Predicting NBA Win Percentage by First X Games

@jonmichaelstroh

10/20/2021

Get Overall Records

Edit Standings for Merging

```
for (x in 1:nrow(standings)){
  standings$slugTeam[x] <- paste('ss_', standings$slugTeam[x], "_", standings$slugSeason[x], sep = '')
}
```

Single season game records

```
season_schedule <- nbastatR::seasons_schedule(seasons = 2010:2021, season_types = "Regular Season", box_
```

```
## Acquiring NBA basic team game logs for the 2009-10 Regular Season
## Acquiring NBA basic team game logs for the 2010-11 Regular Season
## Acquiring NBA basic team game logs for the 2011-12 Regular Season
## Acquiring NBA basic team game logs for the 2012-13 Regular Season
## Acquiring NBA basic team game logs for the 2013-14 Regular Season
## Acquiring NBA basic team game logs for the 2014-15 Regular Season
## Acquiring NBA basic team game logs for the 2015-16 Regular Season
## Acquiring NBA basic team game logs for the 2016-17 Regular Season
## Acquiring NBA basic team game logs for the 2017-18 Regular Season
## Acquiring NBA basic team game logs for the 2018-19 Regular Season
## Acquiring NBA basic team game logs for the 2019-20 Regular Season
## Acquiring NBA basic team game logs for the 2020-21 Regular Season

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'
```

```
## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'
```

```
## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'
```

```
## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'

## Warning: All elements of '...' must be named.
## Did you want 'dataTables = c(typeSeason, dateGame, idGame, numberGameTeamSeason, nameTeam,
##     idTeam, isB2B, isB2BFirst, isB2BSecond, locationGame, slugMatchup,
##     slugTeam, countDaysRestTeam, countDaysNextGameTeam, slugOpponent,
##     slugTeamWinner, slugTeamLoser, outcomeGame, isWin, fgmTeam,
##     fgaTeam, pctFGTeam, fg3mTeam, fg3aTeam, pctFG3Team, pctFTTeam,
##     hasVideo, fg2mTeam, fg2aTeam, pctFG2Team, minutesTeam, ftmTeam,
##     ftaTeam, orebTeam, drebTeam, trebTeam, astTeam, stlTeam,
##     blkTeam, tovTeam, pfTeam, ptsTeam, plusminusTeam, urlTeamSeasonLogo)'?

## Warning: 'cols' is now required when using unnest().
## Please use 'cols = c(dataTables)'
```

```
season_schedule <- season_schedule[c('slugSeason', 'dateGame', 'slugMatchup', 'slugTeamWinner', 'slugTea
```

So we have the ability to, from 2010 to 2021, judge how strong of a predictor the first N number of NBA games is on the rest of a team's season:

N game records

N = 1

How about we start with 1 game:

```r
years <- c()
teams <- c()
ngameperc <- c()
n <- 1
for (df_name in names(teamDF1)){
  years <- append(years, teamDF1[[df_name]]$slugSeason[1])
  teams <- append(teams, df_name)
  ngameperc <- append(ngameperc, teamDF1[[df_name]]$runningWins[n]/n)
}

ngamepercdf <- data.frame(years, teams, ngameperc)
```

dataframe 2

```r
years <- c()
teams <- c()
ngameperc <- c()
for (df_name in names(teamDF2)){
  years <- append(years, teamDF2[[df_name]]$slugSeason[1])
  teams <- append(teams, df_name)
  ngameperc <- append(ngameperc, teamDF2[[df_name]]$runningWins[n]/n)
}

ngamepercdf2 <- data.frame(years, teams, ngameperc)
```

dataframe 3

```r
years <- c()
teams <- c()
ngameperc <- c()
for (df_name in names(teamDF3)){
  years <- append(years, teamDF3[[df_name]]$slugSeason[1])
  teams <- append(teams, df_name)
  ngameperc <- append(ngameperc, teamDF3[[df_name]]$runningWins[n]/n)
}

ngamepercdf3 <- data.frame(years, teams, ngameperc)
```

merge both dfs

```r
ngamepercdf  <- rbind(ngamepercdf, ngamepercdf2)
ngamepercdf <- rbind(ngamepercdf, ngamepercdf3)
```

```r
combinedDF <- merge(standings, ngamepercdf, by.x="slugTeam", by.y="teams")
```
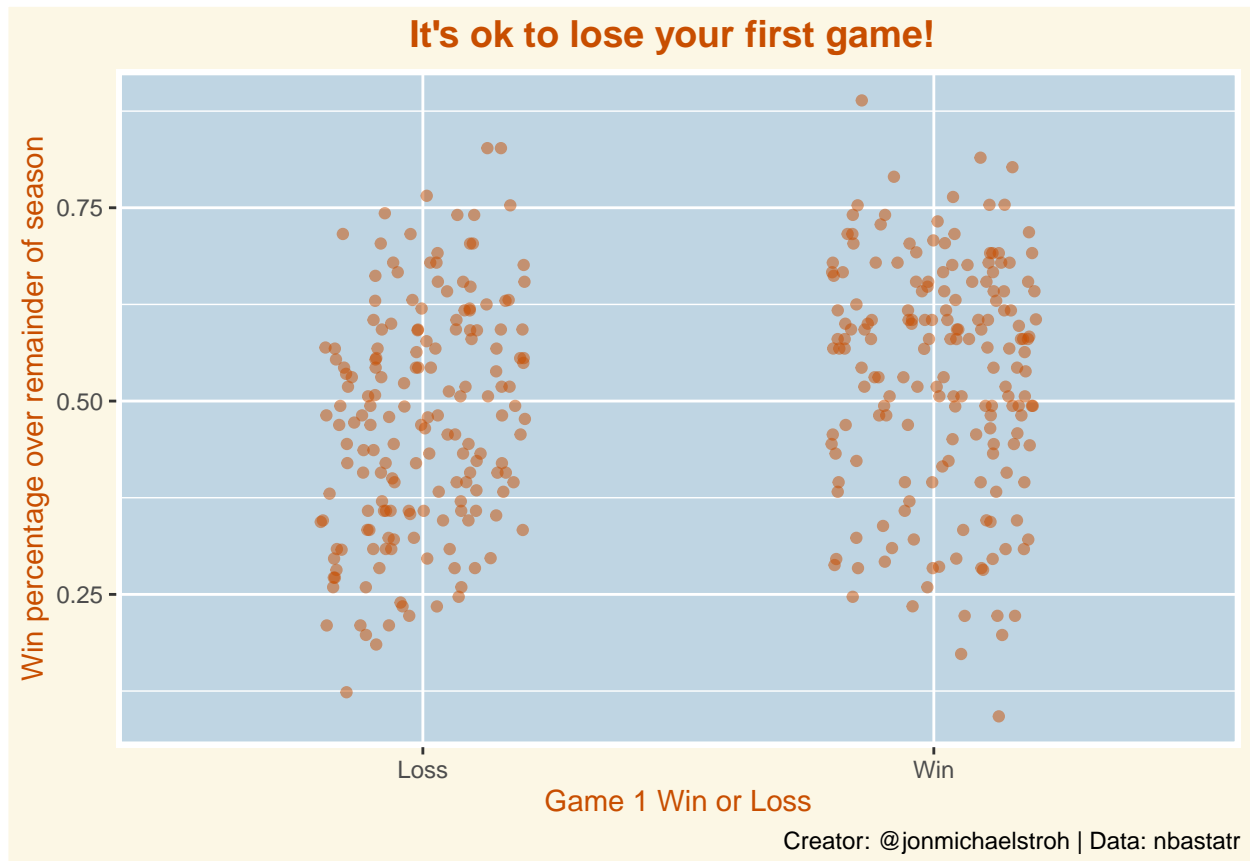
#Get percent after n games

```
combinedDF <- within(combinedDF, recordOverall<-data.frame(do.call('rbind', strsplit(as.character(record
```

```
combinedDF <- combinedDF %>% mutate(percentAfter = (strtoi(recordOverall$X1) - (n * ngameperc))/((strto
```

```
combinedDF <- combinedDF %>% mutate(win = if_else(ngameperc == 1, 'Win', 'Loss'))
```

```
ggplot(combinedDF, aes(x = win, y = percentAfter)) + geom_jitter(alpha=0.5, color = '#C84E00',position
size = 2, linetype = "solid"), panel.grid.major = element_line(size = 0.5, linetype = 'solid', colour =
```
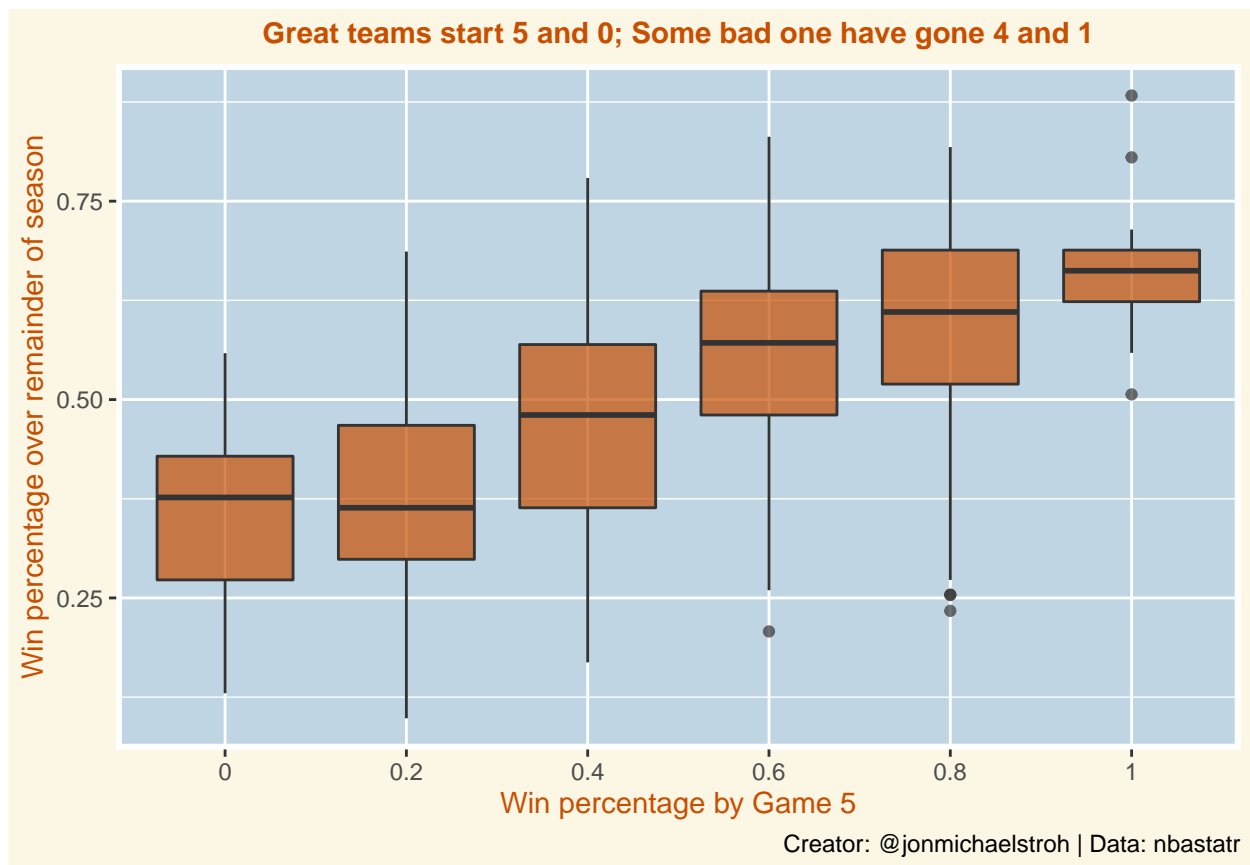

N = 5

What about after 5 games?

dataframe 2

dataframe 3

merge both dfs

#Get percent after n games

```
ggplot(combinedDF, aes(x = as.factor(round(ngameperc, 2)), y = percentAfter)) + geom_boxplot(fill = '#C8
size = 2, linetype = "solid"), panel.grid.major = element_line(size = 0.5, linetype = 'solid', colour =
```

Great teams start 5 and 0; Some bad one have gone 4 and 1

Creator: @jonmichaelstroh | Data: nbastatr

N = 15

And finally, 15 games...

```
years <- c()
teams <- c()
ngameperc <- c()
n <- 15
for (df_name in names(teamDF1)){
  years <- append(years, teamDF1[[df_name]]$slugSeason[1])
  teams <- append(teams, df_name)
  ngameperc <- append(ngameperc, teamDF1[[df_name]]$runningWins[n]/n)
}

ngamepercdf <- data.frame(years, teams, ngameperc)
```

dataframe 2

```r
years <- c()
teams <- c()
ngameperc <- c()
for (df_name in names(teamDF2)){
  years <- append(years, teamDF2[[df_name]]$slugSeason[1])
  teams <- append(teams, df_name)
  ngameperc <- append(ngameperc, teamDF2[[df_name]]$runningWins[n]/n)
}

ngamepercdf2 <- data.frame(years, teams, ngameperc)
```

dataframe 3

```r
years <- c()
teams <- c()
ngameperc <- c()
for (df_name in names(teamDF3)){
  years <- append(years, teamDF3[[df_name]]$slugSeason[1])
  teams <- append(teams, df_name)
  ngameperc <- append(ngameperc, teamDF3[[df_name]]$runningWins[n]/n)
}

ngamepercdf3 <- data.frame(years, teams, ngameperc)
```

merge both dfs

```r
ngamepercdf  <- rbind(ngamepercdf, ngamepercdf2)
ngamepercdf <- rbind(ngamepercdf, ngamepercdf3)
```

```r
combinedDF <- merge(standings, ngamepercdf, by.x="slugTeam", by.y="teams")
```
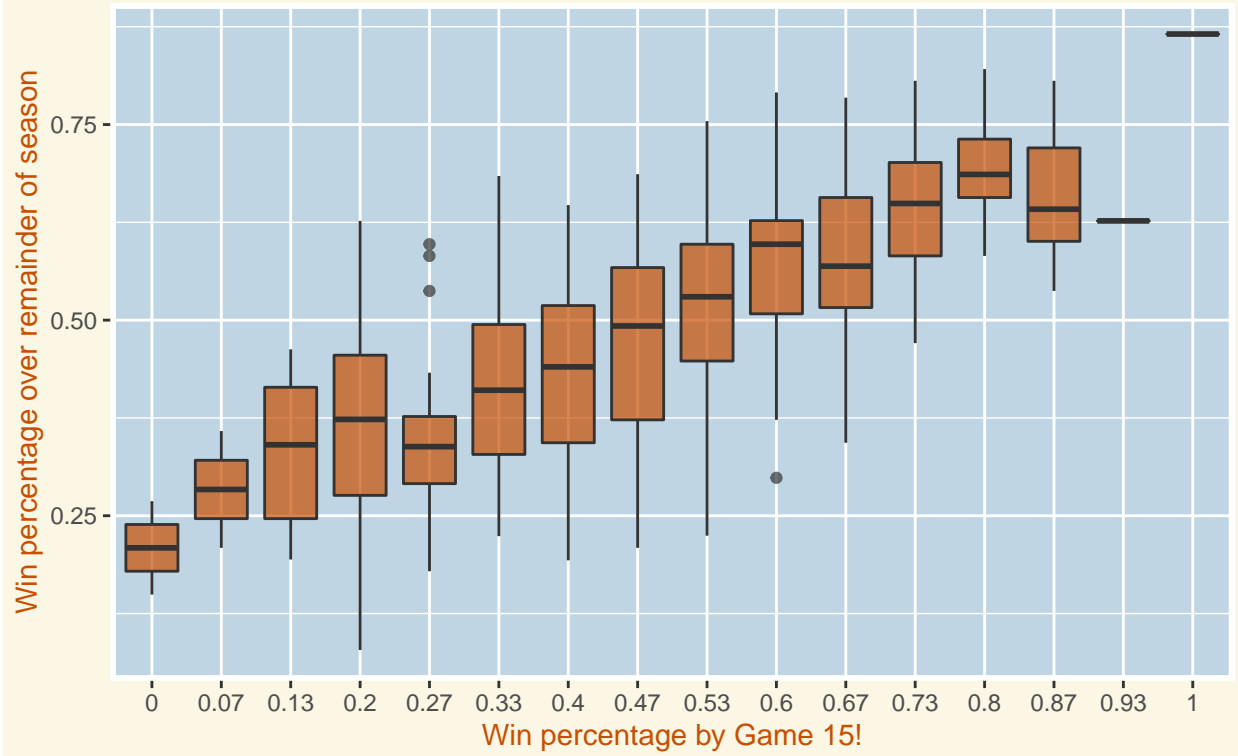
#Get percent after n games

```r
combinedDF <- within(combinedDF, recordOverall<-data.frame(do.call('rbind', strsplit(as.character(recor
```

```r
combinedDF <- combinedDF %>% mutate(percentAfter = (strtoi(recordOverall$X1) - (n * ngameperc))/((strto
```

```r
ggplot(combinedDF, aes(x = as.factor(round(ngameperc, 2)), y = percentAfter)) + geom_boxplot(fill = '#C8
size = 2, linetype = "solid"), panel.grid.major = element_line(size = 0.5, linetype = 'solid', colour =
```

Things seem pretty set by Game 15

Win percentage over remainder of season

Win percentage by Game 15!

Creator: @jonmichaelstroh | Data: nbastatr