

Topic 12 $\frac{1}{2}$:Filling out Brackets

Yes it is time to put all of our expertise to work and fill out brackets (and of course don't forget to celebrate St. Patrick's Day when all of the hard work is over).

Where to fill out your bracket

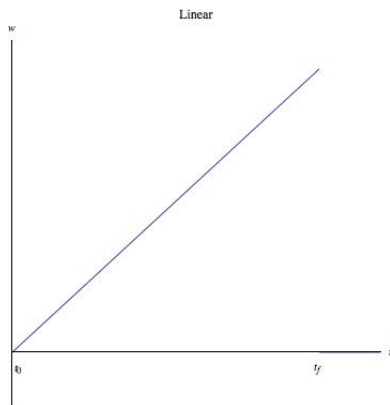
I will set up a group on ESPN and send you a link to our group which will be called **Math In Sports ND 2019**.

If you are an athlete, please download a copy of the bracket and fill it out on paper and submit the paper copy to me (There may be some issue with you filling out a bracket).

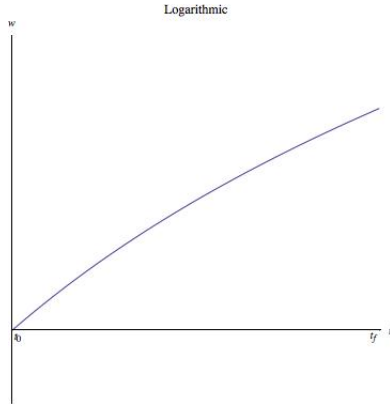
Using Tim Chartier's Applet

The applet is at this web address: <http://marchmathness.davidson.edu/>

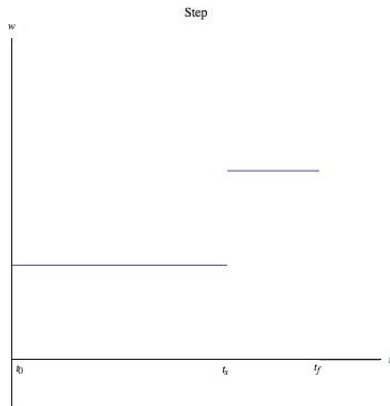
- Click on **Start Ranking** and you get to choose whether you wish to use the Massey or the Colley method.
- When you click on your preferred method, you will be asked if you will be given four options for weighting the games with respect to the time at which they were played. Explanations of the methods are given on the page below the links in the applet.
- After you choose a method to weight with respect to time, you will be asked to choose a weight for home away and neutral games in addition to the necessary parameters for your time weighting method. If you do not wish to weight home and away games, you can just give everything home, away and neutral games a weight of 1. Explanations of all parameters are given below the links in the applet.
- After you have made all of your choices, click **submit** and you will get a list of the teams ordered by rank.
- You can use this ranking to fill out your bracket directly from the webpage or if you wish to use it in a Borda Count, you can save it to a text file or an html file and import it into Excel for further use.
- **Uniform Weighting** With this method all games throughout the season are weighted equally.
- **Linear Weighting** Games are weighted based on the day they were played. The value of games on each day are given by a line through the origin and a slope you provide. Games on the last day are equal to the slope while games on any other day are the slope multiplied by the percent of days in the season that have passed.



- **Log weighted time** Weights are equal to the log of (1+the percent of days in the season that have passed). Therefore, the weight on the first day of the season is equal to $\log(1)$ and the weight on the day you rank the teams is equal to $\log(2)$. The weights on the days between take the shape of a log curve.



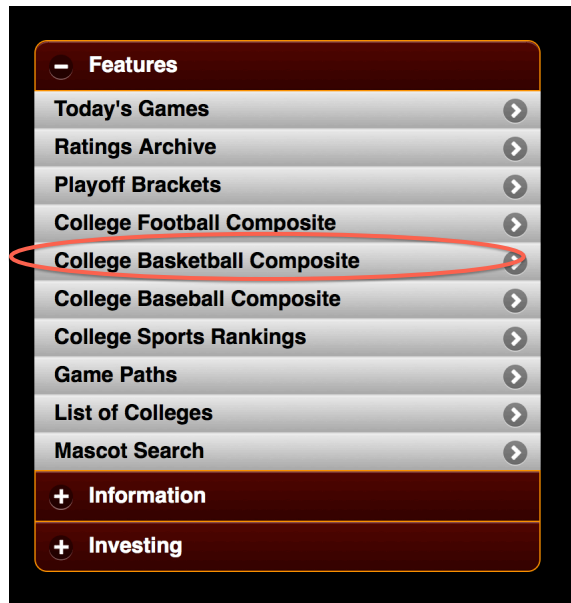
- **Intervals** The season is divided into equal sized intervals, according to the days in the season, with a game in each interval weighted the same. The weights used in each interval are provided by you.



Average of Popular Good Rankings (Equivalent to a Borda Count)

Recall that we can download many of the popular rankings relating to the NCAA mens basketball tournament from Massey's site:

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

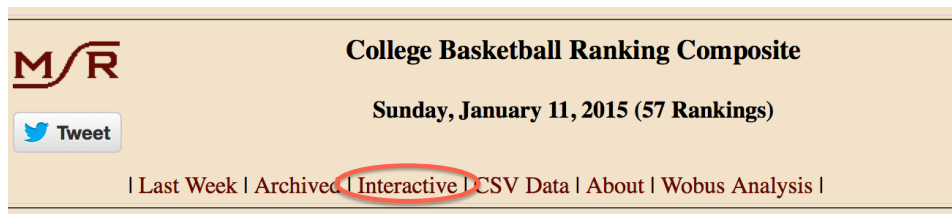


This brings you to a page with current rankings for College Basketball:

<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.

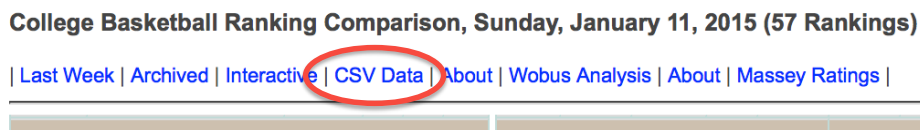
Personally from here I like to go to the interactive page by clicking on `interactive` in the panel above the rankings:



<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.

We can now open the data with Excel by clicking on: `CSV Data` at the top of the page.



We save it as a csv file in a file called `compare.csv` in our working directory.

We clean the data up in Excel first, deleting the glossary of abbreviations at the top. We leave just the abbreviations of the rankings on the top line.

To create an average rank using only your favorite rankings, we first import this file into R using the command :

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

(note this time we set `header` to "TRUE" to convert the top line to column names). Now we would like to take a few of our most respected rankings and calculate the average rank for each team. If we use AP rankings or the Coaches Poll, which rank the top 25, we will have to fill in the blank spaces in those columns with 26's.

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) ?

We can create a new data frame with only these rankings called `myrankings` with the command (Section 4.2 of [1]):

```
myrankings<-rankings[,c("Team", "AP", "USA", "SAG", "MAS", "POM")]
```

Note that the missing ranks in the AP polls and the Coaches poll appear as NA entries. We can identify where these are in the data frame `myrankings` with the command (Section 4.2, Page 159 of [1])

```
is.na(myrankings)
```

As you can see this gives us a logical vector of "TRUE"s and "FALSE"s. We can set the data entries corresponding to "TRUE" (where the NA's appear) to 26 with the command:

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

We can now simply apply the `mean` function to every row with the `apply` function:

```
myrankings$ave<-apply(myrankings[,2:6],1,mean)
```

Finally we can order our data in increasing order of average rankings using the command:

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

(<http://www.statmethods.net/management/sorting.html>)

A Few Notes

- The following website tracks the accuracy of ranking systems throughout the season. [Wobus Sports](#) (under [Ranking Analyses](#))
If you are going to use a Borda Count, this might help you pick good ranking systems to include.
- Unfortunately, the names of the teams in the file above do not match the names in the files we use to compute rankings from the game data. Therefore it is best to reduce the with the rankings created using all of the data to rankings for the 64 teams before merging (by hand).

Downloading Stats from a webpage

There are a number of webpages that have team stats and individual stats for players. This site called TEAMRANKINGS <https://www.teamrankings.com/ncb/stats/>

has a lot of data that one would like, however I'm not sure how accurate it is. Other sources of data are

ESPN <http://www.espn.com/mens-college-basketball/statistics>

(Five Thirty Eight should also have some useful predictions) and

CBS SPORTS <http://www.cbssports.com/collegebasketball/stats>

and

<https://www.sports-reference.com/cbb/seasons/2019-school-stats.html>

Some type of regression might help. One of the big problems is putting together game-by-game statistics with Global Season statistics such as the four factors and average age/height/weight for the teams. Typically the data comes from different websites and the teams have different names. By the end of Spring break, I may be able to create a file so that we can run a regression with the point differential as the dependent variable and the above factors as the independent variables. If I don't

get to make this file (for example if I decide that St. Patrick's day should be St. Patrick's week and celebrate accordingly :)), then we can always use each statistic individually to make a ranking and throw them in with our other selected rankings above.