	Tournament Name
Group2	tournamen	TBL		4	surface	text(65535)	NULL	Tournament Surface
Group2	tournamen	TBL		5	draw_size	int(10)	NULL	draw size
Group2	tournamen	TBL		6	MatchDate	text(65535)	NULL	Match date

# DATA QUERYING & ANALYSIS METHODS

## SQL - Data Querying

We used SQL to retrieve data from a normalized table to analyse players data in R studio and tableau

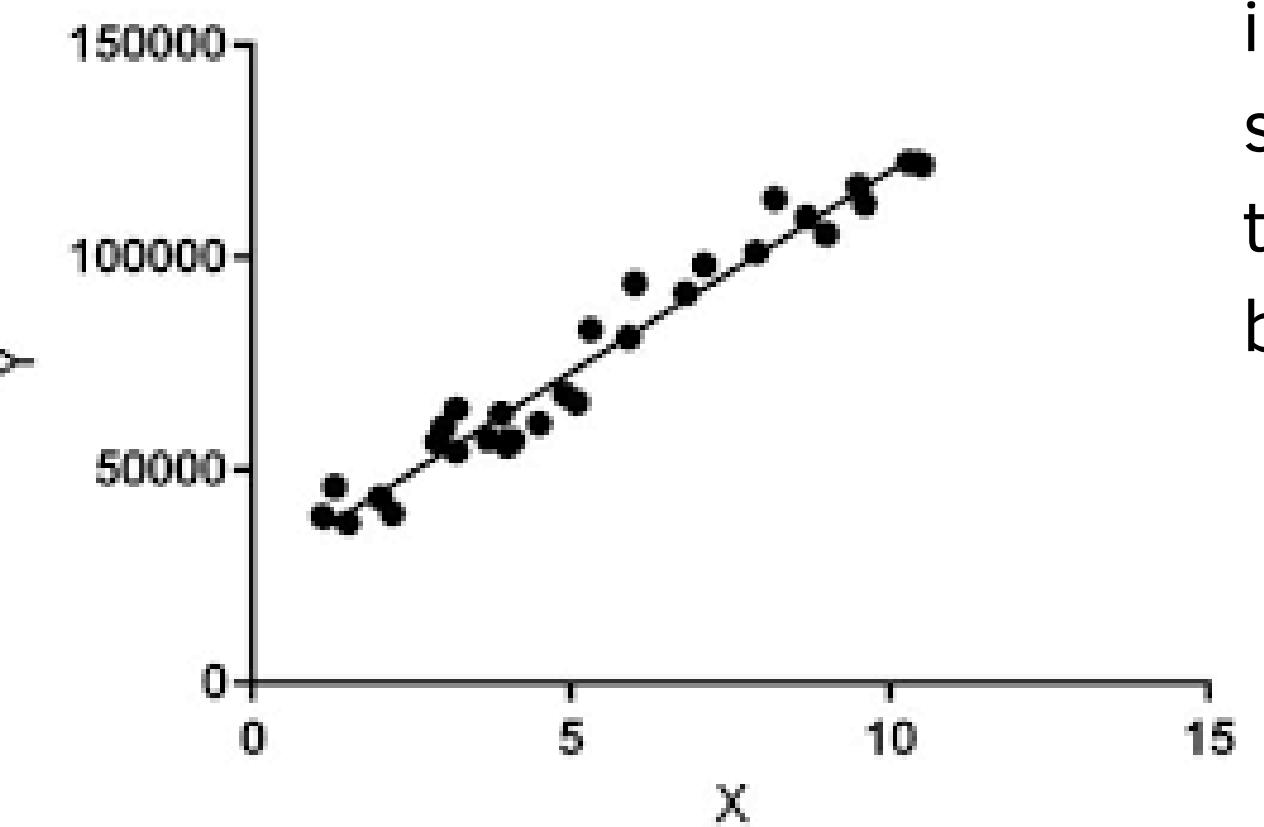


The screenshot shows the MySQL Workbench interface. On the left, the 'SCHEMAS' tree shows 'Group2' selected. Under 'Tables', there is an 'atpMatch' table. The 'Result Grid' shows the following data:

TID_MID	tourney_id	winner_id	loser_id
2000-3011	2000-301	103163	101541
2000-3010	2000-301	102494	103082
2000-30111	2000-301	103292	102221
2000-30112	2000-301	102562	102031
2000-30113	2000-301	101820	102950
2000-30114	2000-301	102770	102161
2000-30115	2000-301	102998	101731
2000-30116	2000-301	102796	102231
2000-30117	2000-301	102607	103161
2000-30118	2000-301	103507	103251
2000-30119	2000-301	102021	102101
2000-3012	2000-301	102607	102641
2000-30120	2000-301	101320	102561
2000-30121	2000-301	102854	102491
2000-30122	2000-301	103292	102561
2000-30123	2000-301	101820	102771
2000-30124	2000-301	102774	103001

## Regression in R

We used linear regression to analyze the relation between win probability vs player height, break point saved % vs match win %



## Player analysis in Tableau

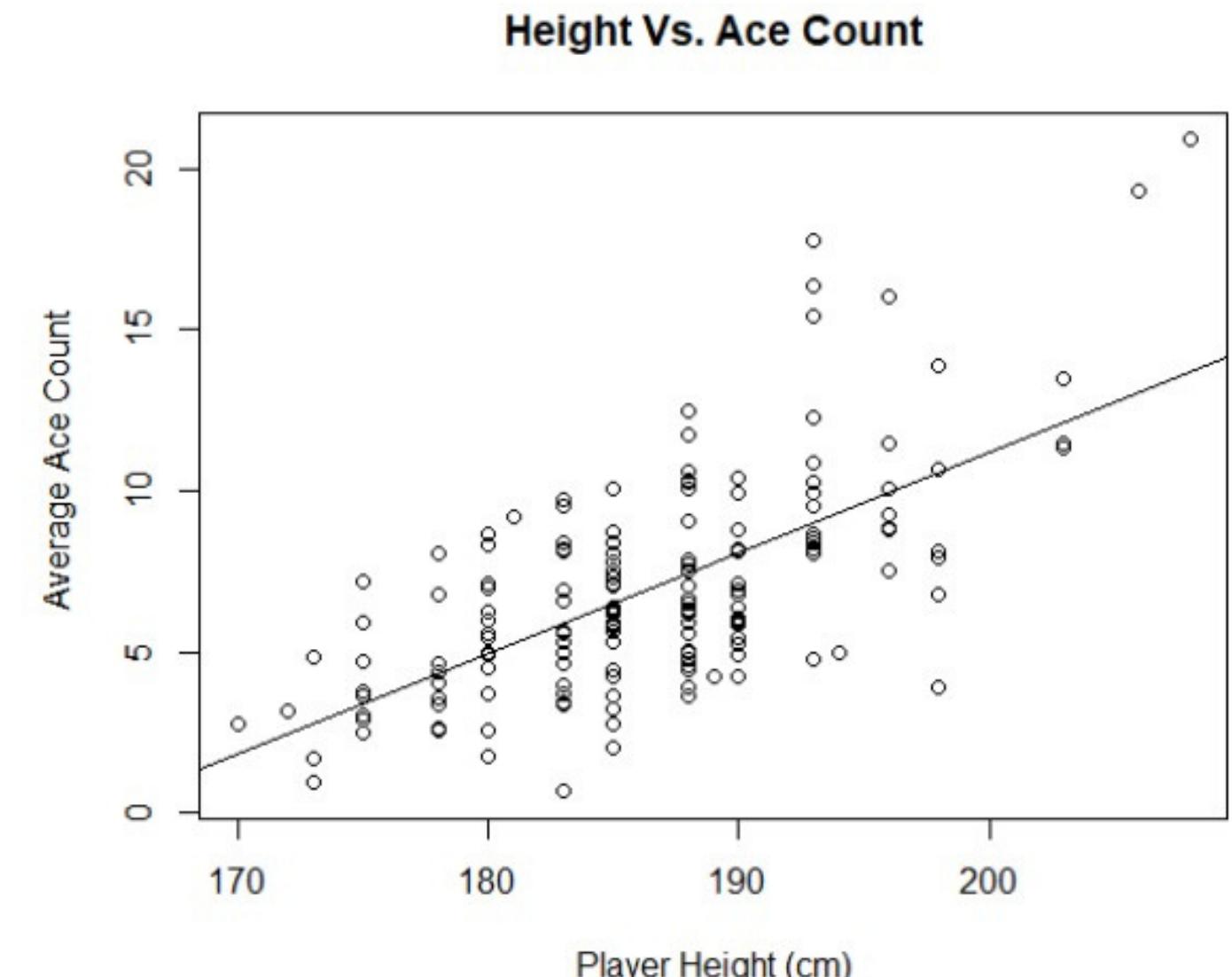
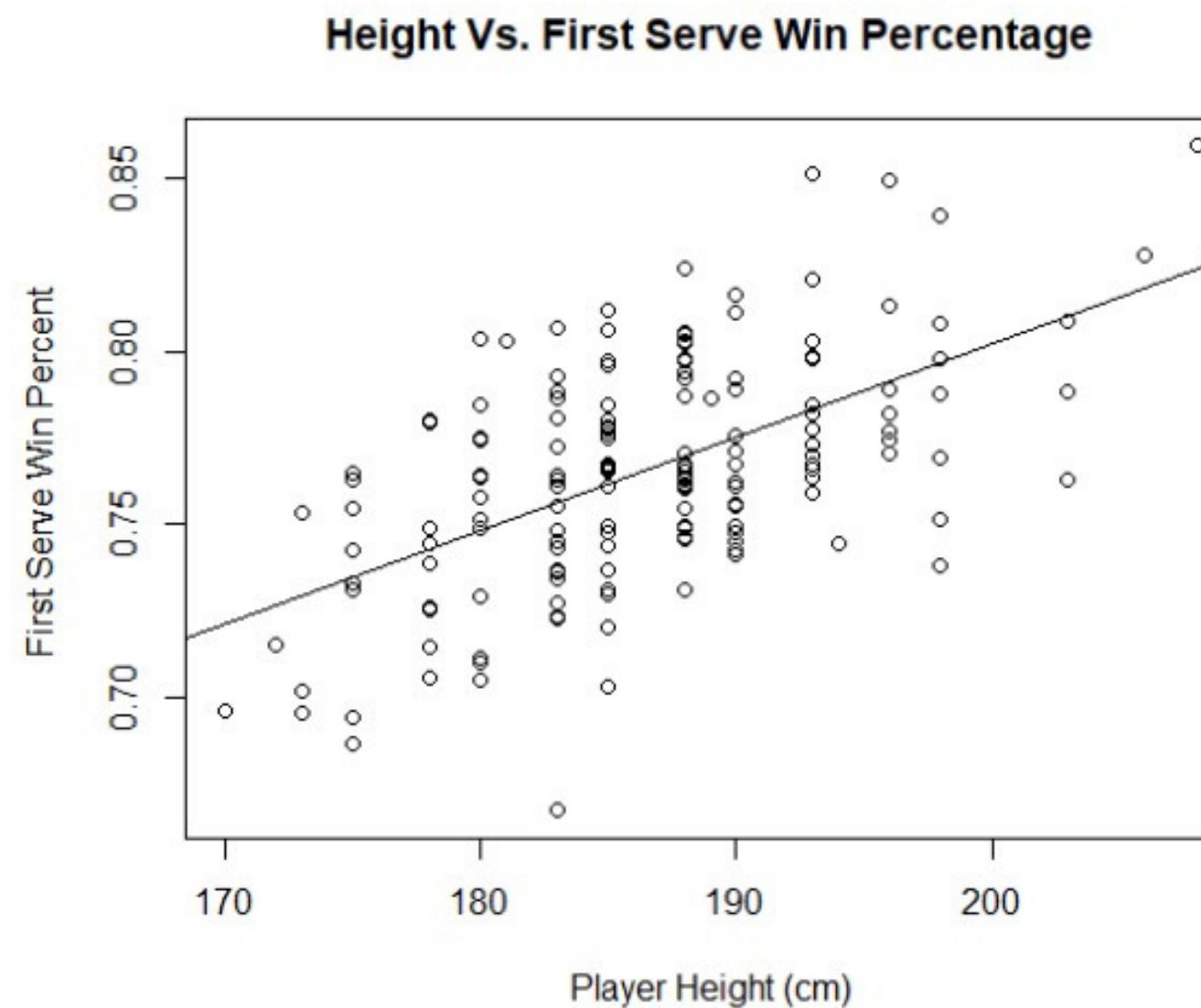
We analyzed data between Rafael Nadal and David Ferrer for the tournament BNP Paribas played in the year 2013 which gave us some interesting insights into the strengths and weaknesses of each of the players. on a count of pitch, breakpoints, height, etc.

# RESEARCH QUESTIONS



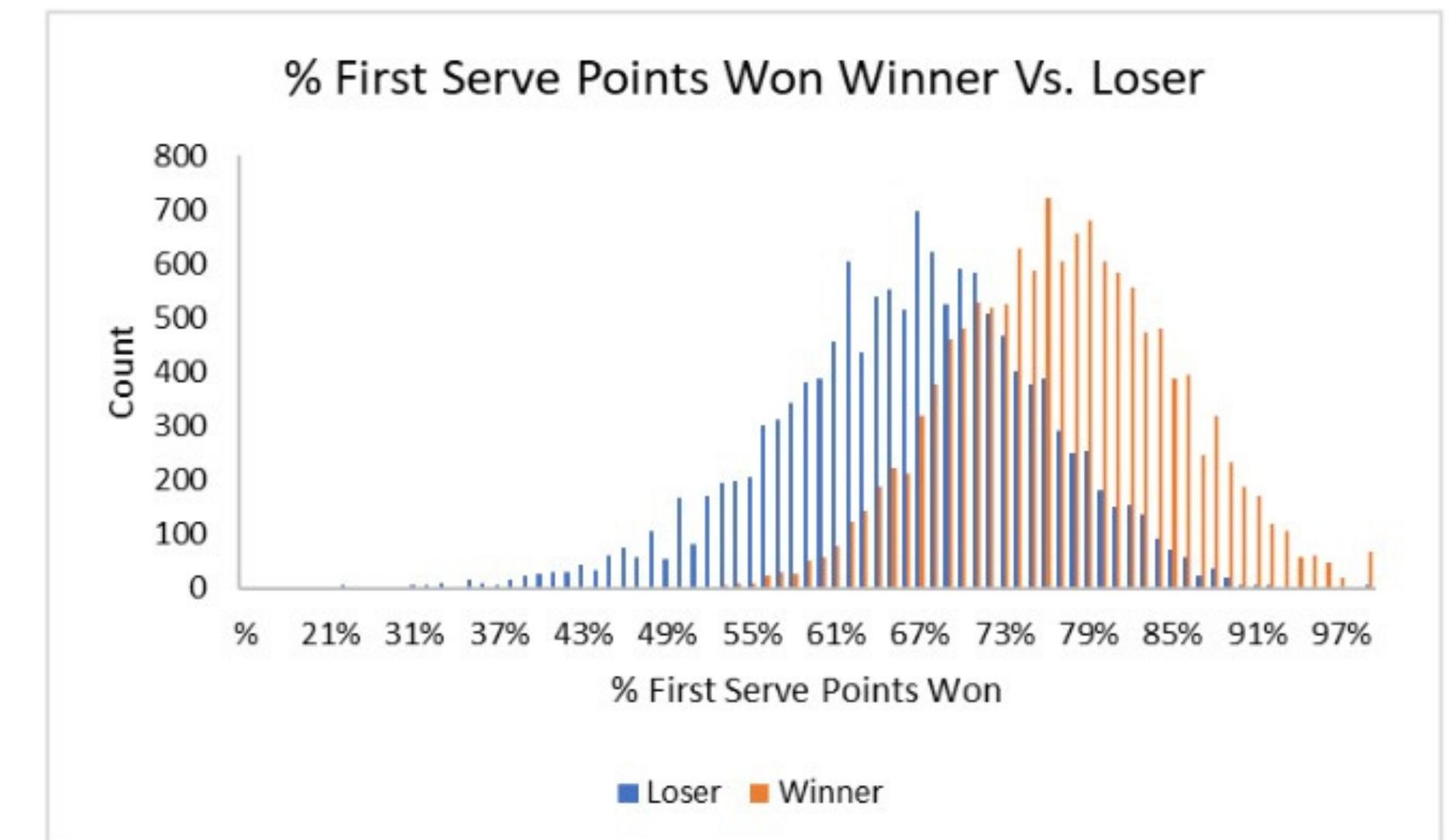
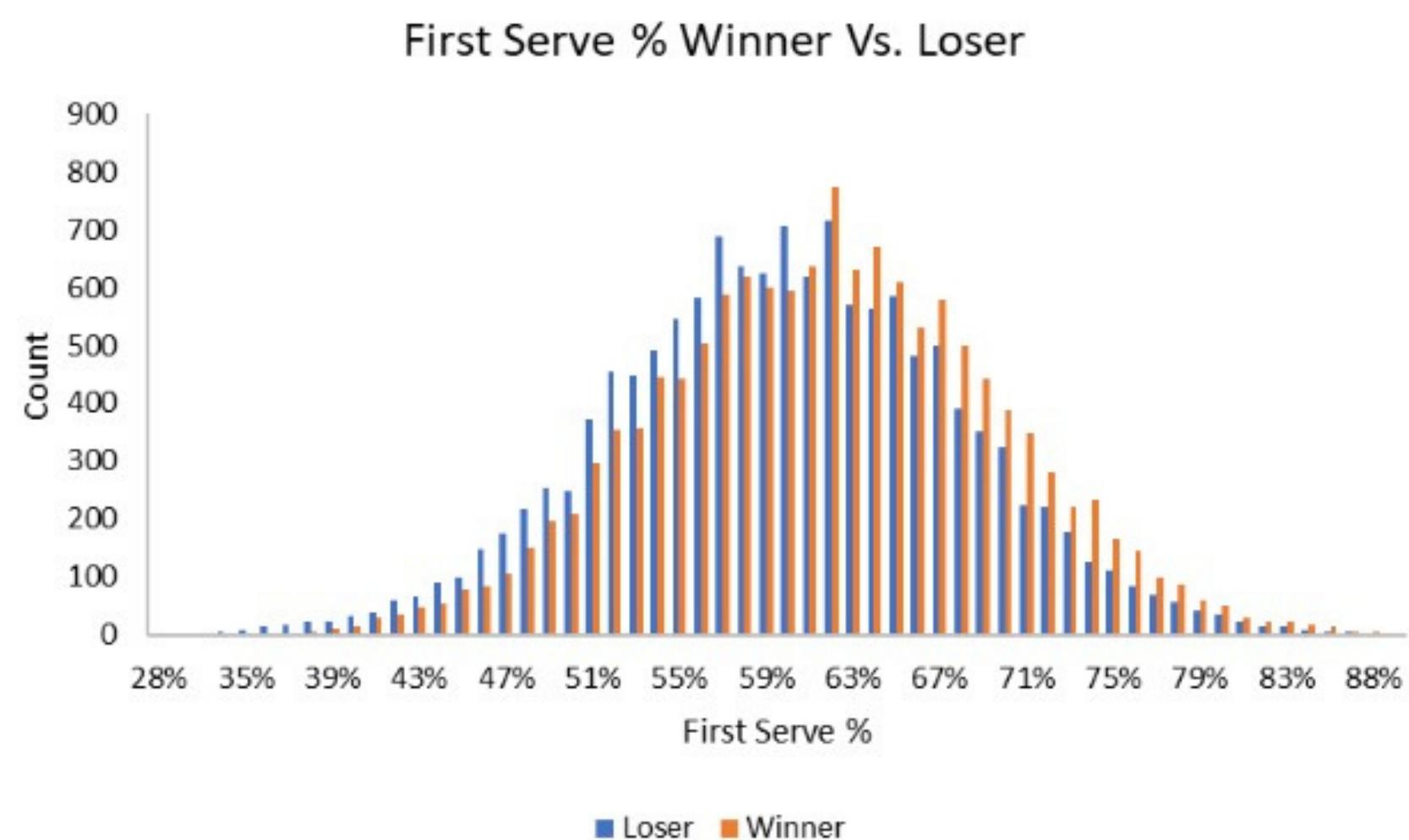
1. Is there a relationship between a player's winning probability and his height?
2. Is there a relationship between a player's winning probability and the breakpoints saved?
3. On a specific match where betting agencies made a high bet on the wrong side, what elements were neglected that we can figure out in our data?

# Height vs. Serve Measures

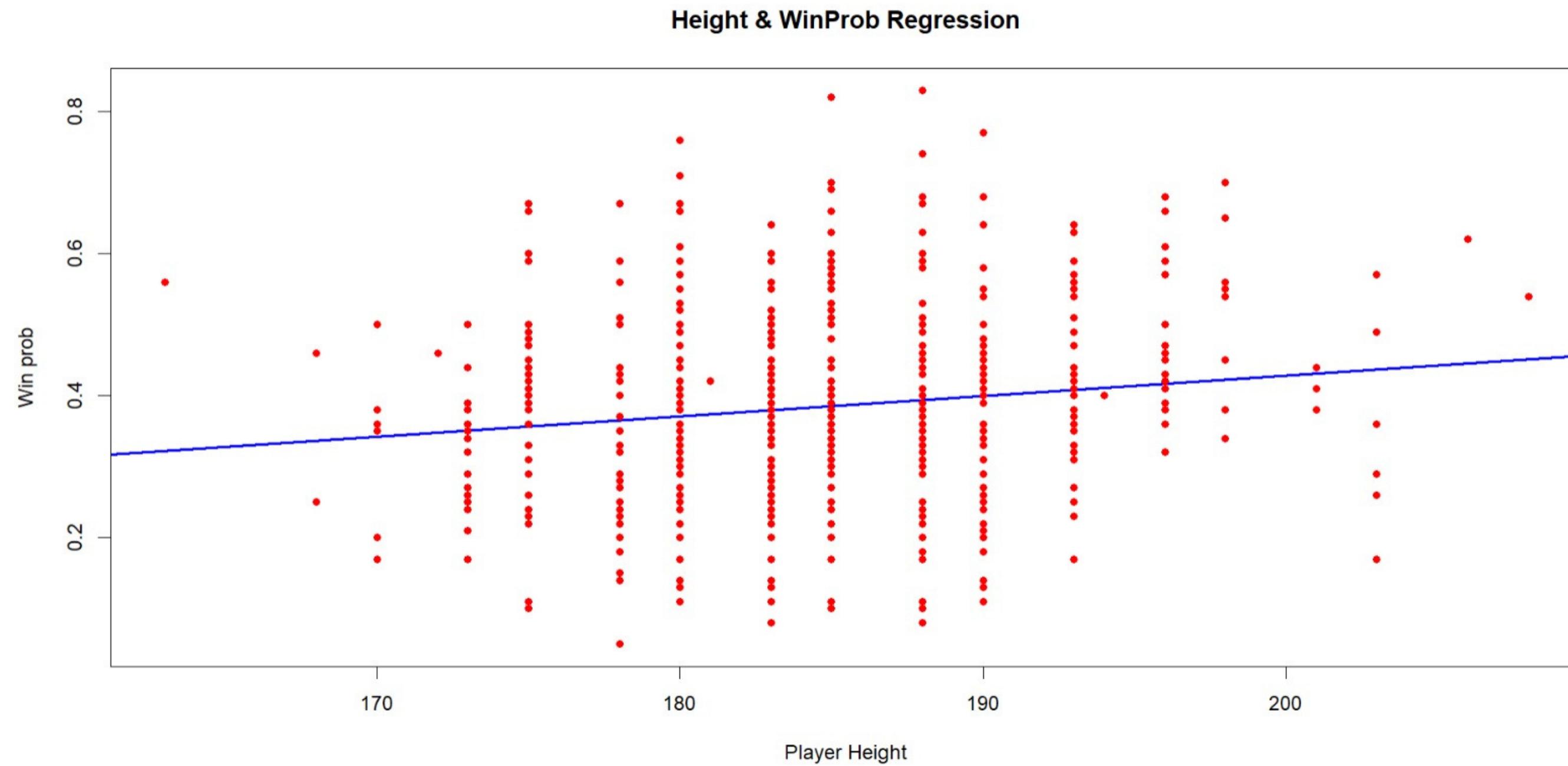


First Serve Win Percentage: Percent of points a player wins when first serve is in play, highlights effectiveness of a serve.

# Winner Vs. Loser Serve Comparison



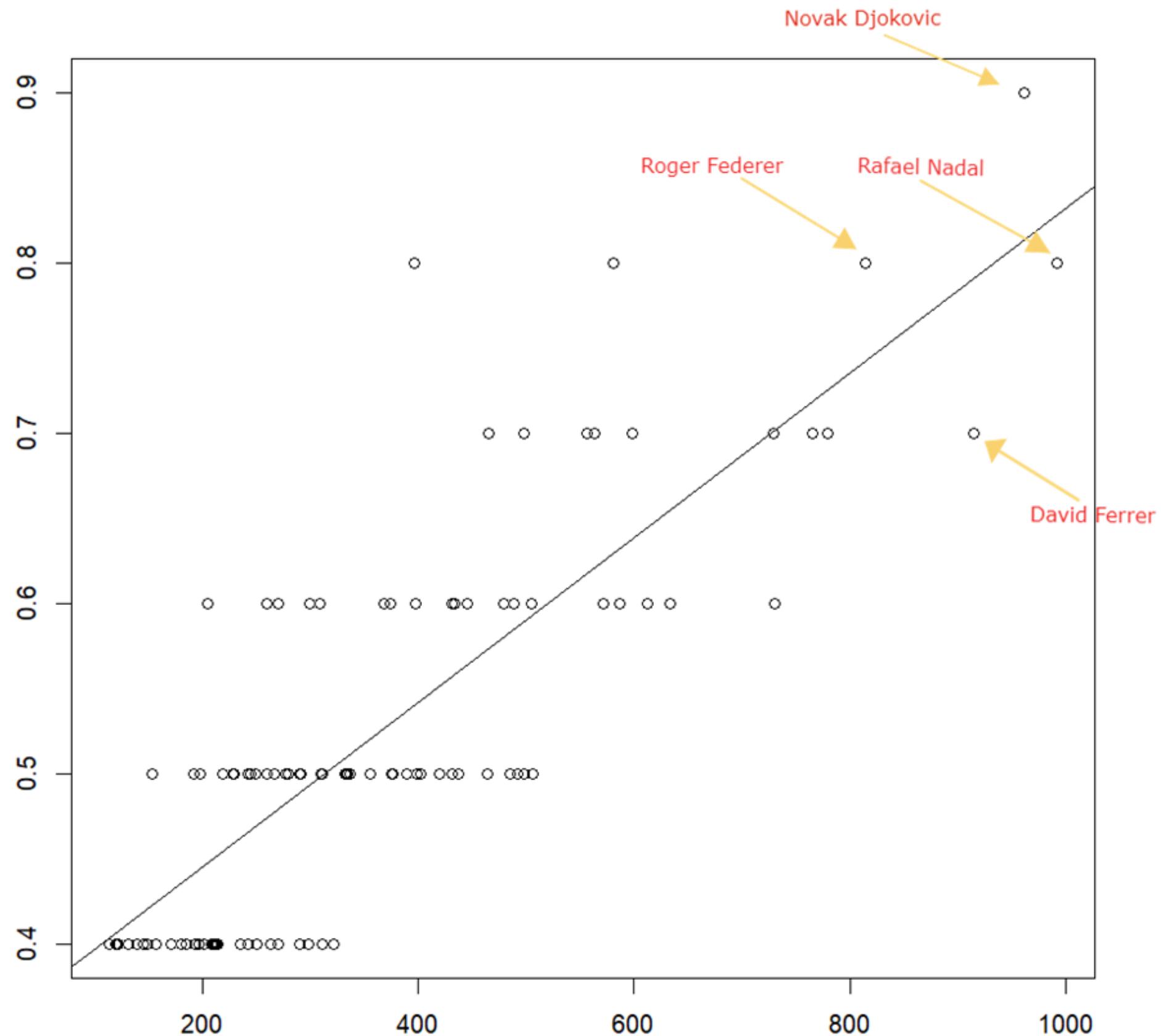
# Regression of Player Height and Win Probability



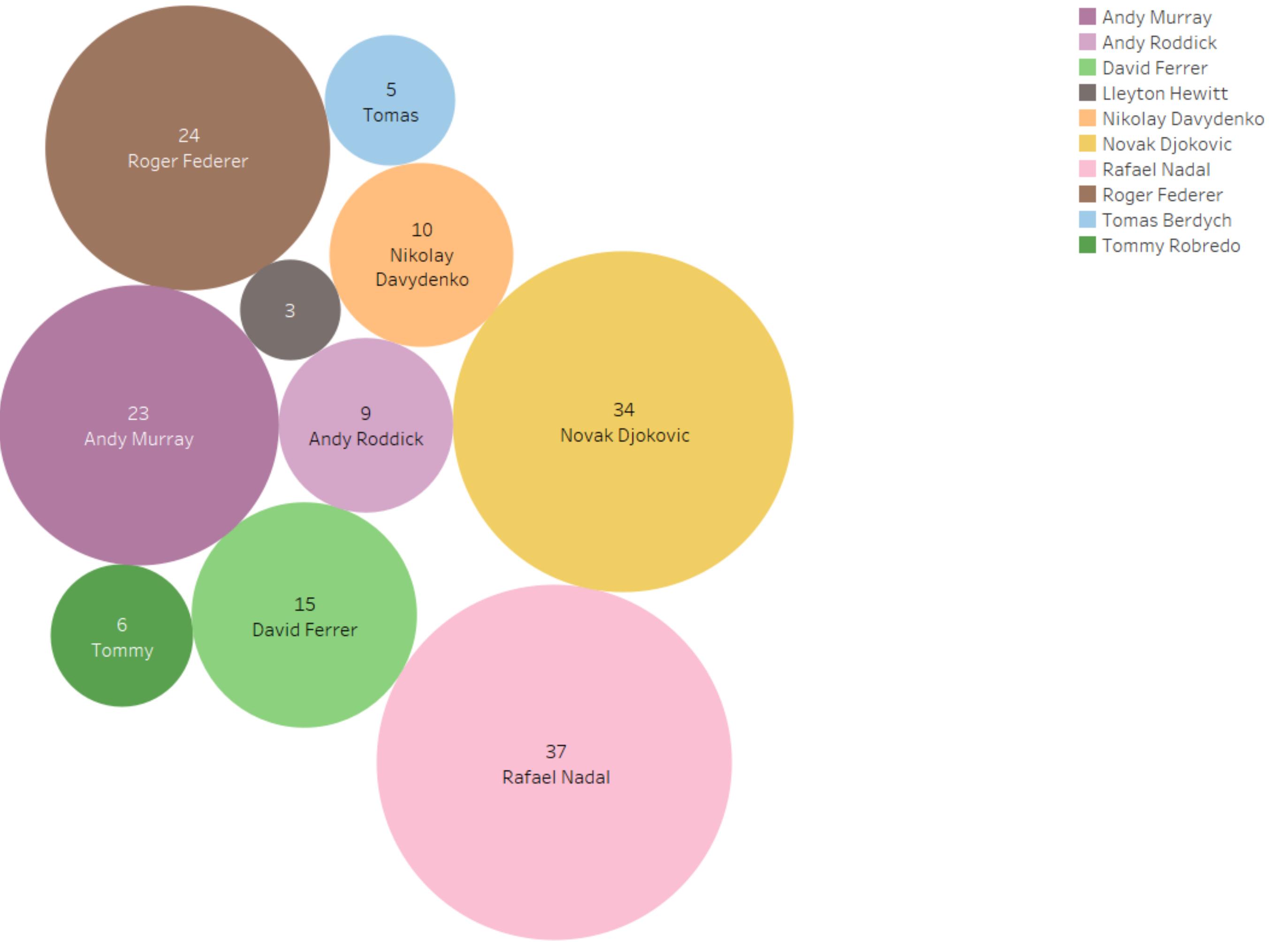
# Breakpoints Saved\* and Match Wins

- R squared: 0.65
- A higher number of breakpoints saved resulted in a player having a higher win percentage.
- The top 20 players save on average 66% of breakpoints saved.
- When betting on a player outside of the top 20 finding an individual with a similar breakpoint saved percentage may increase the chances of a successful bet.

\*Breakpoints Saved: Points where the player serving must win the point or lose the game, being broken means you lose the game while serving.



# Tournament Winners 2008-2013



- Andy Murray
- Andy Roddick
- David Ferrer
- Lleyton Hewitt
- Nikolay Davydenko
- Novak Djokovic
- Rafael Nadal
- Roger Federer
- Tomas Berdych
- Tommy Robredo

# David Ferrer Vs. Rafael Nadal

**Background:** Rafael Nadal was upset at the BNP Paribas 2013 Paris Masters 1000 event by David Ferrer despite being a heavy favorite. Ferrer had only beaten Nadal 4 times out of the 24 times they played until then. The tournament is held in late October after the Grand Slams have all been played.

**Goal:** Using the database we hope to uncover potential indicators that Nadal was over favored and evaluate whether our data could have been used to set more reasonable betting odds.

	winner_name	loser_name	NewDate	Round	Tournament	PSL	PSW	B365L	B365W
▶	David Ferrer	Rafael Nadal	2013-10-28	Semifinals	BNP Paribas Masters	1.15	6.65	1.11	6.5

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

David Ferrer			
Surface	Wins	ServePoints Won	Avg. W Ace
Clay	22	1,550	2.77
Grass	3	170	2.33
Hard	20	1,616	5.15
Grand Total	45	3,336	3.80

Rafael Nadal			
Surface	Wins	ServePoints Won	Avg. W Ace
Clay	25	1,799	2.48
Grass	2	208	5.50
Hard	14	1,164	3.79
Grand Total	41	3,171	3.07

Wins - Hard Surface			
Surface	Wins	ServePoints Won	Avg. W Ace
Hard	20	1,616	5.15
Grand Total	20	1,616	5.15

Wins - Hard Surface			
Surface	Wins	ServePoints Won	Avg. W Ace
Hard	14	1,164	3.79
Grand Total	14	1,164	3.79

Losses - Hard Surface			
Surface	Losses	ServePoints Won	Avg. L Ace
Hard	12.0	936.0	2.0

Losses - Hard Surface			
Surface	Losses	ServePoints Won	Avg. L Ace
Hard	11.0	856.0	2.7

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	856.0	2.7	

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

**Win % 67%**

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	856.0	2.7	

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

**Win % 67%**

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	856.0	2.7	

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

**Win % 67%**

<b>David Ferrer</b>			
Surface	Wins	ServePoints Won	Avg. W Ace
Clay	22	1,550	2.77
Grass	3	170	2.33
Hard	20	1,616	5.15
Grand Total	45	3,336	3.80

**Win % 72%**

<b>Rafael Nadal</b>			
Surface	Wins	ServePoints Won	Avg. W Ace
Clay	25	799	2.48
Grass	2	208	5.50
Hard	14	1,164	3.79
Grand Total	41	3,171	3.07

## **Wins - Hard Surface**

Surface	Wins	ServePoints Won	Avg. W Ace
Hard	20	1,616	5.15
Grand Total	20	1,616	5.15

## **Wins - Hard Surface**

Surface	Wins	ServePoints Won	Avg. W Ace
Hard	14	1,164	3.79
Grand Total	14	1,164	3.79

## **Losses - Hard Surface**

Surface	Losses	ServePoints Won	Avg. L Ace
Hard	12.0	936.0	2.0

## **Losses - Hard Surface**

Surface	Losses	ServePoints Won	Avg. L Ace
Hard	11.0	856.0	2.7

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	856.0	2.7	

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

**Win % 63%**

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	856.0	2.7	

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

**Win % 63%**

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	866.0	2.7	

# Uncovering Indicators of Upset

Service and Win data from the prior two seasons including Hard court only for Nadal and Ferrer

<b>David Ferrer</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	22	1,550	2.77	
Grass	3	170	2.33	
Hard	20	1,616	5.15	
Grand Total	45	3,336	3.80	

<b>Rafael Nadal</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Clay	25	1,799	2.48	
Grass	2	208	5.50	
Hard	14	1,164	3.79	
Grand Total	41	3,171	3.07	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	20	1,616	5.15	
Grand Total	20	1,616	5.15	

<b>Wins - Hard Surface</b>				
Surface	Wins	ServePoints Won	Avg. W Ace	
Hard	14	1,164	3.79	
Grand Total	14	1,164	3.79	

<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	12.0	936.0	2.0	

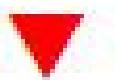
<b>Losses - Hard Surface</b>				
Surface	Losses	ServePoints Won	Avg. L Ace	
Hard	11.0	860.0	2.7	

**Win % 63%**

**Win % 56%**

## David Ferrer

All Surface	Svpt
Total matches	67
Total wins	45
Total loses	22
Win %	67%
Loss %	33%



## Rafael Nadal

All Surface	RN - Svpt
Total matches	57
Total wins	41
Total loses	16
Win %	72%
Loss %	28%



Hard Surface	Svpt
Total matches	32
Total wins	20
Total loses	12
Win %	63%
Loss %	38%



Hard Surface	RN - Svpt
Total matches	25
Total wins	14
Total loses	11
Win %	56%
Loss %	44%



# How could Odds have been improved?

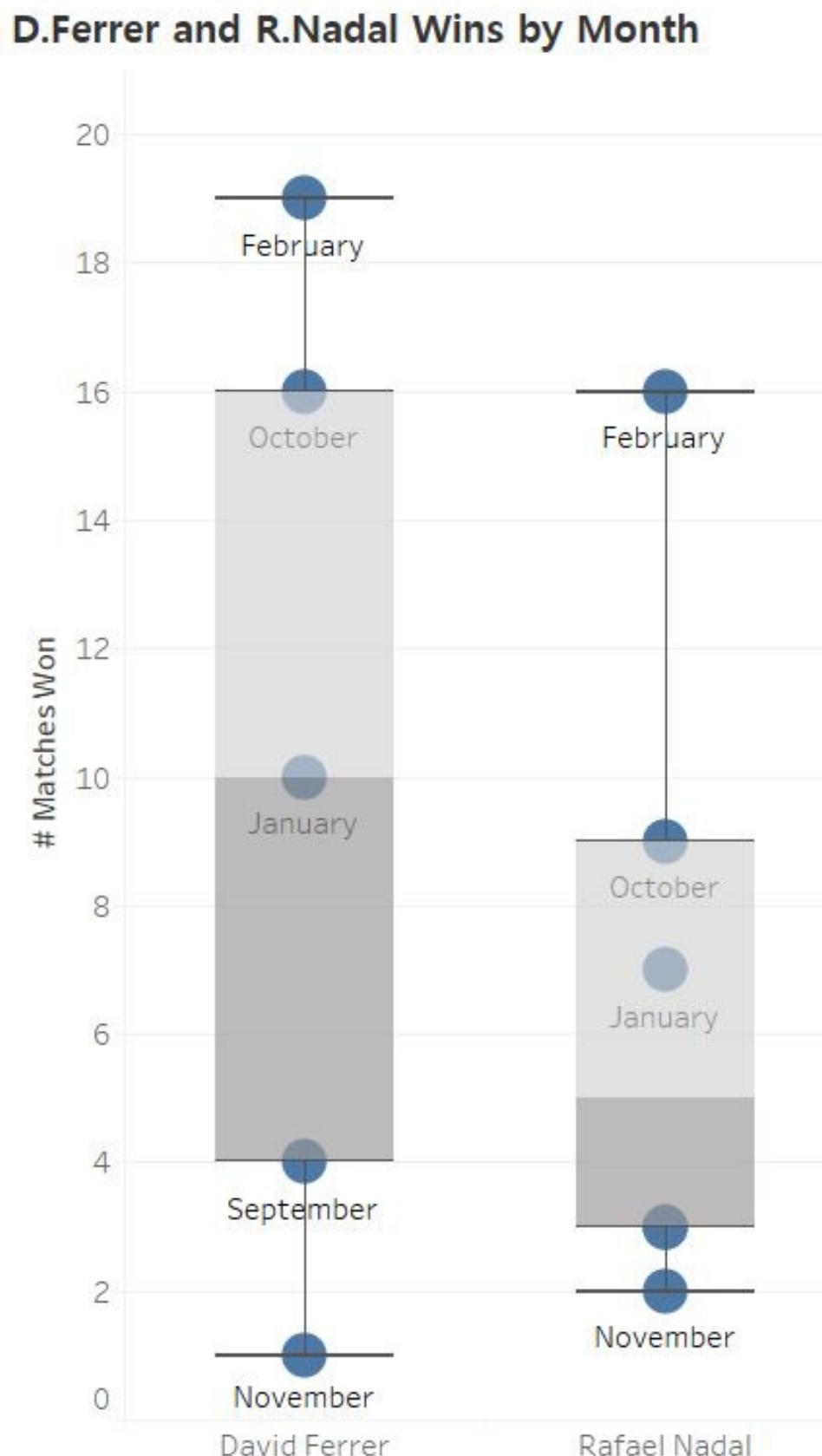
- Rafael Nadal has tailed off at the end of the ATP year with most of his wins coming before the U.S. Open.

	2004	2005	2006	2007	2008	2009	2010	2011	2013	2014	2015	2016
Cincinnati Masters	R64	R64	QF	R32	SF	SF	QF	QF			R16	R16
Paris Masters				F	QF	SF			SF		QF	
Shanghai Masters						F	R16	R16	SF	R32	SF	R32

- Nadal has won only 2 of 33 Masters 1000 tournaments after US Open those usually held between October and November.

# How could Odds have been improved?

- Nadal's consistency on main knockout games on a HARD court is lesser in Masters (losing 58% of games after regular rounds).
  - Ferrer was the defending champion of the tournament, so precedent for him to play well.
  - Ferrer had a higher win percentage on hardcourts over the previous two years during the same period.
  - Ultimately, Ferrers better performance during the previous two years on hard courts, in addition to the time of year the tournament was held, he likely should not have been as large of an underdog



# Conclusion

- Height is not a very significant factor in deciding the win % of a player.
- We must consider serve points saved, ace counts, strengths and weaknesses of each player on the ground of surface in several tournament to predict the win % and to decide the betting odds.



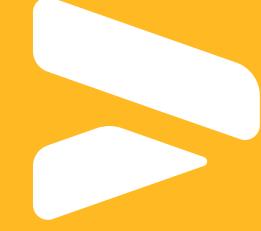
# Discussion

Potential questions our database could address in the future:

1. How does a player's ability to consistently serve Aces affect their chances of winning a match?
2. Does a change in court surface(Clay, Grass, Hard Court) affect the importance of various metrics(number of aces, first serve percentage, etc.) in match outcomes?
3. Does tournament level(Grand Slam, Masters 1000, Masters 500) affect the importance of the various measures?
4. Understand how different court surfaces effect betting odds for a specific player, and evaluate whether they accurately represent a players chances of winning given the historical data(Break Points Saved, First Serve Win Percentage etc.)on each surface.

# ANNEXURE-Tables Retrieved through SQL





# Thank you!

Feel free to approach us  
if you have any questions.