Learning Goals

1. Borda's Method
2. Calculating a simple Borda count.
3. Calculating a simple Borda count with `R`.
4. Calculated a Modified Borda Count.
5. Calculating a Modified Borda count with `R` (dealing with missing data).

Topic 6 : Borda's method: A Scoring System

In the plurality and runoff methods discussed in the previous lecture, we do not take into account the voter's relative preferences for *all* of the candidates. We do not, for example take into account which candidate was ranked last by each voter.

Borda's Method

Lets assume that each judge or voter ranks the candidates in order of preference. The Borda count determines the outcome of a competition or the winner of an election by giving each candidate, for each ballot, a number of points corresponding to the number of candidates ranked lower(or equal).

Simple Borda Count 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 If a Borda count is used to rank the the three basketball players from our example in the previous lecture, who is Number 1?

	Off. Ranking UPROXX	Sports Ill.	Was. Post	Fantasy BB Power R	FG%	=/-	OE
Kevin Durant	1	2	3	1	1	2	2
LeBron James	2	1	1	3	2	3	1
Stephen Curry	3	3	2	2	3	1	3

Example In a survey, squash players were asked to rank brands of squash racquets. The results are shown below:

# Players →	**33**	**3**	**10**	**20**	**7**	**27**
Dunlop	1	1	2	3	2	3
Black Knight	2	3	1	1	3	2
Prince	3	2	3	2	1	1

(a) Which brand would win using Borda's method?

(b) Which brand would win using the Plurality method?

(c) Which brand would win using the Plurality method with a runoff between the first and second place finishers?

Example R If there are a lot of ballots, you may want to use a computer to calculate the average. The csv file in Sakai caleed `Ballots.csv` contains the results of a ballot that required voters to rank their preferences for 5 different types of pets. (where 1 denotes the most preferred pet type and 5, the least preferred).

Import the data

```
Ballots<- read.csv(file="Ballots.csv", header=FALSE, sep=",")
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'Ballots.csv'
```

look at the data

```
head(Ballots)
```

```
##        V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
## 1    Dogs  1  1  1  1  1  1  1  1   1   1   1   2   2   2   2   2   3   3
## 2 Parrots  2  2  2  2  3  3  4  5   5   5   5   1   3   3   5   5   1   1
## 3    Fish  3  3  4  5  4  5  5  2   2   3   4   3   4   5   1   4   2   5
## 4    Cats  4  5  3  3  5  2  2  3   4   2   3   4   1   1   3   1   4   4
## 5   Goats  5  4  5  4  2  4  3  4   3   4   2   5   5   4   4   3   5   2
##   V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36
## 1   3   3   3   3   3   3   4   4   4   4   4   4   5   5   5   5   5
## 2   2   2   4   5   5   5   1   3   3   5   5   5   1   2   4   4   4
## 3   4   5   5   2   4   4   3   1   2   1   1   3   4   1   1   2   3
## 4   1   1   1   1   1   2   2   2   1   2   3   1   3   3   3   3   1
## 5   5   4   2   4   2   1   5   5   5   3   2   2   2   4   2   1   2
```

notice R attached names to the columns; V1-V36 we can find out how many columns there are in our data with

```
ncol(Ballots)
```

```
## [1] 36
```

apply mean(average) function to the 36 numbers in columns 2-36 for each row

```
Ballots$Average<-apply(Ballots[ ,2:36],1,mean)
```

Order by variable Average

```
Ballots1<-Ballots[order(Ballots$Average),]
```

Show names alongside ordered average rank

```
Ballots2<-data.frame(Ballots1$V1,Ballots1$Average)
Ballots2
```

```
##   Ballots1.V1 Ballots1.Average
## 1        Cats         2.428571
## 2        Dogs         2.685714
## 3        Fish         3.142857
## 4     Parrots         3.342857
## 5       Goats         3.400000
```

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 Suppose, we had our 10 judges rank the movies nominated for best picture in the Oscars (in 2017). Now perhaps the judges had not seen all of the movies and everyone just ranked the ones they had seen and the results were as follows:

# Voters	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
The Big Short	4	5	8	3	3	3	1	2	
Bridge of Spies	1	3	7	2	4	2		1	2
Brooklyn	3	2	6	1			3	3	1
Mad Max: Fury Road	2	1	5	4	2	1	2		
Bring Him Home			4	5					4
The Revenant	5		3	6	1		4		5
Room		4	2	7					6
Spotlight			1	8					3

Replace the empty spaces with the appropriate numbers, calculate averages and find the resulting rankings.

Advantages, Disadvantages

The **advantages** of Borda's method (average) over plurality methods is

1. that voters are able to express their opinions about candidates other than just their first choice. This means that a candidate who is ranked highly but not necessarily first by many voters has a good chance of winning when using Borda's method.

2. The average vote for a large amount of data can be computed easily on a computer.

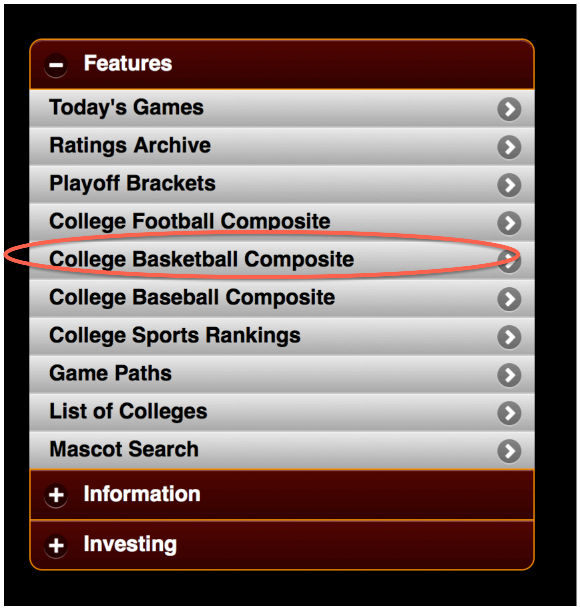
The **disadvantage** of using Borda's method (average) is

1. that it is more susceptible to strategic voting than either the Plurality or Runoff Plurality methods (it is very easy to increase the average vote of a candidate you do not like).

Using a Borda Count to make a Ranking for Basketball:

Let's download some "biggish" data on rankings from the internet relating to the NCAA mens basketball tournament and calculate the averages with R. We can then sort the file with the averages in increasing order to get our own ranking.

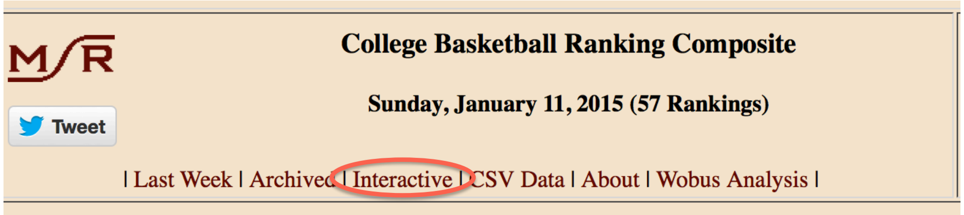
To get some data, we visit Massey's webpage: `http://www.masseyratings.com/`
click on College Basketball Composite in the panel on the right



This brings you to a page with current rankings for College Basketball:
`http://www.masseyratings.com/cb/compare.htm`

This page shows 55 or so current rankings(updated each Sunday during the season), the full names of which are shown across the top of the page. It also shows the average, median and standard deviation of the rankings at the end. We will study two types of computer rankings in greater detail in later lectures, namely the Colley and Massey rankings. As you can see, they give different results.

Personally from here I like to go to the interactive page by clicking on `interactive` in the panel above the rankings:

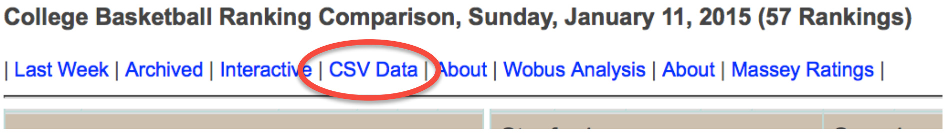

College Basketball Ranking Composite

Sunday, January 11, 2015 (57 Rankings)

M/R
Tweet
| Last Week | Archived | Interactive | CSV Data | About | Wobus Analysis |

http://www.masseyratings.com/cb/comp.htm

On this page, you can choose a particular ranking by clicking on the top of the associated column and the teams will be sorted accordingly.


College Basketball Ranking Comparison, Sunday, January 11, 2015 (57 Rankings)

| Last Week | Archived | Interactive | CSV Data | About | Wobus Analysis | About | Massey Ratings |

We can now open the data with Excel by clicking on: `CSV Data` at the top of the page. To avoid some messiness, we will clean the data up a little in Excel first(here the data at the top has a smaller number of columns than the data below it causing `R` to create less columns than needed for the rankings. We see that it has a glossary of abbreviations at the beginning, we delete this or save it in a separate file. We leave just the abbreviations of the rankings and the teams and their rankings. We will then save it as a csv file in a file called `compare.csv`.

Example What is the ranking resulting from a Borda count (average rank) using the following polls: AP, Coaches Poll(USA), Sagarin(SAG), Massey(MAS) and Pomeroy(POM) ?

BASKETBALL RANKINGS

Import the data:

```
rankings<-read.csv("compare.csv",header=TRUE)
```

Create data frame with preferred rankings

```
myrankings<-data.frame(rankings$Team,rankings$AP,
                      rankings$USA,rankings$SAG,rankings$COL,rankings$MAS,rankings$POM)
#or
myrankings<-data.frame(Team=rankings$Team,AP=rankings$AP,USA=rankings$USA,
                       SAG=rankings$SAG,COL=rankings$COL,MAS=rankings$MAS,POM=rankings$POM)
```

We see that the missing rankings in the AP and USA column have been replaced by NA's

```
tail(myrankings)
```

```
##                    Team AP USA SAG COL MAS POM
## 348 Alabama A&M       NA  NA 349 347 343 346
## 349 Chicago St        NA  NA 351 342 349 348
## 350 UNC Asheville     NA  NA 345 352 353 351
## 351 MS Valley St      NA  NA 350 349 350 350
## 352 MD E Shore        NA  NA 352 350 351 353
## 353 Delaware St       NA  NA 353 353 352 352
```

The command "is.na(myrankings)" runs though the file and creates a new data.frame with TRUE whereever the NA's are and FALSE whereever there are numbers:

```
checkitout<-is.na(myrankings)
tail(checkitout)
```

```
##         Team   AP  USA   SAG   COL   MAS   POM
## [348,] FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [349,] FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [350,] FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [351,] FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [352,] FALSE TRUE TRUE FALSE FALSE FALSE FALSE
## [353,] FALSE TRUE TRUE FALSE FALSE FALSE FALSE
```

Now we want to replace the NA's by 26's since both the AP and USA rankings stop at 25. This command takes the subset of all places where is.na(myrankings) is equal to TRUE in the data.frame myrankings and relaces the NA's by 26

```
myrankings[is.na(myrankings)]<-26
```

```
checkitoutnow<-is.na(myrankings)
tail(checkitoutnow)
```

```
##         Team    AP   USA   SAG   COL   MAS   POM
## [348,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [349,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [350,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [351,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [352,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [353,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
tail(myrankings)
```

```
##                    Team AP USA SAG COL MAS POM
## 348 Alabama A&M       26  26 349 347 343 346
## 349 Chicago St        26  26 351 342 349 348
## 350 UNC Asheville     26  26 345 352 353 351
## 351 MS Valley St      26  26 350 349 350 350
## 352 MD E Shore        26  26 352 350 351 353
## 353 Delaware St       26  26 353 353 352 352
```

Calculate the mean of the rows:

```
myrankings$ave<-apply(myrankings[,2:7],1,mean)
```

Order the teams:

```
myrankings1<-myrankings[order(myrankings$ave),]
```

Print out the top 25(rows 1-20) team names and average rankings in columns 1 and 8 to get your top 25 teams:

```
myrankings1[1:20,c(1,8)]
```

```
##                  Team       ave
## 1  Duke               2.333333
## 2  Virginia           2.666667
## 4  Tennessee          3.166667
## 3  Michigan St        4.333333
## 5  Gonzaga            4.833333
## 6  Michigan           5.166667
## 9  Kentucky           7.833333
## 7  North Carolina     9.166667
## 8  Kansas             9.500000
## 11 Nevada            12.333333
## 12 Virginia Tech     13.333333
## 13 Houston           14.333333
## 10 Purdue            14.833333
## 16 Villanova         14.833333
## 14 Louisville        15.333333
## 21 Marquette         16.500000
## 19 Texas Tech        17.833333
## 15 Iowa St           18.666667
## 17 LSU               19.500000
## 20 Wisconsin         20.166667
```

The command myrankings1[1:20,c(1,8)] shows the first 20 rows of the data frame and the first and eighth column. The command 1:20 creates a vector which includes all numbers 1 through 20.

R commands for importing Data and manipulating it in Dataframes.

R commands: New Tricks

1. `a<-read.csv("b.csv")` : read the data in the csv file `b.csv` into a data.frame called `a` in `R`.

2. `df[c(2,3),c(4,5)]` : gives the data in rows 2 and 3 of the data frame `df` which is also in columns 4 and 5.

3. `ab[-7, ]` : gives all rows of the data frame `ab` except `7`.

4. `cols135<-mydata[ ,c(1,3,5)]`: where `mydata` is a dataframe, gives a new dataframe which is a subset of `mydata` consisting of the data in columns 1, 3, and 5 of `mydata`.

5. `namedcols<-mydata[ ,c("A","B","C")]`: where `mydata` is a dataframe, gives a new dataframe which is a subset of `mydata` consisting of the data in columns corresponding to the variables (named columns) `A, B` and `C` of `mydata`.

6. `is.na(mydata)`: gives a logical vector of `"TRUE"`s and `"FALSE"`s of same dimensions a the dataframe `mydata` showing `"TRUE"`s in the positions where the `NA` entries appear.

7. `mydata[is.na(mydata)]<-0`: replaces all of the `NA`'s in your data by 0.

8. `newdata <- mtcars[order(mtcars$mpg),]` ; sorts the data frame `mtcars` by the variable `mpg` in ascending order and saves result to a new dataframe `newdata`.

9. `apply(X, MARGIN, FUN, ...)` ; applies the function `FUN` to each row of the data frame `X` if `MARGIN` is `1` and to each column if `MARGIN` is `2`.

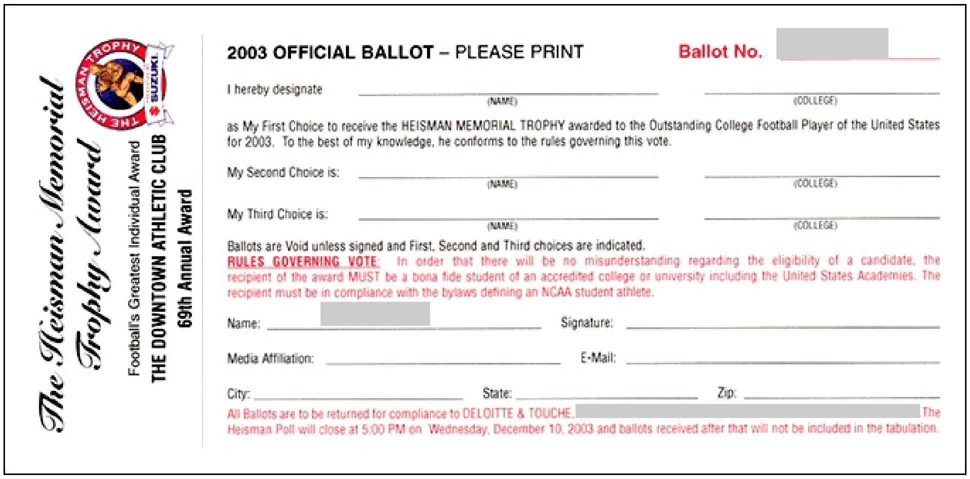
Extras: Variations of The Borda Count in Sports

In sports polls where this form of voting is commonly used, the voters may know a lot about the top teams or players and be able to rank them, but may not know enough to rank all eligible candidates, so lumping all but the top candidates together with 0 points simplifies the process for voters.

Heisman Trophy

A variation of this method is used to decide the winner of the Heissman Trophy. On the Heisman ballot voters are asked to **rank only their top three choices from among all college football players in the United States**. The Borda count for each player is computed by giving 3 points for each first place vote, 2 points for each second place vote and 1 point for each third place vote. The winner is declared to be the candidate with the highest Borda count.

THE 2003 OFFICIAL HEISMAN BALLOT



The Heisman Memorial Trophy Award

Football's Greatest Individual Award

THE DOWNTOWN ATHLETIC CLUB

69th Annual Award

2003 OFFICIAL BALLOT – PLEASE PRINT Ballot No. [redacted]

I hereby designate ______________________ (NAME) ______________________ (COLLEGE)

as My First Choice to receive the HEISMAN MEMORIAL TROPHY awarded to the Outstanding College Football Player of the United States for 2003. To the best of my knowledge, he conforms to the rules governing this vote.

My Second Choice is: ______________________ (NAME) ______________________ (COLLEGE)

My Third Choice is: ______________________ (NAME) ______________________ (COLLEGE)

Ballots are Void unless signed and First, Second and Third choices are indicated.

RULES GOVERNING VOTE: In order that there will be no misunderstanding regarding the eligibility of a candidate, the recipient of the award MUST be a bona fide student of an accredited college or university including the United States Academies. The recipient must be in compliance with the bylaws defining an NCAA student athlete.

Name: [redacted] ______________________ Signature: ______________________

Media Affiliation: ______________________ E-Mail: ______________________

City: ______________ State: ______________ Zip: ______________

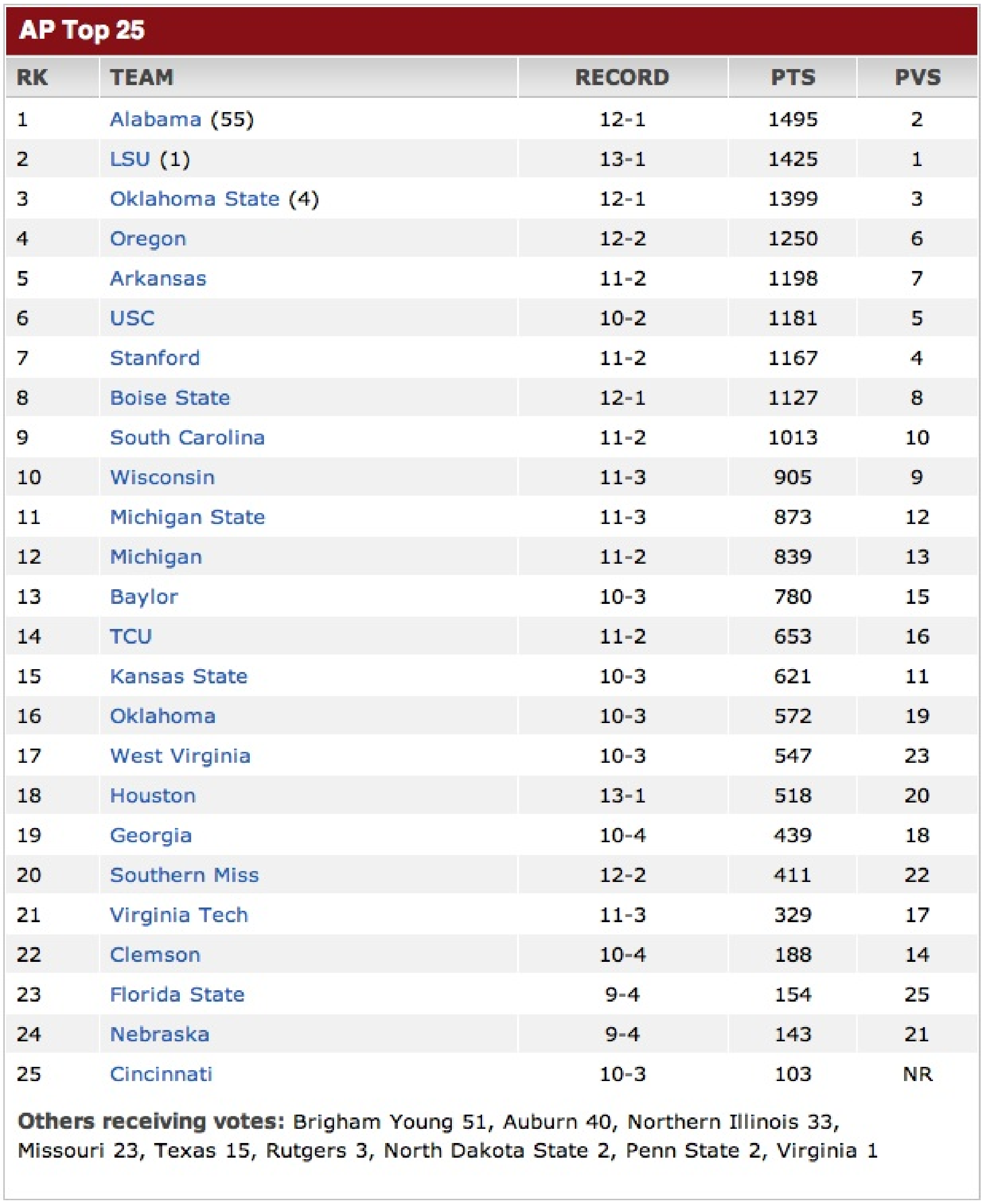
All Ballots are to be returned for compliance to DELOITTE & TOUCHE, [redacted] The Heisman Poll will close at 5:00 PM on Wednesday, December 10, 2003 and ballots received after that will not be included in the tabulation.

NBA most valuable player (uses wider spread of points)

To decide on the winner of the National Basketball Association Most Valuable Player award, **116 members of the media list their first through fifth choices for the award.** Each first place vote receives 10 points, each second place vote receives 7 points, each third place vote receives 5 points, each fourth place vote receives 3 points and each fifth place vote receives 1 point.

AP Polls

The Associated Press (AP) college football poll gives a ranking for the top teams in college football. The voters in the AP poll are newspaper, radio and television sports reporters throughout the country. In 2011 there were 60 voters and the final rankings are given in the table below.



AP Top 25

RK	TEAM	RECORD	PTS	PVS
1	Alabama (55)	12-1	1495	2
2	LSU (1)	13-1	1425	1
3	Oklahoma State (4)	12-1	1399	3
4	Oregon	12-2	1250	6
5	Arkansas	11-2	1198	7
6	USC	10-2	1181	5
7	Stanford	11-2	1167	4
8	Boise State	12-1	1127	8
9	South Carolina	11-2	1013	10
10	Wisconsin	11-3	905	9
11	Michigan State	11-3	873	12
12	Michigan	11-2	839	13
13	Baylor	10-3	780	15
14	TCU	11-2	653	16
15	Kansas State	10-3	621	11
16	Oklahoma	10-3	572	19
17	West Virginia	10-3	547	23
18	Houston	13-1	518	20
19	Georgia	10-4	439	18
20	Southern Miss	12-2	411	22
21	Virginia Tech	11-3	329	17
22	Clemson	10-4	188	14
23	Florida State	9-4	154	25
24	Nebraska	9-4	143	21
25	Cincinnati	10-3	103	NR

Others receiving votes: Brigham Young 51, Auburn 40, Northern Illinois 33, Missouri 23, Texas 15, Rutgers 3, North Dakota State 2, Penn State 2, Virginia 1