MODULE 4: PICKING THE BEST PLAYERS FOR FANTASY FOOTBALL

The basic problem in Fantasy Football is the following:

- You must pick 16 players from given lists of players of various types, Quarterbacks (QB), Running backs (RB), Wide receivers (WR), Tight ends (TE), Defense/Special teams (D/ST) and Place kickers (K).
- You must choose a minimum number of players of each type say 1QB, 2RB, 2WR, 1TE, 1 D/ST, 1K.
- Each player makes points throughout the season for successful plays, such as touch downs, rushing yards and receiving yards. Players also lose points for interceptions thrown or fumbles. Thus each player has a **value** or a potential value.
- Players are recruited with a draft system where teams make sequential choices from the roster. Trading players during the season is sometimes allowed.
- Each player has a **cost**.
- As a team manager, you have a **maximum** amount to spend, so the total cost of your players must be less than that specific amount.

The basic problem is to maximize the number of points your team will earn subject to budget constraints and constraints on the number of players. One must also keep in mind that the most desirable players may not be available when your turn to pick a player arrives. Determining the value of a player is an important part of the process and identifying undervalued players helps increase gain at a lower cost.

We will start by thinking about how we might come up with a value for a player. We consider how we might amalgamate the rankings of the experts or alternatively how we might attempt to forecast the number of points each player will make. We will think a little about the solution to the optimization problem which is analogous to `the knapsack problem` in optimization theory. Although we will not consider it in this section due to lack of time, one can apply game theory to the drafting process see [2].

1. Ranking

One way to get an idea of the comparative value of players and perhaps identify undervalued players is to rank the players in order. We might use points scored last season or some other statistic to rank our players or we might avail of the rankings by fantasy football experts which are widely available. Most likely different statistics and different experts will give us different rankings for our players and you may wish to combine some of the best rankings to get your own ranking. This is the first topic that we discuss and we look at methods from voting theory to help us combine rankings. We also look at Arrow's theorem which says that there is no perfect method of combining rankings.

1.1. **Ratings and Rankings.** **Note** A **ranking** refers to a rank-ordered list of competitors and a **rating** gives us a list of numerical scores. Every rating gives us a ranking for the competitors.

Example 1.1. *Suppose we rank Quarterbacks according to the number of completed passes in the previous season. The number of completed passes (COMP) gives us a rating as shown below for the 2014 season. We can convert this to a ranking, the top ranked (# 1 player) being the one with the most completed passes. The top ten quarterbacks for the 2014 season according to this ranking are shown on the left below.*

Completions Leaders - All Players

RK	PLAYER	TEAM	COMP
1	Drew Brees, QB	NO	456
2	Matt Ryan, QB	ATL	415
3	Ben Roethlisberger, QB	PIT	408
4	Peyton Manning, QB	DEN	395
5	Ryan Tannehill, QB	MIA	392
6	Andrew Luck, QB	IND	380
7	Eli Manning, QB	NYG	379
	Philip Rivers, QB	SD	379
9	Tom Brady, QB	NE	373
10	Jay Cutler, QB	CHI	370

Passing Yards Leaders - All Players

RK	PLAYER	TEAM	COMP	ATT	PCT	YDS
1	Drew Brees, QB	NO	456	659	69.2	4,952
	Ben Roethlisberger, QB	PIT	408	608	67.1	4,952
3	Andrew Luck, QB	IND	380	616	61.7	4,761
4	Peyton Manning, QB	DEN	395	597	66.2	4,727
5	Matt Ryan, QB	ATL	415	628	66.1	4,694
6	Eli Manning, QB	NYG	379	601	63.1	4,410
7	Aaron Rodgers, QB	GB	341	520	65.6	4,381
8	Philip Rivers, QB	SD	379	570	66.5	4,286
9	Matthew Stafford, QB	DET	363	602	60.3	4,257
10	Tom Brady, QB	NE	373	582	64.1	4,109

Example 1.2. *On the other hand, we could rank Quartebacks according to the number of passing yards they have acquired in in the previous season. The number of passing yards (YDS) gives us a rating as shown below for the 2014 season. We can convert this to a ranking, the top ranked (# 1 player) being the one with the most passing yards. The top ten quarterbacks for the 2014 season according to this ranking are shown the right above. Note that it gives a different ranking than the one we got using completions.*

There are also a number of combined statistical ratings to choose from, for example the **Expected Points Added (EPA)** from Football Analytics.com (shown on the left below or ESPN's Total Quarterback rating (Total QBR) as shown on the right below.

Rank	Player	Team	G	WPA	EPA
1	12-A.Rodgers	GB	18	5.87	223.8
2	12-T.Brady	NE	19	4.16	176.8
3	9-T.Romo	DAL	17	4.94	166.5
4	7-B.Roethlisberger	PIT	17	4.03	153.1
5	5-J.Flacco	BLT	18	3.07	148.8
6	9-D.Brees	NO	16	2.81	137.2
7	18-P.Manning	DEN	17	3.69	126.7
8	12-A.Luck	IND	19	1.86	126.4
9	3-R.Wilson	SEA	19	3.38	121.6
10	17-P.Rivers	SD	16	3.25	109.0
11	2-M.Ryan	ATL	16	2.92	107.6
12	9-M.Stafford	DET	17	3.72	76.0
13	17-R.Tannehill	MIA	16	2.53	70.7
14	11-A.Smith	KC	15	1.33	63.4
15	10-E.Manning	NYG	16	3.17	62.6
16	1-C.Newton	CAR	16	2.12	53.1
17	7-C.Kaepernick	SF	16	1.36	51.0
18	3-C.Palmer	ARZ	6	1.96	48.2
19	14-R.Fitzpatrick	HST	12	0.22	41.0
20	5-T.Bridgewater	MIN	13	1.33	38.4

Season Leaders · Weekly Leaders · Best Games · Best Teams · All-Time Best Seasons* · All-Time Best Games*

2014 Regular Season NFL Leaders

RK	PLAYER	PASS EPA	RUN EPA	SACK EPA	PEN EPA	TOTAL EPA	ACT PLAYS	QB PAR	QB PAA	TOTAL QBR
1	Tony Romo, DAL	86.2	3.3	-18.8	5.1	75.7	537	103.9	61.1	**82.7**
2	Aaron Rodgers, GB	91.9	11.7	-17.0	6.6	93.2	648	125.0	73.4	**82.6**
3	Peyton Manning, DEN	91.3	-3.3	-10.8	5.4	82.5	704	118.6	62.4	**77.2**
4	Tom Brady, NE	81.6	3.6	-12.5	6.2	78.9	699	109.6	53.9	**74.3**
5	Ben Roethlisberger, PIT	91.8	-2.1	-19.2	5.1	75.5	713	107.0	50.1	**72.5**
6	Drew Brees, NO	89.7	2.4	-17.6	7.8	82.4	765	112.2	51.2	**71.6**
7	Eli Manning, NYG	84.4	1.8	-16.4	4.6	74.4	701	101.3	45.4	**70.9**
8	Joe Flacco, BAL	65.5	-0.9	-13.9	12.9	63.7	659	87.2	34.6	**67.3**
9	Matt Ryan, ATL	81.6	2.9	-18.6	1.5	67.3	741	97.1	38.0	**67.0**
10	Philip Rivers, SD	81.0	-0.1	-20.0	7.2	68.1	707	92.3	35.9	**66.8**

In addition to rankings created from ratings derived from statistics, many sports analysts publish rankings created from algorithms (which produce ratings and in turn rankings) and opinions. We show some examples for such Quarterback rankings below.

Jamey Eisenberg
Senior Fantasy Writer
Follow

Updated: Mon, Mar 16, 2015 12:01 am

Quarterbacks	
1	Aaron Rodgers GB
2	Andrew Luck IND
3	Peyton Manning DEN
4	Tom Brady NE
5	Matt Ryan ATL
6	Cam Newton CAR
7	Ben Roethlisberger PIT
8	Russell Wilson SEA
9	Drew Brees NO
10	Matthew Stafford DET
11	Tony Romo DAL
12	Eli Manning NYG
13	Ryan Tannehill MIA

Dave Richard
Senior Fantasy Writer
Follow

Updated: Fri, Mar 20, 2015 2:51 pm

Quarterbacks	
1	Aaron Rodgers GB
2	Andrew Luck IND
3	Tom Brady NE
4	Peyton Manning DEN
5	Matt Ryan ATL
6	Cam Newton CAR
7	Russell Wilson SEA
8	Tony Romo DAL
9	Drew Brees NO
10	Eli Manning NYG
11	Matthew Stafford DET
12	Ryan Tannehill MIA
13	Philip Rivers SD

Christopher Harris' 2015 QB rankings

Rank	Player
1	Aaron Rodgers
2	Andrew Luck
3	Peyton Manning
4	Tom Brady
5	Russell Wilson
6	Ben Roethlisberger
7	Matt Ryan
8	Drew Brees
9	Cam Newton
10	Ryan Tannehill
11	Tony Romo
12	Matthew Stafford
13	Philip Rivers
14	Sam Bradford
15	Eli Manning
16	Colin Kaepernick
17	Carson Palmer
18	Jay Cutler
19	Joe Flacco
20	Andy Dalton
21	Teddy Bridgewater
22	Robert Griffin III
23	Alex Smith
24	Nick Foles
25	Derek Carr
26	Blake Bortles
27	Ryan Fitzpatrick
28	Ryan Mallett
29	Zach Mettenberger
30	Josh McCown

2. Combining Rankings

None of the above ranking systems will be 100% accurate in predicting points earned per player in the upcoming season. Some sites keep track of accuracy levels of the various rankings such as `Fantasy Pros`. We might therefore consider amalgamating some of the top rankings to make an overall ranking. This can be done in a number of ways which are studied in voting theory. There is however no perfect method of amalgamating rankings. We will go through some of the most common methods below. We will use some terminology from voting theory so that we will not confuse the overall ranking we wish to achieve with the rankings we are combining. We will refer to

- the ranking systems or statistics we use for a ranking as voters,
- each individual ranking will be referred to as a ballot
- and the rank given to each player as a vote.

2.1. **Plurality.** This is a method for choosing a single "winning" candidate from a a number of rankings or voters who have ranked the candidates. The winning candidate is the one with the most number 1 votes. In the case of a tie, some back up plan is needed to decide between the winners.

Example 2.1. *Suppose we want to choose a wide receiver from the following 5 for our fantasy football team. We have ranked the players using EPA/P (Expected points added per play), CR(Catch Rate), YPR(Yards Per reception), RecTD(Received Touch Downs) and Rec(Number of receptions) for the 2014 season from the site `Advanced Football Analytics`.*

Rankings − >	EPA	CR	YPR	RecTD	Rec
E. Sanders	1	1	4	3	1
T. Williams	4	5	1	1	5
S. Smith	2	2	3	4	2
M. Floyd	3	4	2	5	4
M. Wallace	5	3	5	2	3

According to the plurality method Sanders would be our top choice, since he has 3 number one votes.

2.2. **Splitting the vote.** In addition to the problem with ties, the plurality method only takes into account the number 1 votes. This can lead to a situation where the number 1 votes are split among athletes of similar similar high ability and it is conceivable that an athlete who is least preferred by a majority of the voters could be the winner of the plurality vote as in the following example:

Example 2.2. *Lets assume we are considering picking one of the five wide receivers listed below as W.R. 1- W.R. 5. Lets also assume that we are using the plurality method with eight good ranking systems RA- RG to make our choice. We have the following results:*

Rankings − >	RA	RB	RC	RD	RE	RF	RG	RF
W.R. 1	1	1	2	3	2	2	2	2
W.R. 2	3	4	3	2	4	4	5	5
W.R. 3	2	2	1	1	3	3	3	3
W.R. 4	4	3	4	4	1	5	4	4
W.R. 5	5	5	5	5	5	1	1	1

If one uses the Plurality method to put together these rankings, the winner is W.R. 5 despite the fact that W.R. 5 is the lowest ranked athlete in a majority of the rankings. One could get around this problem by having a **runoff** between the top two athletes. (If there is a tie for second place, we eliminate all but those in first and second place or use tie breaking rules if we must).
We redistribute the number 1 votes for the eliminated athletes by assigning them to the athlete ranked highest among the remaining candidates.
We continue the process until we have a winner.

In the above example, this would lead to the following result

Rankings − >	RA	RB	RC	RD	RE	RF	RG	RH	#1 Votes in Round 1	#1 Votes in Round 2	#1 Votes in Round 3
W.R. 1	1	1	2	3	2	2	2	2	2	3	5(winner)
W.R. 2	3	4	3	2	4	4	5	5	0 (eliminate)		
W.R. 3	2	2	1	1	3	3	3	3	2	2(eliminate)	
W.R. 4	4	3	4	4	1	5	4	4	1 (eliminate)		
W.R. 5	5	5	5	5	5	1	1	1	3	3	3

2.3. **Instant Runoff.** This is the same as plurality with runoff described above, where we eliminate the athlete with the least number of number one votes in each round.

To apply the Instant Runoff method:

(1) Each voter ranks the list of athletes in order of preference; the athletes are ranked in ascending order with a “1” next to the most preferred candidate, a “2” next to the second most preferred candidate/athlete and so forth.
(In some implementations, the voter ranks as many or as few choices as they wish while in others they are required to rank all of the athletes or a prescribed number of them.)

(2) In the initial count, the first preference of each voter is counted and used to order the athletes. Each first preference counts as one vote for the appropriate athlete.

(3) Once all the first preferences are counted, if one athlete holds a majority (more than 50% of votes cast), that athlete wins. Otherwise the athlete who holds the fewest first preferences is eliminated. (If there is an exact tie for last place in numbers of votes, tie-breaking rules determine which candidate to eliminate.)

(4) Ballots assigned to eliminated athletes are recounted and assigned to one of the remaining athletes based on the next preference on each ballot.

(5) The process repeats until one candidate achieves a majority (more than 50%) of votes cast for continuing candidates. Ballots that 'exhaust' all their preferences (all its ranked candidates are eliminated) are set aside.

Example 2.3. *Suppose we wish to choose one of the following Quarterbacks who are ranked by ranking systems A through E as follows:*

Rankings − >	RA	RB	RC	RD	RE	RF	#1 Votes in Round 1	#1 Votes in Round 2	#1 Votes in Round 3
Cam Newton	1	1	2	3	5	6	2	2	2
Drew Brees	2	2	1	1	2	4	2	2	4 (winner)
Ben Roethlisberger	3	3	5	5	4	3	0(eliminated)		
Matt Ryan	4	4	4	2	1	5	1	1	
Eli Manning	5	6	3	4	6	2	0(eliminated)		
Matthew Stafford	6	5	6	6	3	1	1	1	

2.4. **Simple Borda Count.** Another method of amalgamating votes is to simply to obtain a rating by averaging the ranks for any given athlete and use the resulting ranking.

If **voters rank the entire list of candidates** or choices in order of preference from the first choice to the last choice then using **Borda's method** the votes are tallied as follows:

- On a particular ballot, the lowest ranking candidate is given 1 point, the second lowest is given 2 points, and so on, the top candidate receiving points equal to the number of candidates .
- The number of points given to each candidate is summed across all ballots. This is called the **Borda Count** for the candidate.
- The winner is the candidate with the highest Borda count.

A Simple Borda Count results in the same ranking as that derived from average rankings: As it turns out the ranking we get for the candidates from a Borda Count is the same as the ranking we get from the ratings calculated by averaging the votes, where a lower rating is given a better(lower) ranking.

Example 2.4. *Let us consider the rankings from the above example and suppose that we wish to choose two of the following Quarterbacks who are ranked by ranking systems A through E by using the Borda Method.*

Rankings − >	RA	RB	RC	RD	RE	RF	Average Rank	Borda Rank
Cam Newton	1	1	2	3	5	6	3	2
Drew Brees	2	2	1	1	2	4	2	1
Ben Roethlisberger	3	3	5	5	4	3	3.83	4
Matt Ryan	4	4	4	2	1	5	3.33	3
Eli Manning	5	6	3	4	6	2	4.33	5
Matthew Stafford	6	5	6	6	3	1	4.5	6

Example 2.5. *The number 1 athlete from the Borda Method does not necessarily always coincide with the number 1 player that we get from the instant runoff method. Consider the following example:*

Borda Method

Rankings − >	RA	RB	RC	RD	RE	RF	RG	RH	Average Rank	Borda Rank
Player 1	1	1	1	3	2	6	5	6	3.125	2
Player 2	2	2	2	2	1	2	2	3	2	1
Player 3	5	6	3	6	3	1	1	2	3.375	4
Player 4	3	4	5	1	4	5	3	1	3.25	3
Player 5	4	3	4	5	5	4	4	5	4.25	5
Player 6	6	5	6	4	6	3	6	4	5	6

vskip .1in

Instant Runoff

We break ties by eliminating the person with the largest vote sum.

	RA	RB	RC	RD	RE	RF	RG	RH	#1's R 1	# 1's R 2	# 1's R 3	# 1's R 4
P 1	1	1	1	3	2	6	5	6	3	3	4	4(wins)
P 2	2	2	2	2	1	2	2	3	1	1*		
P 3	5	6	3	6	3	1	1	2	2	2	2*	
P 4	3	4	5	1	4	5	3	1	2	2	2	4
P 5	4	3	4	5	5	4	4	5	0*			
P 6	6	5	6	4	6	3	6	4	0*			

2.5. **Modified Borda Count:** If there are N candidates and some voters do not rank the entire list of candidates, instead ranking their top M candidates, where $M < N$ we can modify the Borda count in the following way:

- On a particular ballot where the candidates are ranked 1 through M, the lowest ranking candidate is given 1 point, the second lowest is given 2 points, and so on, the top candidate receiving points equal to M . The unranked candidates are given 0 points.
- The number of points given to each caniddate is summed across all ballots. This is called the **Borda Count** for the candidate.
- The winner is the candidate with the highest Borda count.

The **ranking resulting from a modified Borda Count** for an election with N candidates is equivalent to the **ranking resulting from an averaging process** modified as follows:

- On a particular ballot where the candidates are ranked 1 through M and $M < N$, the unranked individuals all receive a rank of M + 1.
- the average across all ballots is then taken for each candidate.

Example 2.6. *To find a ranking for the top quarterbacks, I might choose the top twenty quarterbacks from some of the most accurate ranking methods as measured by* `Fantasy Pros.` *Below I show some preseason rankings for 2015, showing the top twenty players for each ranking if available.*

Rankings $->$	A	B	C	D	E	F
Aaron Rodgers	1	2	1	1	1	1
Andrew Luck	2	1	16	13	3	2
Russell Wilson	3	7	7	4	6	6
Peyton Manning	4	5	4	3	2	3
Tom Brady	5	3	20	19	5	4
Cam Newton	6	8	5	6	12	7
Drew Brees	7	9	3	2	4	5
Ben Roethlisberger	8	10	10	8	11	8
Matt Ryan	9	13	9	5	7	10
Eli Manning	10	17	6	7	14	15
Matthew Stafford	11	15	11	11	9	11
Tony Romo	12	4	2	9	8	9
Ryan Tannehill	13	12	12	10	13	13
Philip Rivers	14	11	15	16	10	12
Carson Palmer	15	19	8		18	19
Colin Kaepernick	16	16	19	20	16	14
Sam Bradford	17	20		12		
Jay Cutler	18	14	14	15	17	16
Robert Griffin III	19		17	17		20
Nick Foles	20	18			19	
Joe Flacco		6				17
Andy Dalton			18	18	20	
Teddy Bridgewater			13	14	15	18

We see here that the top 20 quarterbacks are not the same group for each ranking system. One method of dealing with this is to use the modified Borda Method above with $N = 23$ and $M = 20$. We can fill in the missing rankings with 21 and calculate the average for each player.

Rankings − >	A	B	C	D	E	F	Average	Borda Method Rank
Aaron Rodgers	1	2	1	1	1	1	1.17	1
Andrew Luck	2	1	16	13	3	2	6.17	5
Russell Wilson	3	7	7	4	6	6	5.5	4
Peyton Manning	4	5	4	3	2	3	3.5	2
Tom Brady	5	3	20	19	5	4	9.3	10
Cam Newton	6	8	5	6	12	7	7.3	6
Drew Brees	7	9	3	2	4	5	5	3
Ben Roethlisberger	8	10	10	8	11	8	9.17	9
Matt Ryan	9	13	9	5	7	10	8.83	8
Eli Manning	10	17	6	7	14	15	11.5	12
Matthew Stafford	11	15	11	11	9	11	11.33	11
Tony Romo	12	4	2	9	8	9	7.33	7
Ryan Tannehill	13	12	12	10	13	13	12.17	13
Philip Rivers	14	11	15	16	10	12	13	14
Carson Palmer	15	19	8	21	18	19	16.67	16
Colin Kaepernick	16	16	19	20	16	14	16.83	17
Sam Bradford	17	20	21	12	21	21	18.67	20
Jay Cutler	18	14	14	15	17	16	15.67	15
Robert Griffin III	19	21	17	17	21	20	19.17	21
Nick Foles	20	18	21	21	19	21	20	23
Joe Flacco	21	6	21	21	21	17	17.83	19
Andy Dalton	21	21	18	18	20	21	19.83	22
Teddy Bridgewater	21	21	13	14	15	18	17	18

2.6. **Head-To-Head Comparisons.** Another, more time consuming, method of amalgamating rankings is by running a **head-to-head comparison** between each pair of athletes. If we know the preference rankings of the voters, we can compare the votes of any two athletes to see which one would win in a plurality election in the absence of the other candidates. Such a comparison is called a **Head-To-Head** comparison of the two candidates.

Each matchup gives either a winner and loser or a draw between the two athletes compared. Thus if we compare every possible pair of candidates in this way, we can treat the results as we would those of a round robin tournament. We could rank candidates by wins minus losses (see **Copeland's method below**) or if we keep track of the magnitude of the differences, we could use the equivalent of the point differential (voters in favor minus voters against).

Condorcet Winner An athlete who is the winner of a head-to-head comparison with every other athlete is called a **Condorcet winner**. If an athlete beats or ties with every other athlete in a head-to-head comparison, that athlete is called a **weak Condorcet winner**. For any set of rankings (assuming all athletes are ranked), there may or may not be a Condorcet winner or a weak Condorcet winner. If a Condorcet winner exists, he/she is unique whereas there may be more than one weak Condorcet winner.

Copeland's method This is a simple method often used to decide a Round Robin Tournament, where every athlete plays every other athlete exactly once. With Copeland's method, athletes are ordered by the number of pairwise victories minus the number of pairwise defeats. This method is easily understood and easy to calculate, however it often leads to ties and puts more emphasis on the number of victories and defeats rather than their magnitude. Note that a Condorcet winner will be the highest ranked athlete when we apply Copeland's method.

Example 2.7. *Consider the following players ranked according to rankings $A - E$.*

Rankings$->$	**A**	**B**	**C**	**D**	**E**
Player 1	3	2	5	5	5
Player 2	1	1	2	2	2
Player 3	5	5	1	1	1
Player 4	4	3	3	3	4
Player 5	2	4	4	4	3

(a) Rank the players using Borda's method

Rankings$->$	**A**	**B**	**C**	**D**	**E**	Average	Borda Rank
Player 1	3	2	5	5	5	4	5
Player 2	1	1	2	2	2	1.6	1
Player 3	5	5	1	1	1	2.6	2
Player 4	4	3	3	3	4	3.4	3
Player 5	2	4	4	4	3	3.4	3

(b) Rank the players using Copeland's method?

We fill in the table below where each entry shows the results of a head-to-head comparison between the row player and the column player with the number of voters in favor of the row player appearing first and the number in favor of the column player appearing second. $W - L$ denotes wins minus losses (for pairwise comparisons) for each player and the final column shows the resulting ranking.

	Player 1	Player 2	Player 3	Player 4	Player 5	W - L	Copeland's Rank
Player 1	–	$0-5$	$2-3$	$2-3$	$1-4$	-4	5
Player 2	$5-0$	–	$2-3$	$5-0$	$5-0$	2	2
Player 3	$3-2$	$3-2$	–	$3-2$	$3-2$	4	1
Player 4	$3-2$	$0-5$	$2-3$	–	$3-2$	0	3
Player 5	$4-1$	$0-5$	$2-3$	$2-3$	–	-2	4

(c) Which player would you choose using the plurality method with a runoff between first and second place winners?

We see that Player 3 already has more than 50% of the number 1 votes, therefore Player 3 is the winner using the Instant Runoff method or the plurality method with a runoff between the top two players.

(d) Which player, if any, is the Condorcet winner?

Since Player 3 beats all other players in a head-to-head comparison, Player 3 is a Condorcet winner.

3. A Perfect Voting System

In this section, the term *voting system* will refer to a method of amalgamating rankings such as the ones we have discussed above, the Borda method, the instant runoff method or Copeland's method. We saw that there are drawbacks to each type of voting system we studied. The plurality method with three or more candidates may lead to a winner who is undesirable to the majority of voters. We saw that a Condorcet winner may not be the winner using the Borda method and in fact may not even make it into the runoff in a Plurality voting system with runoff. Furthermore a Condorcet winner may not even exist.

It is natural to try to find a perfect (or good) voting system. In order to explore the possibilities, we must first define what we mean by perfect or good. We follow the ideas of Nobel Laureate Kenneth Arrow who, beginning in the 1940's explored methods of ordering choices among public policies. We first begin with a list of properties that most people would consider desirable in a voting system.

Universal Domain Any ordering of the candidates/players is allowed, that is, there are no restrictions placed on the ranking of the candidates/players a voter may choose.

Pareto optimality If all voters prefer candidate/player A to candidate/player B, then the group choice should not prefer candidate/player B to candidate/player A.

Non-Dictatorship No one individual voter preference totally determines the group choice.

Independence from irrelevant alternatives If a group of voters choose candidate/player A over candidate/player B, then the addition or subtraction of other candidates/players should not change the group choice to B.
This requirement is the most debatable. On the one hand a choice between A and B should not depend on what other choices are available, on the other hand however, it is only by comparison with other possibilities that voters' perception of differences between candidates/players can be brought to light.

Example 3.1. *Plurality with runoff method violates the independence from irrelevant alternatives condition:*

Rankings− >	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Player 1	1	1	1	1	2	2	2
Player 2	2	2	2	2	1	1	1

Rankings− >	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Player 1	2	2	2	1	2	2	2
Player 2	3	3	3	2	1	1	1
Player 3	1	1	1	3	3	3	3

We see in the example on the left above that 4 out of 7 voters prefer Player 1 toPlayer 2 and that Player 1 is destined to win. On the right we introduce a third player, Player 3, and three of those who initially ranked Player 1 first, now rank Player 3 as number 1, but still prefer Player 1 to Player 2. We see that Player 2 now wins (getting 4 number 1 votes after Player 1 is eliminated), which means that the introduction of the irrelevant alternative Player 3 reversed the outcome.

Arrow's Impossibility Theorem There is no voting system based on **rankings** that satisfies the properties of universal domain, Pareto optimality, non-dictatorship and independence from irrelevant alternatives.

It is important to note here that Arrows theorem applies in the context of rankings, where voters give an ordering of their preferences. It does not apply to the situation where voters give a measure of the worth or utility or strength of performance of each candidate. Basically it says that it is impossible to find a function or rule that will amalgamate a sequence of individual rankings or ballots (represented by lists or lists with ties) in a reasonable way. We also note that there are many other desirable properties of voting systems which are not listed above and should be considered when choosing a way to amalgamate ballots.

Example 3.2. *Copeland's method violates the independence from irrelevant alternatives criterion. Suppose you wish to amalgamate the given rankings for the following players using Copeland's method in order to rank players from this category for your draft picks:*

#Voters	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Player 1	1	1	2	2	2	2	3	3	2	5
Player 2	4	4	1	1	1	1	2	4	5	4
Player 3	3	3	3	3	5	5	1	1	4	2
Player 4	2	2	4	4	3	3	4	2	1	3
Player 5	5	5	5	5	4	4	5	5	3	1

We get

Rankings$->$	Player 1	Player 2	Player 3	Player 4	Player 5	W - L	Copeland's Rank
Player 1	–	$4-6$	$7-3$	$7-3$	$9-1$	2	1^*
Player 2	$6-4$	–	$4-6$	$5-5$	$8-2$	1	2
Player 3	$3-7$	$6-4$	–	$5-5$	$6-4$	1	2
Player 4	$3-7$	$5-5$	$5-5$	–	$9-1$	0	4
Player 5	$1-9$	$2-8$	$4-6$	$1-9$	–	-4	5

Now right before the draft, one of the players (Player 3) gets injured and is out for the season. Since Player 3 was not the top choice, this fact should have no effect on the choice of Player 1 as the first draft, or should it ?

Lets see, we rerank the players accordingly and calculate Copeland's ranking:

#Voters	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Player 1	1	1	2	2	2	2	2	2	2	4
Player 2	3	3	1	1	1	1	1	3	4	3
Player 4	2	2	3	3	3	3	3	1	1	2
Player 5	4	4	4	4	4	4	4	6	3	1

We get

Rankings$->$	Player 1	Player 2	Player 4	Player 5	W - L	Copeland's Rank
Player 1	–	$4-6$	$7-3$	$9-1$	1	2
Player 2	$6-4$	–	$5-5$	$8-2$	2	1^*
Player 4	$3-7$	$5-5$	–	$9-1$	0	3
Player 5	$1-9$	$2-8$	$1-9$	–	-3	4

4. Assigning a Value To a Player

Ideally we would like to assign the points that the player will score throughout the upcoming season as the value of the player. Unfortunately, these statistics are not available to us before the season begins.

4.1. **Using Last Season's Points.** We could use last season's points for each player as their value. First we might look at how accurate the previous season's points were as a predictor for the current season's points from historical data.

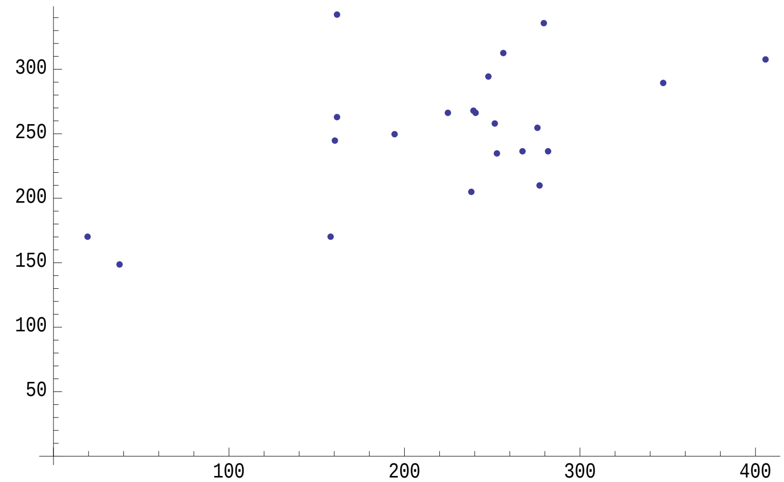
In this table we show the scores for two consecutive seasons for 22 quarterbacks.

	2013	2014
Aaron Rodgers	162	342
Andrew Luck	279	336
Russell Wilson	256	312
Peyton Manning	406	307
Ben Roethlisberger	248	295
Drew Brees	348	290
Matt Ryan	239	268
Tom Brady	241	267
Ryan Tannehill	225	266
Eli Manning	162	263
Tony Romo	252	258
Philip Rivers	276	254
Joe Flacco	194	249
Jay Cutler	160	244
Matthew Stafford	267	237
Cam Newton	282	237
Colin Kaepernick	253	234
Andy Dalton	277	210
Alex Smith	238	205
Kyle Orton	19	171
Ryan Fitzpatrick	158	171
Brian Hoyer	38	149

We can make a picture called a scatterplot from this data, which helps us determine if there is a strong relationship between them. We plot the points from the 2013 season on the horizontal axis and those from the 2014 season on the vertical axis.

- If the dots (roughly) lie along a line sloping upwards then there is a positive relationship between them and higher points in 2013 is related to a higher number of points in 2014 and vice versa. A steeper slope to the line indicates a stronger relationship between the two variables.
- If the points on the scatterplot lie along a line sloping downwards there is a negative relationship between the 2 variables and higher points in 2013 are related to lower points in 2014.
- If the points lie along a horizontal line then all of the players got roughly the same number of points in 2014 no matter what the outcome in 2013.
- There is no relationship between the two variables if the points form a disc like shape or a nonlinear shape.

The scatterplot for our data lis as follows:



4.2. **Fitting a line to the data : The Least squares line.** Given a set of data points in the xy-plane, $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \quad \dots \quad, (x_n, y_n)\}$ such as those shown above, we can find an equation for the line which best fits the data using the method of least squares. This line minimizes the squares of the difference between the y values on the line and the y values for the points in the data. This equation gives us a linear formula which estimates the relationship between the variable x and the variable y which we can use for predictions.

Recall that the equation of a line is of the form $y = \beta_0 + \beta_1 x$ where β_0 and β_1 are constants. The idea is to find values of β_0 and β_1 so that the sum

$$\text{SSE} = ((y_1 - y(x_1))^2 + (y_2 - y(x_2))^2 + \cdots + (y_n - y(x_n))^2)$$

is minimal where $y(x_i)$ is tha value corresponding to x_i from the formula $y = \beta_0 + \beta_1 x$ and y_i is the value corresponding to x_i in the datapoint (x_i, y_i). (SSE stands for the sum of the squared errors.)

We see from this interactive `demonstration` on `Wolfram Alpha Demonstrations` that for some lines the sum of the squared errors is larger than for others.

Example 4.1. *Consider the data for our quarterbacks above. Let x_i be the number of fantasy points scored by quarterback i in 2013 and let y_i denote the number of fantasy points scored by quarterback i is 2014. Lets calculate the sum of the squared error (SSE) for a particular line $y = 150 + (0.5)x$.*

points 2013	points 2014	predicted values	Errors	Squared Errors
x_i	y_i	$y(x_i) = 150 + (0.5)x_i$	$y_i - y(x_i)$	$(y_i - y(x_i))^2$
162	342	231	111	12321
279	336	289.5	46.5	2162.25
256	312	278	34	1156
406	307	353	-46	2116
248	295	274	21	441
348	290	324	-34	1156
239	268	269.5	-1.5	2.25
241	267	270.5	-3.5	12.25
225	266	262.5	3.5	12.25
162	263	231	32	1024
252	258	276	-18	324
276	254	288	-34	1156
194	249	247	2	4
160	244	230	14	196
267	237	283.5	-46.5	2162.25
282	237	291	-54	2916
253	234	276.5	-42.5	1806.25
277	210	288.5	-78.5	6162.25
238	205	269	-64	4096
19	171	159.5	11.5	132.25
158	171	229	-58	3364
38	149	169	-20	400
			$SE = -225$	$SSE = 43122.25$

From the table above , we see that for the line $y = 150+(0.5)x$, the sum of the errors is an unreliable statistic in measuring how well the line fits the data due to cancellation. We avoid this problem by squaring the error and use the Sum of Squares of The Error (SSE) to measure how well the line fits the data. Naturally a smaller SSE will indicate that a line is a better fit for the data.

There is a unique line for which SSE is at a minimum. This line is called the Least Squares Line. The methodology used to obtain the equation of this line is called the method of least squares.

We can solve for the coefficients β_0 and β_1 of such a line, for a particular set of data, by using calculus to find the minimum of the function

$$SSE = \Sigma[y_i - (\beta_0 + \beta_1 x_i)]^2,$$

for the variable β_0 and β_1.

Definition 4.1. *The least squares line $y = \beta_0 + \beta_1 x$, for a set of data, is the unique line with the following properties:*

(1) The sum of errors equals 0; SE = 0

(2) The sum of squared errors (SSE) is smaller than that for any other straight line model.

The values of β_0 and β_1 for the least squares line are given by the following formulas (where $\bar{x}$ and $\bar{y}$ demote the means of the data sets $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$ respectively):

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}}, \qquad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

where

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

and

$$SS_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

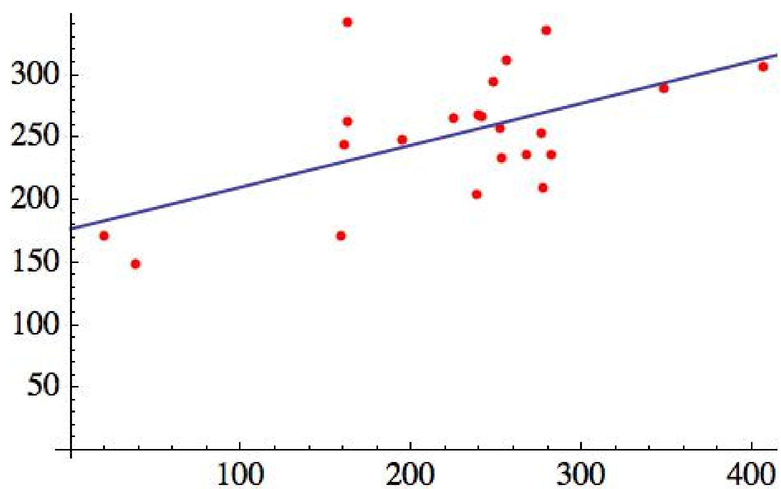
We have

$$SSE = \sum (y_i - (\beta_0 + \beta_1 x_i))^2 = SS_{yy} - \beta_1 SS_{xy}$$

where

$$SS_{yy} = \sum y_i^2 - \frac{\sum (y_i)^2}{n}.$$

Example 4.2. *For the above data, the least squares line is given by $y = 177.06 + 0.33x$ and the sum of the squared errors for this line is $SSE = 36487.3$. This is less than the sum of the squared errors for the line shown above, in fact it is the minimum such sum possible for any line that we might fit to the data.*



We could use this line to estimate the fantasy points that a quarterback will score id 2015 given the number of points he scored in 2014. For example if a quarterback scored 250 points in fantasy football in 2014, we might expect the number of points he scores in 2015 to be roughly $177.06 + 0.33(250) \approx 259.56$.

4.3. **Calculating the Least Squares Line.** We can often use our calculator to get values for $\bar{x}$, $\bar{y}$, $\sum x_i^2$, $\sum x_i y_i$, $\sum x_i$, $\sum y_i$:
For example for the TI83 we follow the following steps:

- Step 1: Put the data sets in L_1 and L_1.
- Step 2: Press `STAT`, `CALC`, `2-Var Stats`, L_1, L_2, `ENTER` to get a list of statistics including those above.
- We then use the formulas given above to calculate values for $\hat{\beta}_0$ and $\hat{\beta}_1$.

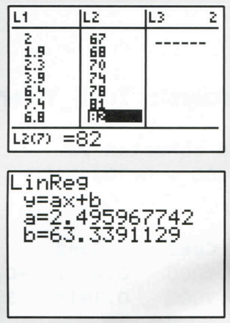
We can also find the values of β_0 and β_1 directly on the TI83:

Step 1 *Enter the data*
Press **STAT** and select **1:Edit**
Note: If a list already contains data, clear the old data. Use the up arrow to highlight the list name, '**L1**' or '**L2**.'

Press **CLEAR ENTER**.
Enter your x-data in **L1** and your y-data in **L2**.

Step 2 *Find the equation*
Press **STAT** and highlight **CALC**
Press 4 for **LinReg(ax + b)**
Press **ENTER**
The screen will show the values for a and b in the equation $\mathbf{y = ax + b}$.

Example The figures below show a table of data entered on the TI-83 and the regression equation obtained using the steps given above.



The method described above is often called **Linear Regression** and you can solve for the coefficients β_0 and β_1 using most statistical packages such as those that come with `Excel` or the packages `stata` or `SPSS`.

4.4. **Taking the effect of Several Variables into Account.** It is likely that a number of variables will have a positive or negative influence the number of points scored in the upcoming season. I have listed a few candidates below:

- H = Height of player
- A = Age of player
- GP = Games Played (career)
- C = Completion percentage (previous season)
- AY = yards per pass attempted (previous season)
- QBR = Total Quarterback rating (ESPN Previous season)
- INT = Interception percentage (previous season)
- FUM = Fumbles (previous season)
- LNG = Longest Pass Play (career)
- PP = Fantasy points previous season.

We can collect data on these variables and consider their relative weight or influence on the points scored in the 2014 season using a generalization of the method of linear regression described above. The method is relatively easy to apply if you have some statistical software. In Chapter 3 his book Mathletics, Winston [3] describes how to run such a regression in `Excel` and gives a compact explanation of how to interpret the results. You will find a more detailed explanation in [1].

4.5. **The Optimal combination.** Although we cannot explore the (knapsack) problem of maximizing expected points scored subject to budget constraints without some techniques from linear programming or some computer programming skills, we can have a quick look at how we might go about setting up the problem in a way that could be solved with the help of a computer. If you use Google to search for helpful sites on solutions to the knapsack problem in fantasy football you will find many helpful sites such as `this one`.

Lets suppose

- you want to choose 3 players from a set of
  3 quarterbacks,
  2 Running Backs,
  3 wide receivers
- You must choose at least one quarterback
- You do not wish to choose more than 1 quarterback
- You must not spend more than $45

The value of each player (the number of points you expect them to earn or last year's points) and their cost is shown in the table below. (Quarterbacks are named Q1-Q3, Running Backs are named RB1, RB2, Wide Receivers are named WR1-WR3)

Player	Points Exp.	Cost
QB1	250	$20
QB2	220	$18
QB3	170	$15
RB1	270	$20
RB2	225	$16
WR1	230	$14
WR2	210	$12
WR3	150	$5

For each player we create a variable which can take one of the values 0 or 1. These variables will represent the number of each player that we will purchase. We call the variables q_1, q_2, q_3 for the quarterbacks, r_1, r_2 for the running backs and w_1, w_2, w_3 for the wide receivers. Let $P(q_1)$ denote the points we expect quarterback 1 to earn next season and let $C(q_1)$ denote their cost etc...

Our problem with our constraints can be summarized as follows:
We wish to find the values of $q_1, q_2, q_3, r_1, r_2, w_1, w_2, w_3$ such that
the sum $q_1P(q_1) + q_2P(q_2) + q_3P(q_3) + r_1P(r_1) + r_2P(r_2) + w_1P(w_1) + w_2P(w_2) + w_3P(w_3)$ is maximized subject to the constraints;
$q_1C(q_1) + q_2C(q_2) + q_3C(q_3) + r_1C(r_1) + r_2C(r_2) + w_1C(w_1) + w_2C(w_2) + w_3C(w_3) \leq 45$.
We must also have that $1 = q_1 + q_2 + q_3$ (exactly one quarterback), and $q_1 + q_2 + q_3 + r_1 + r_2 + w_1 + w_2 + w_3 = 3(a total of 3 players)$.

Now since each one of the 8 variables must have the value 0 or 1, there are 256 possible sequences of 0's and 1's of length 8. However not all of the 256 such sequences will satisfy our constraints. The constraint $1 = q_1 + q_2 + q_3$ means that exactly one of the first three variables must be equal to 1. Putting this together with $q_1 + q_2 + q_3 + r_1 + r_2 + w_1 + w_2 + w_3 = 3$, we gat that exactly two of the 5 variables r_1, r_2, w_1, w_2, w_3 must equal 1. There are exactly 10 such combinations of these 5 variables. We can put any of these 10 combinations together with the 3 choices of quarterback to make 30 possible combinations which fit the last two constraints on the list. Below, we calculate the cost for each of these to check if it is a feasible combination satisfying the budget and then we choose the one which maximizes points.

q_1	q_2	q_3	r_1	r_2	w_1	w_2	w_3	Cost	Points
1	0	0	1	1	0	0	0	56	
1	0	0	1	0	1	0	0	54	
1	0	0	1	0	0	1	0	52	
1	0	0	1	0	0	0	1	45	670 (Max)
1	0	0	0	1	1	0	0	50	
1	0	0	0	1	0	1	0	48	
1	0	0	0	1	0	0	1	41	625
1	0	0	0	0	1	1	0	46	
1	0	0	0	0	0	1	1	37	610
1	0	0	0	0	1	0	1	39	630
0	1	0	1	1	0	0	0	54	
0	1	0	1	0	1	0	0	52	
0	1	0	1	0	0	1	0	50	
0	1	0	1	0	0	0	1	43	640
0	1	0	0	1	1	0	0	48	
0	1	0	0	1	0	1	0	46	
0	1	0	0	1	0	0	1	39	595
0	1	0	0	0	1	1	0	44	660
0	1	0	0	0	0	1	1	35	580
0	1	0	0	0	1	0	1	37	600
0	0	1	1	1	0	0	0	51	
0	0	1	1	0	1	0	0	49	
0	0	1	1	0	0	1	0	47	
0	0	1	1	0	0	0	1	40	
0	0	1	0	1	1	0	0	51	
0	0	1	0	1	0	1	0	47	
0	0	1	0	1	0	0	1	40	545
0	0	1	0	0	1	1	0	41	610
0	0	1	0	0	0	1	1	32	530
0	0	1	0	0	1	0	1	34	550

Thus we see that the feasible solution that gives the maximum number of points is QB1, RB1 and WR3.

There are algorithmic ways of tackling this problem using the Branch and Bound Method for Integer Programming along with Techniques from Linear Programming along with some dynamic programming methods. Computers and computer programming are obviously necessary for finding an optimal solution when larger numbers of players are involved.

References

1. J. McClave and T. Sincich, *Statistics*, Pearson.
2. Phillip D. Straffin, *Game theory and strategy*, The Mathematical Association of America.
3. Wayne Winston, *Mathletics*, Princeton University Press.