

## **RegularSeasonDetailedResults**

This file is a more detailed set of game results, covering seasons 2003-2016. This includes team-level total statistics for each game (total field goals attempted, offensive rebounds, etc.) The column names should be self-explanatory to basketball fans (as above, "w" or "l" refers to the winning or losing team):

- wfgm - field goals made
- wfga - field goals attempted
- wfgm3 - three pointers made
- wfga3 - three pointers attempted
- wftm - free throws made
- wfta - free throws attempted
- wor - offensive rebounds
- wdr - defensive rebounds
- wast - assists
- wto - turnovers
- wstl - steals
- wblk - blocks
- wpf - personal fouls

## **TourneyCompactResults**

This file identifies the game-by-game NCAA tournament results for all seasons of historical data. The data is formatted exactly like the regular\_season\_compact\_results.csv data. Note that these games also include the play-in games (which always occurred on day 134/135) for those years that had play-in games.

- "season" - this is the year of the associated entry in seasons.csv (the year in which the final tournament occurs)
- "daynum" - this integer always ranges from 0 to 132, and tells you what day the game was played on. It represents an offset from the "dayzero" date in the "seasons.csv" file. For example, the first game in the file was daynum=20. Combined with the fact from the "season.csv" file that day zero was 10/29/1984, that means the first game was played 20 days later, or 11/18/1984. There are no teams that ever played more than one game on a given date, so you can use this fact if you need a unique key. In order to accomplish this uniqueness, we had to adjust one game's date. In March 2008, the SEC postseason tournament had to reschedule one game (Georgia-Kentucky) to a subsequent day, so Georgia had to actually play two games on the same day. In order to enforce this uniqueness, we moved the game date for the Georgia-Kentucky game back to its original date.
- "wteam" - this identifies the id number of the team that won the game, as listed in the "teams.csv" file. No matter whether the game was won by the home team or visiting team, "wteam" always identifies the winning team.
- "wscore" - this identifies the number of points scored by the winning team.
- "lteam" - this identifies the id number of the team that lost the game.

- "lscore" - this identifies the number of points scored by the losing team.
- "numot" - this indicates the number of overtime periods in the game, an integer 0 or higher.
- "wloc" - this identifies the "location" of the winning team. If the winning team was the home team, this value will be "H". If the winning team was the visiting team, this value will be "A". If it was played on a neutral court, then this value will be "N". Sometimes it is unclear whether the site should be considered neutral, since it is near one team's home court, or even on their court during a tournament, but for this determination we have simply used the Kenneth Massey data in its current state, where the "@" sign is either listed with the winning team, the losing team, or neither team.