

NBA Data Info

A dataset containing NBA teams by year (from 1980 to 2016), their winning percentage, and other relevant shooting, turnover, rebounding, and free throw statistics. There are 1,014 observations for this dataset.

Variables

Team Name (Team)

Year (Year) – 1980 to 2016

Win Percentage (W.L.%) – Wins / Games Played (for a given season)

NBA Conference (Conference) – Eastern (East) and Western (West) Conferences

Did the team make the playoffs? (MadePlayoffs) – “Yes”, the team made the playoffs or “No”, the team did not make the playoffs.

Effective Field Goal % (EFG%) – $(\text{All Field Goals Made} + .5 * (3\text{P Field Goals Made})) / (\text{All Field Goals Attempted})$

Opponent's Effective Field Goal % (Opp_EFG%) – $(\text{All Opponent's Field Goals Made} + .5 * (\text{Opponent's 3P Field Goals Made})) / (\text{All Opponent's Field Goals Attempted})$

Turnovers Committed per Possession (TTP) – Turnovers / Possession. A possession starts when a team gets the basketball and end when they give up control of the basketball. We have to estimate the possession statistic, Wayne Winston calls for $\text{FGs Made} + \text{Turnovers} + (.44 * \text{FTs Attempted})$.

Defensive Turnovers Caused per Possession (DTTP) – Turnovers Caused / Possession

Offensive Rebound Rate (ORebRate) – Offensive Rebounds / Missed Shots. We calculate missed shots by subtracting FGs Made from FGs Attempted ($\text{FGA} - \text{FG} = \text{Missed Shots}$).

Defensive Rebound Rate (DRebRate) – Defensive Rebounds / Opponents Missed Shots.

Free Throw Rate (FTR) – Free Throws Made / Field Goals Attempted. The FTR is impacted by how often a team gets to the foul line as well as by their free throw percentage.

Opponent's Free Throw Rate (OFTR) – Opponents Free Throws Made / Opponents Field Goals Attempted.

Collection Method

Basketball Reference (<http://www.basketball-reference.com>) has a hoard of basketball related data. I used the “rvest” package to scrape from each year of their seasons pages (i.e. http://www.basketball-reference.com/leagues/NBA_2014.html). The pages are almost (but not completely) identical for each

year. This was mostly simple as the pages have a records table, team's statistics table, and opponent's statistics table there were issues that needed help. An example is below:

```
# Eastern Conference Records Table
base = 'http://www.basketball-reference.com/leagues/NBA_'
year = 1980
all_seasons_east_records = {}

while(year < 2017) {
  url = paste0(base, toString(year), '.html')
  team_table = '//*[@id="E_standings"]'

  nba = url %>%
    html() %>%
    html_nodes(xpath=team_table) %>%
    html_table()
  nba = nba[[1]]

  nba$year = year

  all_seasons_east_records = bind_rows(nba,all_seasons_east_records)
  year = year + 1 }
```

I still needed to clean up certain rows (i.e. Subtotal rows or division labels) and certain columns would have extra data (i.e. Numeric order of finish in parentheses after the team name). I used the code below to help clean up the data.

```
# Eastern Conference
all_seasons_east_records$`Eastern Conference` =
  str_replace_all(all_seasons_east_records$`Eastern Conference`,
    "[(.),1,2,3,4,5,6,7,8,9,0,Â]", "")

all_seasons_east_records$`Eastern Conference` =
  str_replace_all(all_seasons_east_records$`Eastern Conference`,
    "Philadelphia ers", "Philadelphia 76ers")

all_seasons_east_records$`Eastern Conference` =
  str_trim(all_seasons_east_records$`Eastern Conference`,
    side = c("both", "left", "right"))

all_seasons_east_records = filter(all_seasons_east_records,
  `Eastern Conference` != "Central Division",
  `Eastern Conference` != "Atlantic Division",
  `Eastern Conference` != "Southeast Division")

all_seasons_east_records$Conference = "East"
all_seasons_east_records = rename(all_seasons_east_records, Team = `Eastern Conference`)
```

And then finally, I needed to create the fields that Wayne Winston mentioned in his book (Variables made up from box score statistics). A few examples are below.

Effective FG%

$$\text{nba_data\$EFG} = (\text{nba_data\$FG} + (0.5 * \text{nba_data\$`3P`})) / \text{nba_data\$FGA}$$

Opponents Eff FG%

$$\text{nba_data\$Opp_EFG} = (\text{nba_data\$Opp_FG} + (0.5 * \text{nba_data\$Opp_3P})) / \text{nba_data\$Opp_FGA}$$

Turnovers Committed / Possession

$$\text{nba_data\$TTP} = \text{nba_data\$TOV} / (\text{nba_data\$FG} + \text{nba_data\$TOV} + (\text{nba_data\$FTA} * .44))$$

Turnovers Caused / Possession

$$\text{nba_data\$DTTP} = \text{nba_data\$Opp_TOV} / (\text{nba_data\$Opp_FG} + \text{nba_data\$Opp_TOV} + (\text{nba_data\$Opp_FTA} * .44))$$