lab_3.R

salmaelshahawy

2019-09-12

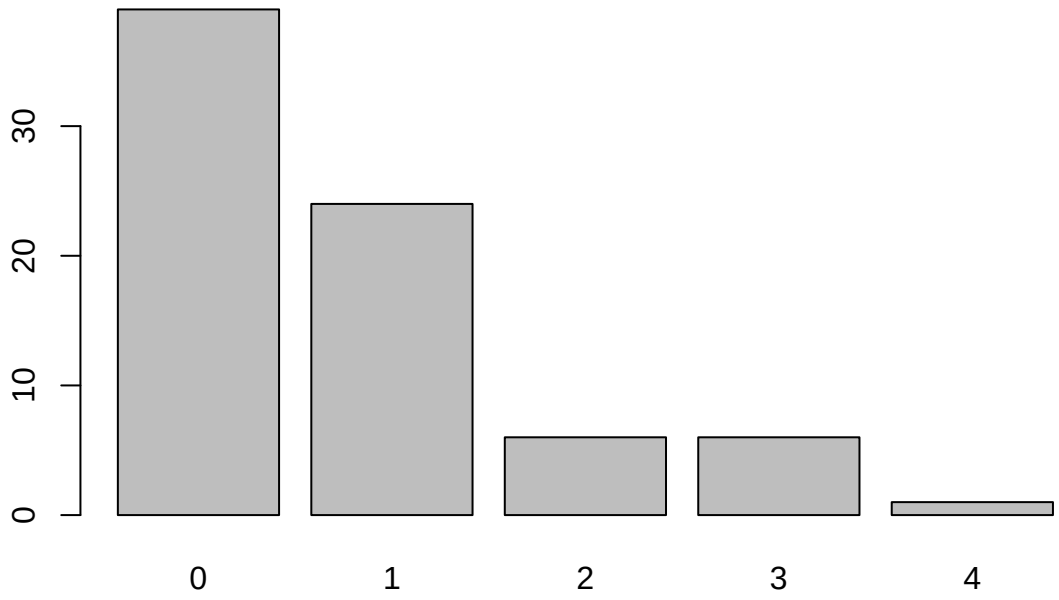
```
load("kobe.rdata")
head(kobe)
```

```
##    vs game quarter time
## 1 ORL    1       1 9:47
## 2 ORL    1       1 9:07
## 3 ORL    1       1 8:11
## 4 ORL    1       1 7:41
## 5 ORL    1       1 7:03
## 6 ORL    1       1 6:01
##                                           description basket
## 1                 Kobe Bryant makes 4-foot two point shot      H
## 2                             Kobe Bryant misses jumper      M
## 3                        Kobe Bryant misses 7-foot jumper      M
## 4 Kobe Bryant makes 16-foot jumper (Derek Fisher assists)      H
## 5                         Kobe Bryant makes driving layup      H
## 6                             Kobe Bryant misses jumper      M
```

```
kobe$basket[1:9]
```

```
## [1] "H" "M" "M" "H" "H" "M" "M" "M" "M"
```

```
# What does a streak length of 1 mean, i.e. how many hits and misses are in a streak of 1? What about a
# 1: means one hit followed by missed hit. one hit - one miss. streak 0 means no hits, only missed hits

kobe_streak <- calc_streak(kobe$basket)
barplot(table(kobe_streak))
```




```
# Describe the distribution of Kobe's streak lengths from the 2009 NBA finals.
  # It was a right-skewed distribution
# What was his typical streak length? How long was his longest streak of baskets?
  # The typical streak length was 0. Streaks of 1 were second with 24, then 6 streaks of 2 and 3, and o

outcomes <- c("heads", "tails")
sample(outcomes, size = 1, replace = TRUE)
```

```
## [1] "tails"
```

```
sim_fair_coin <- sample(outcomes, size = 100, replace = TRUE)
sim_fair_coin
```

```
##   [1] "heads" "heads" "heads" "heads" "tails" "tails" "tails" "heads"
##   [9] "tails" "heads" "tails" "heads" "heads" "tails" "tails" "heads"
##  [17] "heads" "heads" "tails" "heads" "heads" "tails" "heads" "tails"
##  [25] "heads" "heads" "tails" "heads" "heads" "tails" "tails" "heads"
##  [33] "heads" "heads" "heads" "heads" "tails" "tails" "tails" "tails"
##  [41] "heads" "heads" "heads" "tails" "tails" "heads" "heads" "tails"
##  [49] "heads" "tails" "heads" "tails" "heads" "tails" "heads" "heads"
##  [57] "tails" "heads" "tails" "heads" "heads" "heads" "tails" "tails"
##  [65] "heads" "heads" "tails" "tails" "heads" "heads" "heads" "tails"
##  [73] "heads" "heads" "heads" "heads" "tails" "tails" "tails" "tails"
##  [81] "tails" "heads" "heads" "tails" "heads" "tails" "tails" "heads"
##  [89] "tails" "tails" "heads" "tails" "heads" "heads" "tails" "tails"
```

```
##  [97] "heads" "tails" "heads" "tails"
```

```
table(sim_fair_coin)
```

```
## sim_fair_coin
## heads tails
##    55    45
```

```
sim_unfair_coin <- sample(outcomes, size = 100, replace = TRUE, prob = c(0.2, 0.8))
sim_unfair_coin
```

```
##   [1] "tails" "heads" "tails" "heads" "tails" "tails" "tails" "tails"
##   [9] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"
##  [17] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"
##  [25] "tails" "tails" "heads" "heads" "heads" "tails" "tails" "tails"
##  [33] "tails" "tails" "tails" "heads" "tails" "tails" "heads" "tails"
##  [41] "tails" "tails" "heads" "tails" "tails" "tails" "tails" "tails"
##  [49] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"
##  [57] "tails" "tails" "tails" "tails" "heads" "heads" "heads" "tails"
##  [65] "tails" "tails" "heads" "tails" "tails" "heads" "tails" "tails"
##  [73] "tails" "tails" "tails" "tails" "tails" "tails" "heads" "heads"
##  [81] "tails" "heads" "tails" "tails" "tails" "tails" "tails" "tails"
##  [89] "heads" "heads" "heads" "tails" "tails" "tails" "tails" "tails"
##  [97] "tails" "tails" "tails" "heads"
```

```
table(sim_unfair_coin)
```

```
## sim_unfair_coin
## heads tails
##    20    80
```

```
# In your simulation of flipping the unfair coin 100 times, how many flips came up heads?
  # 17 times

?sample

outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 1, replace = TRUE)
table(sim_basket)
```

```
## sim_basket
## H
## 1
```

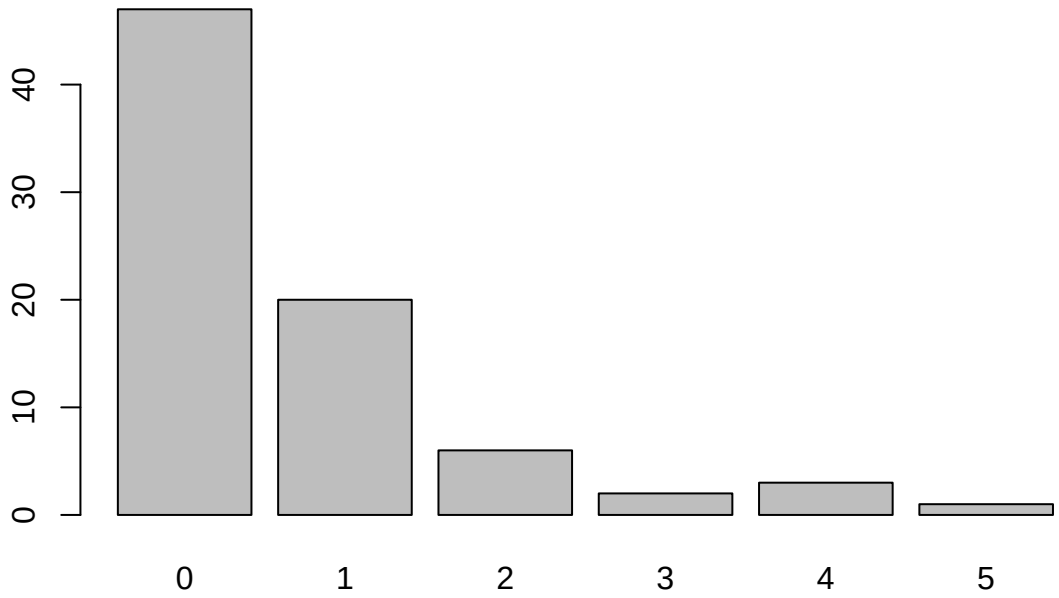
```
# What change needs to be made to the sample function so that it reflects a shooting percentage of 45%?
  # add prob = c(0.45, 0.55) argument to the sample function

# Make this adjustment, then run a simulation to sample 133 shots.
# Assign the output of this simulation to a new object called sim_basket.

outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45, 0.55))
table(sim_basket)
```

```
## sim_basket
##  H  M
## 55 78
```

```
sim_basket
```

```
##   [1] "M" "M" "H" "H" "M" "M" "M" "M" "M" "M" "H" "M" "M" "M" "M" "M" "M"
##  [18] "M" "M" "M" "H" "H" "H" "M" "M" "H" "M" "M" "M" "M" "M" "H" "M" "M"
##  [35] "M" "H" "M" "M" "H" "H" "M" "H" "H" "M" "M" "H" "M" "H" "M" "H" "H"
##  [52] "M" "H" "M" "M" "M" "H" "M" "H" "M" "H" "H" "H" "H" "H" "M" "H" "M"
##  [69] "M" "H" "H" "H" "H" "M" "H" "M" "M" "M" "M" "H" "H" "H" "H" "M" "H" "H"
##  [86] "M" "M" "H" "H" "H" "H" "M" "M" "M" "M" "M" "H" "H" "M" "M" "H" "H"
## [103] "H" "M" "H" "M" "H" "M" "M" "H" "M" "H" "M" "M" "M" "H" "M" "H" "M"
## [120] "M" "H" "M" "M" "H" "M" "M" "M" "M" "H" "M" "M" "M" "M"
```

```
# On your own

ind_streak <- calc_streak(sim_basket)
barplot(table(ind_streak))
```




```
# 1. Describe the distribution of streak lengths.
  # Right-skewed distribution
# What is the typical streak length for this simulated independent shooter with a 45% shooting percenta
  # 0
# How long is the player's longest streak of baskets in 133 shots?
```

```
  # 2

# If you were to run the simulation of the independent shooter a second time,
# how would you expect its streak distribution to compare to the distribution from the question above?
  # I expect to be the same as always the percentage is 45% to 55% for hit and miss respectively.

# How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths f
  # The streak length for the simulated shooter is longer than Kobe Bryant 6 - 4 baised to the simulato
  # The distributions look very similar. Therefore, there doesn't appear to be evidence for Kobe Bryant

table(kobe_streak)
```

```
## kobe_streak
##  0  1  2  3  4
## 39 24  6  6  1
```

```
table(ind_streak)
```

```
## ind_streak
##  0  1  2  3  4  5
## 47 20  6  2  3  1
```

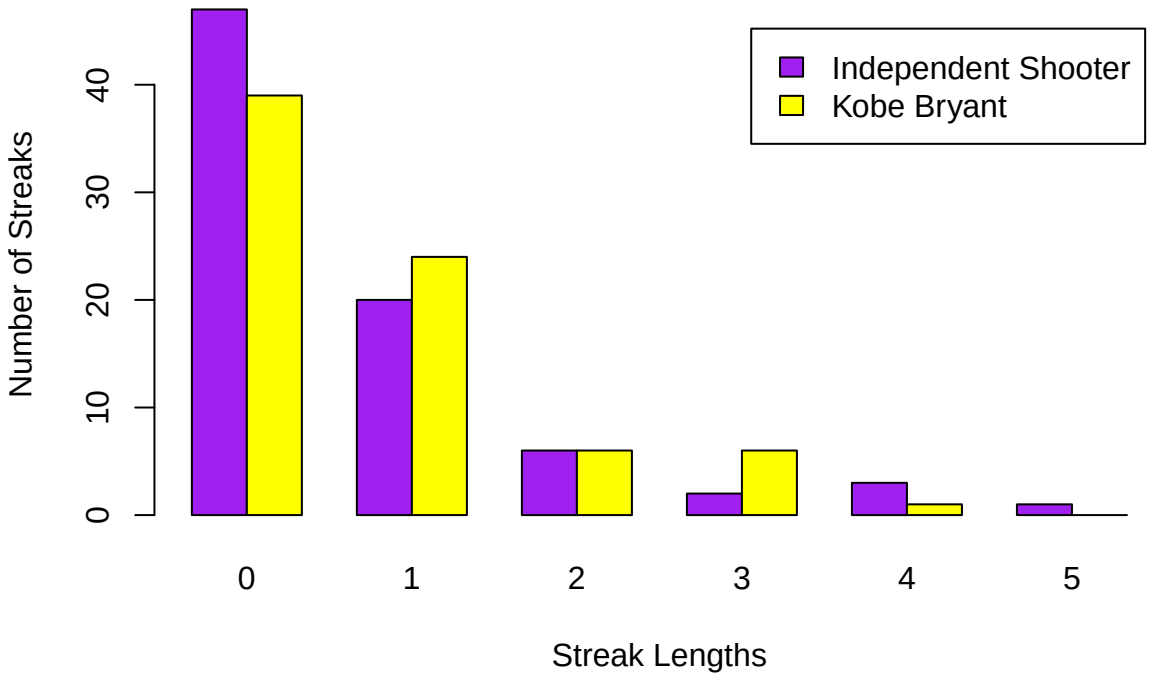
```
# the two tables are not having the same length, so I implemented a logic to put a zero in the table th
# doesn't have a corresponding value.
kobe_df <- unlist(table(kobe_streak))
ind_df <- unlist(table(ind_streak))
if (length(kobe_df) < length(ind_df)) {
  kobe_df <- c(kobe_df,rep(0,length(ind_df) - length(kobe_df)))
} else {
  ind_df <- c(ind_df,rep(0,length(kobe_df) - length(ind_df)))
}

kobe_df
```

```
##  0  1  2  3  4
## 39 24  6  6  1  0
```

```
data_df <- rbind(ind_df,kobe_df)
data_df
```

```
##          0  1 2 3 4 5
## ind_df  47 20 6 2 3 1
## kobe_df 39 24 6 6 1 0
```

```
barplot(data_df ,beside = T,col=c("purple","yellow"),xlab="Streak Lengths",ylab = "Number of Streaks",l
```




```
# to compare the two models, we have to do statistical summary to have a clear comparasion.

# get the median

IQR(kobe_df)
```

```
## [1] 17.25
```

```
IQR(ind_df)
```

```
## [1] 14.25
```

```
mean(kobe_df)
```

```
## [1] 12.66667
```

```
mean(ind_df)
```

```
## [1] 13.16667
```

```
# from analysis above we conclude that there is no major difference between kobe model and independent
# both distributions are close together.
```