The Problem

I could not find a website that had displayed Basketball stats for each team in the format that I wanted.

The Goal

I wanted the stats to be displayed as totals from the regular season of 82 games plus the totals from the play in and playoffs that way I would have the correct data from each teams season and not segmented parts

Ex. 1

What exists

Regular season	Wins	Loses	Games Played	points
Hawks #1	50	32	82	8000
Knicks #10	38	44	82	7250
Nets #9	38	44	82	7500
Lakers #12	20	62	82	6000

Play-in	Wins	Loses	Games Played	points
Hawks #1	Not required	Not required	Not required	Not required
Knicks #10	2	0	2	200
Nets #9	0	1	1	90
Lakers #12	Didn’t qualify	Didn’t qualify	Didn’t qualify	Didn’t qualify

Play-offs	Wins	Loses	Games Played	points
Hawks #1	4	4	8	800

Knicks #10	16	8	24	2000
Nets #9	Didn't qualify	Didn't qualify	Didn't qualify	Didn't qualify
Lakers #12	Didn't qualify	Didn't qualify	Didn't qualify	Didn't qualify

What I want

Total season	Wins	Loses	Games Played	points
Hawks	54	36	90	8800
Knicks	54	52	108	9450
Nets	38	45	83	7590
Lakers	20	62	82	6000

Procedure (before R)

1. find the data
    a. I went to this link (https://www.nba.com/stats/teams/traditional?Season=2023-24&SeasonType=Regular+Season&PerMode=Totals)
        i. This was the regular season data and I still needed the play-in and playoff data
    b. Went to https://www.nba.com/stats/teams/traditional?Season=2023-24&SeasonType=PlayoffsPerMode=Totals for playoffs and https://www.nba.com/stats/teams/traditional?Season=2023-24&SeasonType=PlayIn&PerMode=Totals for play-in
2. Take the data
    a. I copied all the data from the 3 links into a google sheet titled NBA2023(https://docs.google.com/spreadsheets/d/1uqFrLvIKnfme_jXggIqROcs8xWsZQ-D3uBiIxQjCF7c/edit?gid=0#gid=0)
        i. The 2023 is because the data from that season started in 2023
3. Adding the data into 30 rows for each team
    a. I realized how slow this was going to be if I was gonna have to manually add all the data inside the google sheet and delete all the rows that I went through so I resorted to using R code.
        i. To prepare the data before going into R I sorted the data frame by alphabetical order in the team name column

        1. First freeze row 1 and then sort
    ii. I also cleared any column with percentages because the math would not of worked the same way concerning addition
    iii. I also renamed some columns that way It would be easier to work with in R
        1. 3:00pm -> 3p
        2. 3p% -> 3%
    iv. I then created my CSV and put it in a folder

Procedure (in R)

1. Download the csv file into R
```
nba2023 <- read.csv(file.choose(),stringsAsFactors = TRUE)
```
2. Look at the file make sure everything is good + its good to see where you started from
```
View(nba2023)
```
3. Add values from the rows the where team is the same
    a. To do this I created a function
```
process_teams <- function(data) {
  library(dplyr)
  summed_data <- data %>%
    group_by(TEAM) %>%
    summarise(across(where(is.numeric), sum, na.rm = TRUE)) %>%
    mutate(row_type = "SUM")
  data <- data %>%
    mutate(row_type = "ORIGINAL")
  result <- bind_rows(data, summed_data)
  return(result)
}
```
    b. Then you plug in the data frame you want the function to be applied to
```
data <- nba2023
nba2023 <- process_teams(data)
```
    c. Look at the new data frame to make sure everything looks good
```
View(nba2023)
```
4. Now we want to extract the rows with the summed values because those are the totals from the whole season
```
sorted.nba2023 <- nba2023[order(nba2023$row_type), ]
totalnba2023 <- sorted.nba2023[55:84, 1:28 ]
```
    a. To make sure got all 30 teams I used the levels function to check
```
levels(totalnba2023$TEAM)
```
    b. My output looked like which is good because it displays all 30 nba teams
```
     [1] "Atlanta Hawks"          "Boston Celtics"
"Brooklyn Nets"          "Charlotte Hornets"      "Chicago
Bulls"
```

```
      [6] "Cleveland Cavaliers"    "Dallas Mavericks"
"Denver Nuggets"         "Detroit Pistons"        "Golden State
Warriors"
     [11] "Houston Rockets"        "Indiana Pacers"         "LA
Clippers"            "Los Angeles Lakers"     "Memphis
Grizzlies"
     [16] "Miami Heat"             "Milwaukee Bucks"
"Minnesota Timberwolves" "New Orleans Pelicans"   "New York
Knicks"
     [21] "Oklahoma City Thunder"  "Orlando Magic"
"Philadelphia 76ers"     "Phoenix Suns"           "Portland
Trail Blazers"
[26] "Sacramento Kings"       "San Antonio Spurs"      "Toronto
Raptors"        "Utah Jazz"              "Washington Wizards"
```

c. Then to check if everything else is good

```
View(totalnba2023)
```

5. Now that I have the total of that whole season want to get the percentages back
    a. These percentages will go into pre existing empty columns

```
totalnba2023$win. <- (totalnba2023$W/totalnba2023$GP)*100
totalnba2023$FG. <- (totalnba2023$FGM/totalnba2023$FGA)*100
totalnba2023$X3. <- (totalnba2023$X3p/totalnba2023$X3PA)*100
totalnba2023$FT. <- (totalnba2023$FTM/totalnba2023$FTA)*100
```

    b. To check if everything is still good

```
View(totalnba2023)
```

6. Making a new data frame with the total values divided by the games played for that season

```
NBAstats23.GP <- data.frame(totalnba2023$TEAM, totalnba2023$GP,
totalnba2023$W, totalnba2023$L, totalnba2023$win.,
totalnba2023$MIN, totalnba2023$PTS/totalnba2023$GP,
totalnba2023$FGM/totalnba2023$GP,totalnba2023$FGA/totalnba2023$
GP, totalnba2023$FG. ,totalnba2023$X3p/totalnba2023$GP,
totalnba2023$X3PA/totalnba2023$GP, totalnba2023$X3.,
totalnba2023$FTM/totalnba2023$GP,
totalnba2023$FTA/totalnba2023$GP, totalnba2023$FT.,
totalnba2023$OREB/totalnba2023$GP,
totalnba2023$DREB/totalnba2023$GP,
totalnba2023$REB/totalnba2023$GP,
totalnba2023$AST/totalnba2023$GP,
totalnba2023$TOV/totalnba2023$GP,
totalnba2023$STL/totalnba2023$GP,
totalnba2023$BLK/totalnba2023$GP,
totalnba2023$BLKA/totalnba2023$GP,
totalnba2023$PF/totalnba2023$GP,
totalnba2023$PFD/totalnba2023$GP,
totalnba2023$X.ERROR./totalnba2023$GP)
```

    a. To make sure everything still looks good

```
View(NBAstats23.GP)
```

7. To add some more values that could be useful into the data frame. This adds 5 new columns at the end

```
NBAstats23.GP$new_column1 <-
(totalnba2023$OREB/totalnba2023$REB)*100
NBAstats23.GP$new_column2 <-
(totalnba2023$DREB/totalnba2023$REB)*100
NBAstats23.GP$new_column3 <-
(totalnba2023$AST-totalnba2023$TOV)/totalnba2023$GP
NBAstats23.GP$new_column4 <-
(totalnba2023$BLK-totalnba2023$BLKA)/totalnba2023$GP
NBAstats23.GP$new_column5 <-
(totalnba2023$PF-totalnba2023$PFD)/totalnba2023$GP
```

8. To make the data frame look better I want to change the names of the columns into something smaller and give unnamed columns names

```
colnames(NBAstats23.GP)[1:32] <- c("Team","GP","W","L","win%",
                                   "MIN",
                                   "PTS",
                                   "FGM","FGA","FG%",
                                   "x3","x3A","x3%",
                                   "FTM","FTA","FT%",
                                   "OREB","DREB", "REB",
                                   "AST","TOV",
                                   "STL",
                                   "BLK","BLKA",
                                   "PF","PFD",
                                   "plus.minus",
                                  "OREB%", "DREB%", "AST-TOV",
      "BLK-BLKA", "PF-PFD")
```

    a. Lets see the final data frame

```
View(NBAstats23.GP)
```

How to do it again

I want more than one season. The goal is from 1996 (the first season the nba.com has taken data for) to 2023

1. I copied and pasted steps 1-8 from Procedure (in R) into a google doc
    a. https://docs.google.com/document/d/1x8l62tMqma8FMBJZ1Z7pHYLXfYwMDxEw5i43eX9LF5o/edit?tab=t.0
2. Command f to replace “23” (the season I took my data for the first go through) with “22” (the next season)
    a. This will go on until “00”

    b. Then ill have to replace “2000” with “1999”
    c. Then i can go back to replacing “99” with “98” and so on until 1996
    d. Apparently R can do this as well so going forward that is what will be used
3. Repeat all steps Procedure (before R) to this point

Changes

2020: less games played due to covid
2019: only 2 play-in teams/ less games played due to covid
2018: no play-in
2013: Charlotte Hornets <- Charlotte Bobcats
2012: New Orleans Pelicans <- New Orleans Hornets
2012: Brooklyn Nets <- New Jersey Nets
2007: Oklahoma City Thunder <- Seattle SuperSonics
2004-1999: for some reason the NBA did not feel like tracking personal fouls drawn
2003: - Charlotte Bobcats
2001: best of 7 to best of 5 playoffs => less games
2000: Memphis Grizzlies <- Vancouver Grizzlies
1998: 50 game regular season

ReSources

Google docs
Google Sheets
R
Github
https://www.nba.com/stats/teams/traditional?Season=2023-24&SeasonType=Regular+Season&PerMode=Totals
https://www.nba.com/stats/teams/traditional?Season=2023-24&SeasonType=PlayoffsPerMode=Totals
https://www.nba.com/stats/teams/traditional?Season=2023-24&SeasonType=PlayIn&PerMode=Totals