

Exercise 1

Exercise 2

Exercise 3

Exercise 4

Exercise 5

Exercise 6

Exercise 7

Exercise 8

Lab 3

Code ▼

Julian Adames-Ng

2022-02-20

Hide

```
library(tidyverse)
library(openintro)
```

Exercise 1

What does a streak length of 1 mean, i.e. how many hits and misses are in a streak of 1? What about a streak length of 0?

-A streak of length 1 means that you made the very first shot, but missed the second so the streak count is reset. A streak of length zero means you missed the first shot on a given run. The streak count resets after each miss.

Hide

```
glimpse(kobe_basket)
```

```
## Rows: 133
## Columns: 6
## $ vs      <fct> ORL, ORL, ORL, ORL, ORL, ORL, ORL, ORL, ORL, ORL, ORL, ORL...
## $ game    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ quarter <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3...
## $ time     <fct> 9:47, 9:07, 8:11, 7:41, 7:03, 6:01, 4:07, 0:52, 0:00, 6:35...
## $ description <fct> Kobe Bryant makes 4-foot two point shot, Kobe Bryant misse...
## $ shot     <chr> "H", "M", "M", "H", "H", "M", "M", "M", "M", "H", "H", "H"...
```

Exercise 2

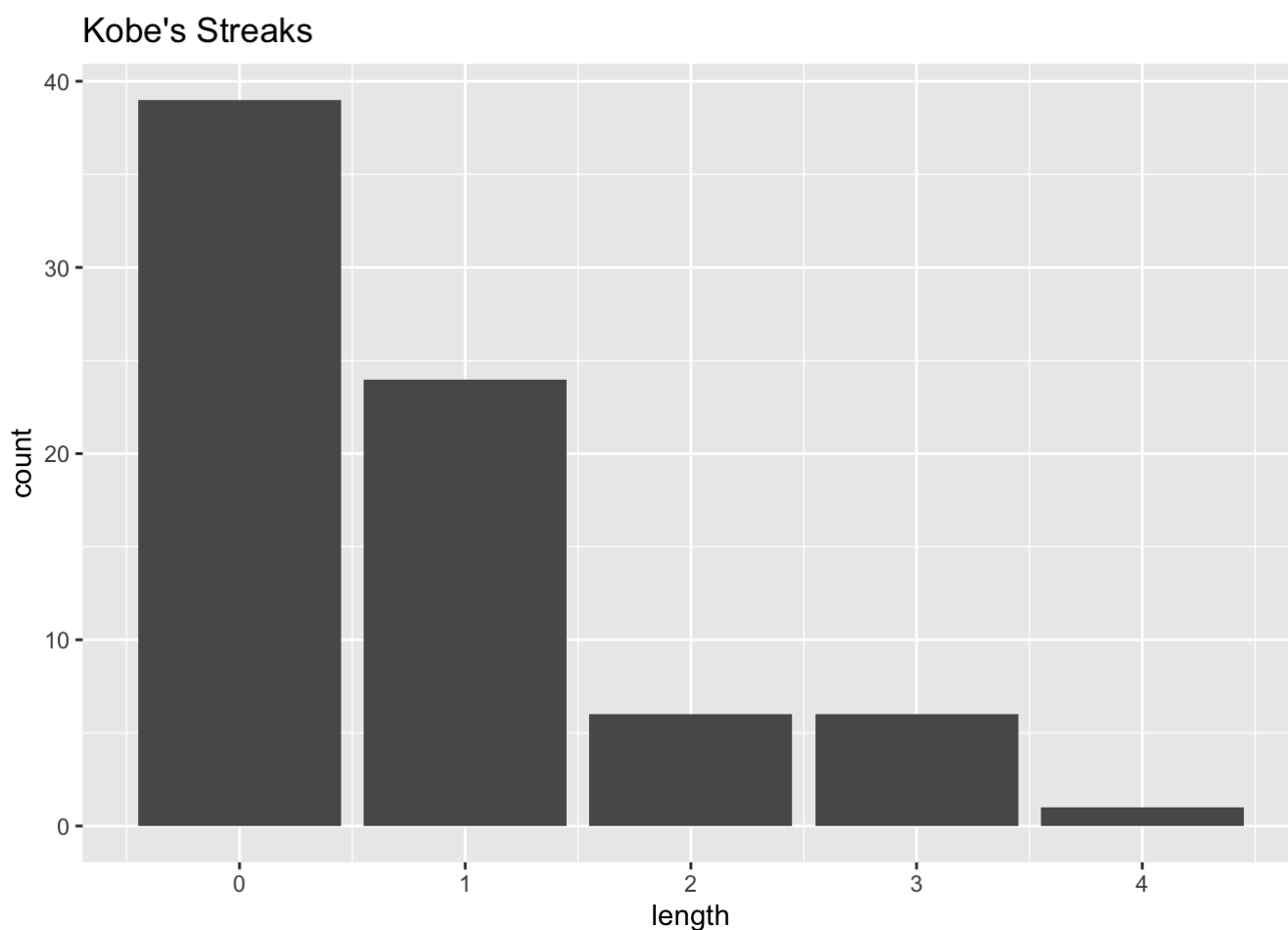
Describe the distribution of Kobe's streak lengths from the 2009 NBA finals. What was his typical streak length? How long was his longest streak of baskets? Make sure to include the accompanying plot in your answer.

-The distribution is skewed right. Kobe's typical streak length was 0 since that is the mode of the data set. His longest streak length was 4.

Hide

```
kobe_streak <- calc_streak(kobe_basket$shot)

ggplot(data = kobe_streak, aes(x = length)) + geom_bar() + ggtitle("Kobe's Streaks")
```



Exercise 3

In your simulation of flipping the unfair coin 100 times, how many flips came up heads? Include the code for sampling the unfair coin in your response. Since the markdown file will run the code, and generate a new sample each time you Knit it, you should also "set a seed" before you sample. Read more about setting a seed below.

-In my simulation, heads came up 26 times.

Hide

```

coin_outcomes <- c("heads", "tails")
#sample(coin_outcomes, size = 1, replace = TRUE)

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

#sim_fair_coin

#table(sim_fair_coin)

#unfair coin
set.seed(35797)
sim_unfair_coin <- sample(coin_outcomes, size = 100, replace = TRUE, prob = c(0.2,
0.8))

table(sim_unfair_coin)

```

```

## sim_unfair_coin
## heads tails
##      26      74

```

Exercise 4

What change needs to be made to the sample function so that it reflects a shooting percentage of 45%? Make this adjustment, then run a simulation to sample 133 shots. Assign the output of this simulation to a new object called `sim_basket`.

-To change the shooting percentage to 45%, we must assign the probabilities of the respective outcomes as 0.45 for "H" and 0.55 for "M".

Hide

```

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

sim_basket

```

```

##      [1] "H" "H" "H" "M" "M" "M" "H" "H" "H" "M" "M" "H" "M" "H" "M" "M" "M" "M"
##     [19] "M" "M" "M" "M" "H" "H" "M" "M" "M" "H" "H" "H" "H" "M" "M" "M" "M" "M"
##     [37] "M" "M" "M" "M" "M" "H" "H" "M" "H" "M" "M" "M" "M" "M" "H" "H" "H" "H"
##     [55] "H" "H" "M" "M" "H" "M" "M" "M" "M" "M" "M" "M" "H" "H" "M" "M" "M" "M"
##     [73] "M" "M" "M" "M" "M" "H" "M" "H" "M" "H" "M" "M" "H" "M" "M" "M" "H" "H"
##     [91] "M" "M" "M" "M" "M" "M" "H" "M" "M" "M" "H" "H" "M" "H" "H" "H" "H" "H"
##    [109] "H" "M" "M" "M" "M" "M" "H" "M" "H" "H" "H" "H" "H" "M" "M" "M" "H" "H"
##    [127] "M" "H" "H" "H" "M" "M" "H"

```

Hide

```

table(sim_basket)

```

```
## sim_basket
## H M
## 53 80
```

Exercise 5

Using `calc_streak`, compute the streak lengths of `sim_basket`, and save the results in a data frame called `sim_streak`.

[Hide](#)

```
sim_streak <- calc_streak(sim_basket)

sim_streak
```

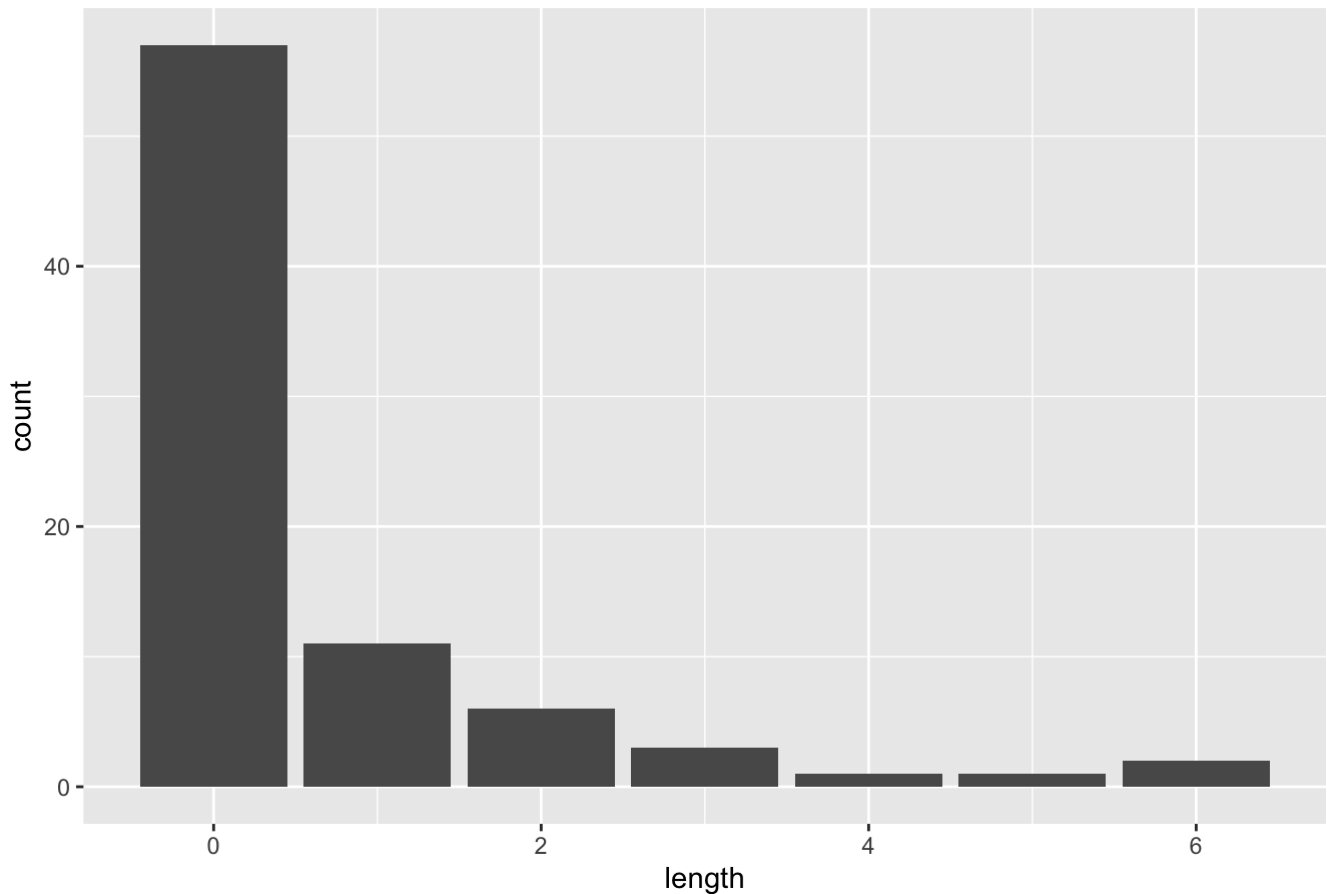
##	length
## 1	3
## 2	0
## 3	0
## 4	3
## 5	0
## 6	1
## 7	1
## 8	0
## 9	0
## 10	0
## 11	0
## 12	0
## 13	0
## 14	0
## 15	2
## 16	0
## 17	0
## 18	4
## 19	0
## 20	0
## 21	0
## 22	0
## 23	0
## 24	0
## 25	0
## 26	0
## 27	0
## 28	2
## 29	1
## 30	0
## 31	0
## 32	0
## 33	0
## 34	6
## 35	0
## 36	1
## 37	0
## 38	0
## 39	0
## 40	0
## 41	0
## 42	0
## 43	2
## 44	0
## 45	0
## 46	0
## 47	0
## 48	0
## 49	0
## 50	0
## 51	0

```
## 52      1
## 53      1
## 54      1
## 55      0
## 56      1
## 57      0
## 58      0
## 59      2
## 60      0
## 61      0
## 62      0
## 63      0
## 64      0
## 65      1
## 66      0
## 67      0
## 68      2
## 69      6
## 70      0
## 71      0
## 72      0
## 73      0
## 74      1
## 75      5
## 76      0
## 77      0
## 78      2
## 79      3
## 80      0
## 81      1
```

Hide

```
ggplot(data = sim_streak, aes(x = length)) + geom_bar() + ggtitle("Independent Shooter's Streaks")
```

Independent Shooter's Streaks



Exercise 6

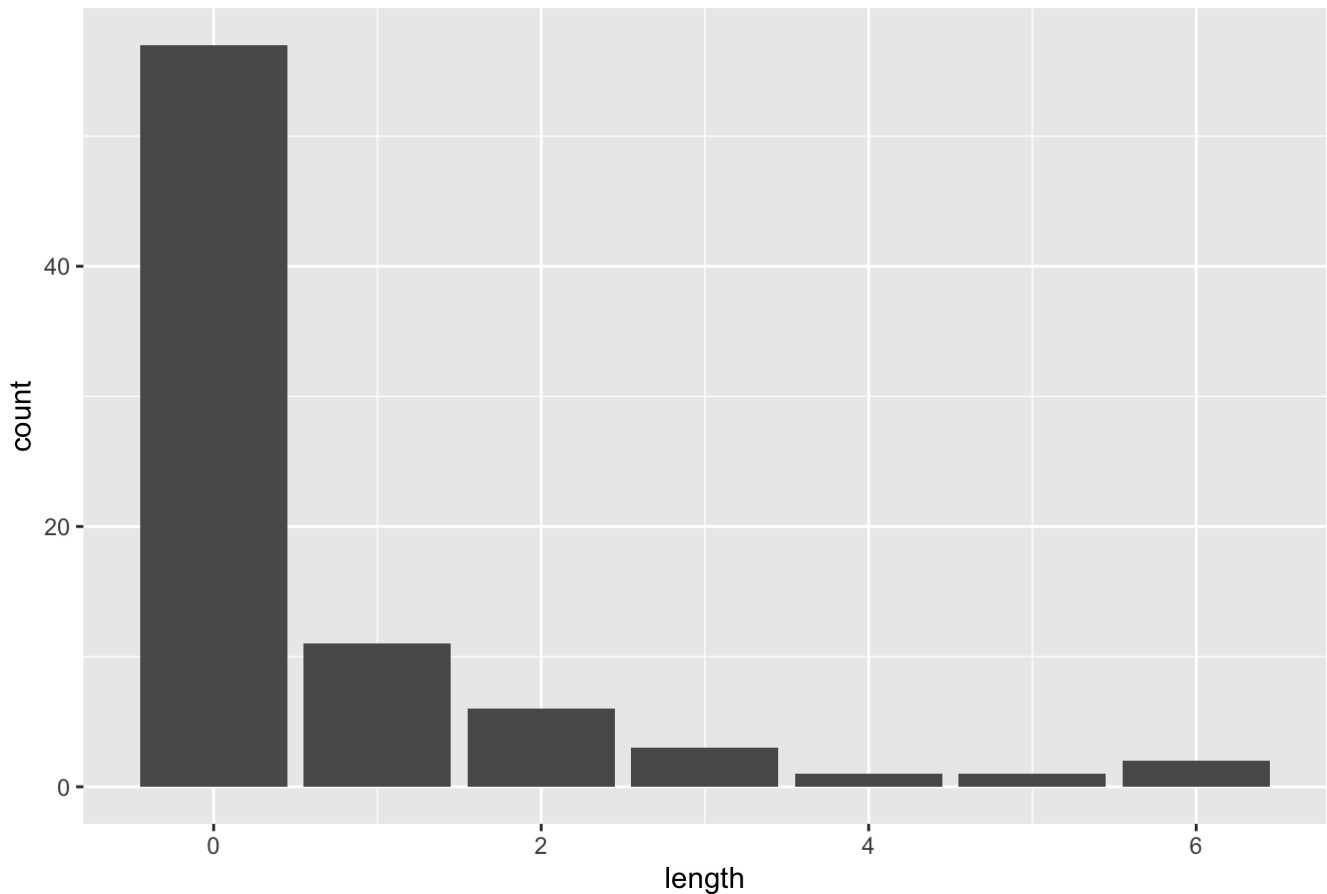
Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots? Make sure to include a plot in your answer.

-This distribution is also skewed right. The typical streak length for the independent shooter is 0 and the longest streak in 133 shots is 14.

Hide

```
set.seed(922)
ggplot(data = sim_streak, aes(x = length)) + geom_bar() + ggtitle("Independent Shooter's Streaks")
```

Independent Shooter's Streaks



Exercise 7

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? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

-Considering the seed function, it would be exactly the same. Disregarding the seed function, it should remain the same or very similar if we're talking about the distribution on the whole. The individual shot streaks may change because each run of the experiment yields different results, but the behavior overall should skew right as in the previous run.

Hide

```
sim_streak1 <- calc_streak(sim_basket)

sim_streak1
```

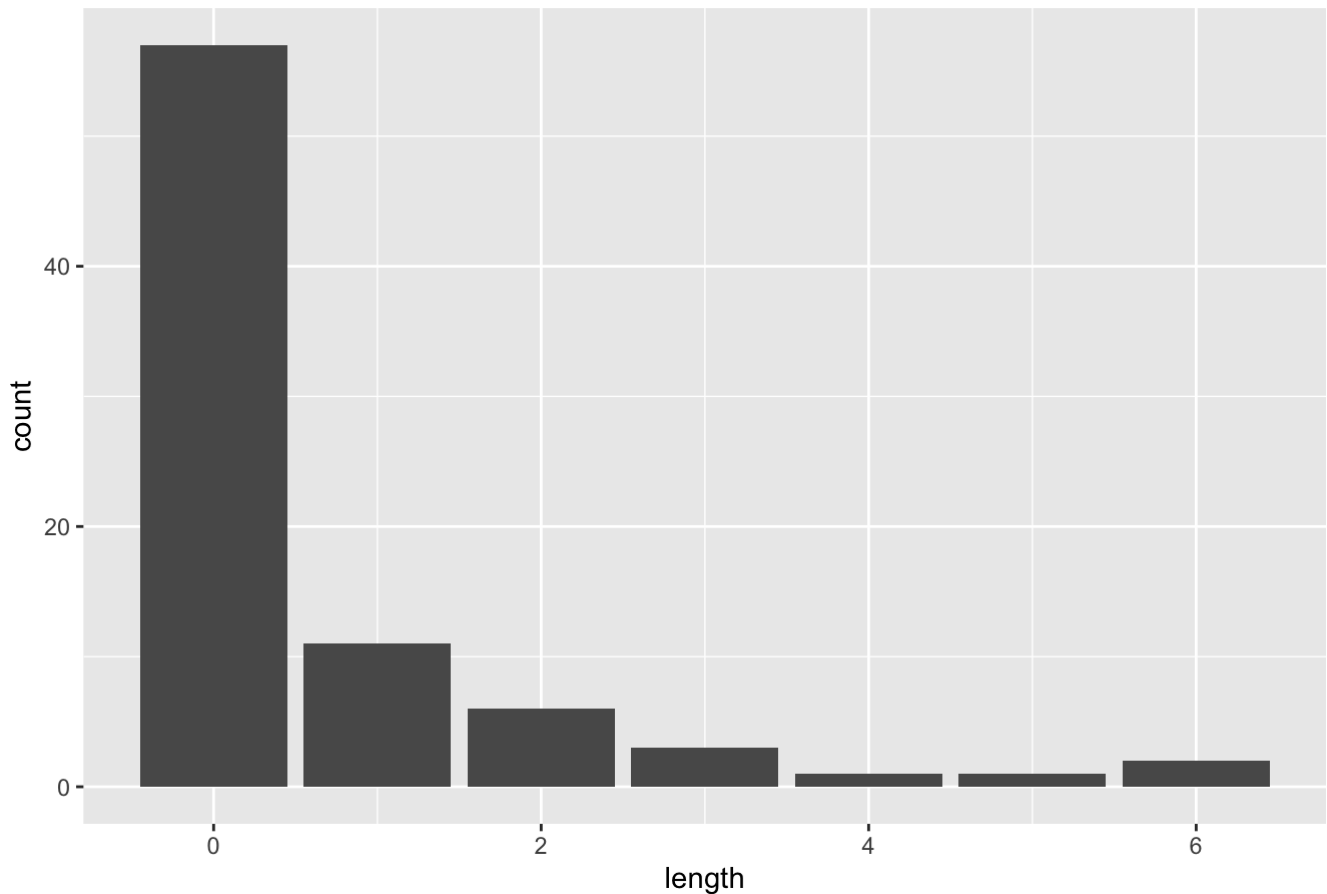

##	length
## 1	3
## 2	0
## 3	0
## 4	3
## 5	0
## 6	1
## 7	1
## 8	0
## 9	0
## 10	0
## 11	0
## 12	0
## 13	0
## 14	0
## 15	2
## 16	0
## 17	0
## 18	4
## 19	0
## 20	0
## 21	0
## 22	0
## 23	0
## 24	0
## 25	0
## 26	0
## 27	0
## 28	2
## 29	1
## 30	0
## 31	0
## 32	0
## 33	0
## 34	6
## 35	0
## 36	1
## 37	0
## 38	0
## 39	0
## 40	0
## 41	0
## 42	0
## 43	2
## 44	0
## 45	0
## 46	0
## 47	0
## 48	0
## 49	0
## 50	0
## 51	0

```
## 52      1
## 53      1
## 54      1
## 55      0
## 56      1
## 57      0
## 58      0
## 59      2
## 60      0
## 61      0
## 62      0
## 63      0
## 64      0
## 65      1
## 66      0
## 67      0
## 68      2
## 69      6
## 70      0
## 71      0
## 72      0
## 73      0
## 74      1
## 75      5
## 76      0
## 77      0
## 78      2
## 79      3
## 80      0
## 81      1
```

Hide

```
ggplot(data = sim_streak, aes(x = length)) + geom_bar() + ggtitle("Independent Shooter's Streaks")
```

Independent Shooter's Streaks



Exercise 8

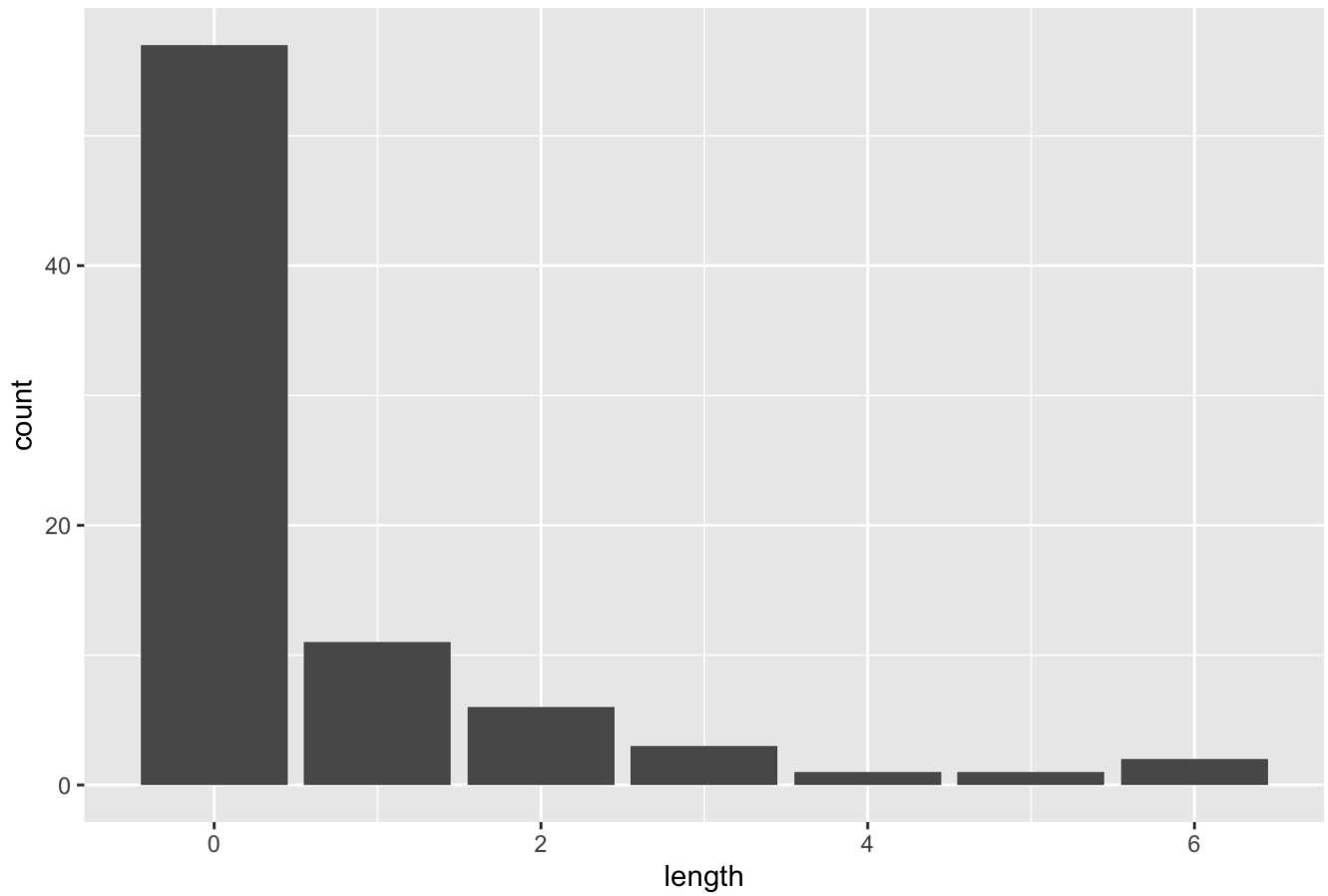
How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe's shooting patterns? Explain.

-Both distributions skew right/have a right tail with a mode of 0 and similar longest streaks. The simulated shooter is a good model that mostly fits Kobe's shooting patterns.

Hide

```
set.seed(927)
ggplot(data = sim_streak, aes(x = length)) + geom_bar() + ggtitle("Independent Shooter's Streaks")
```

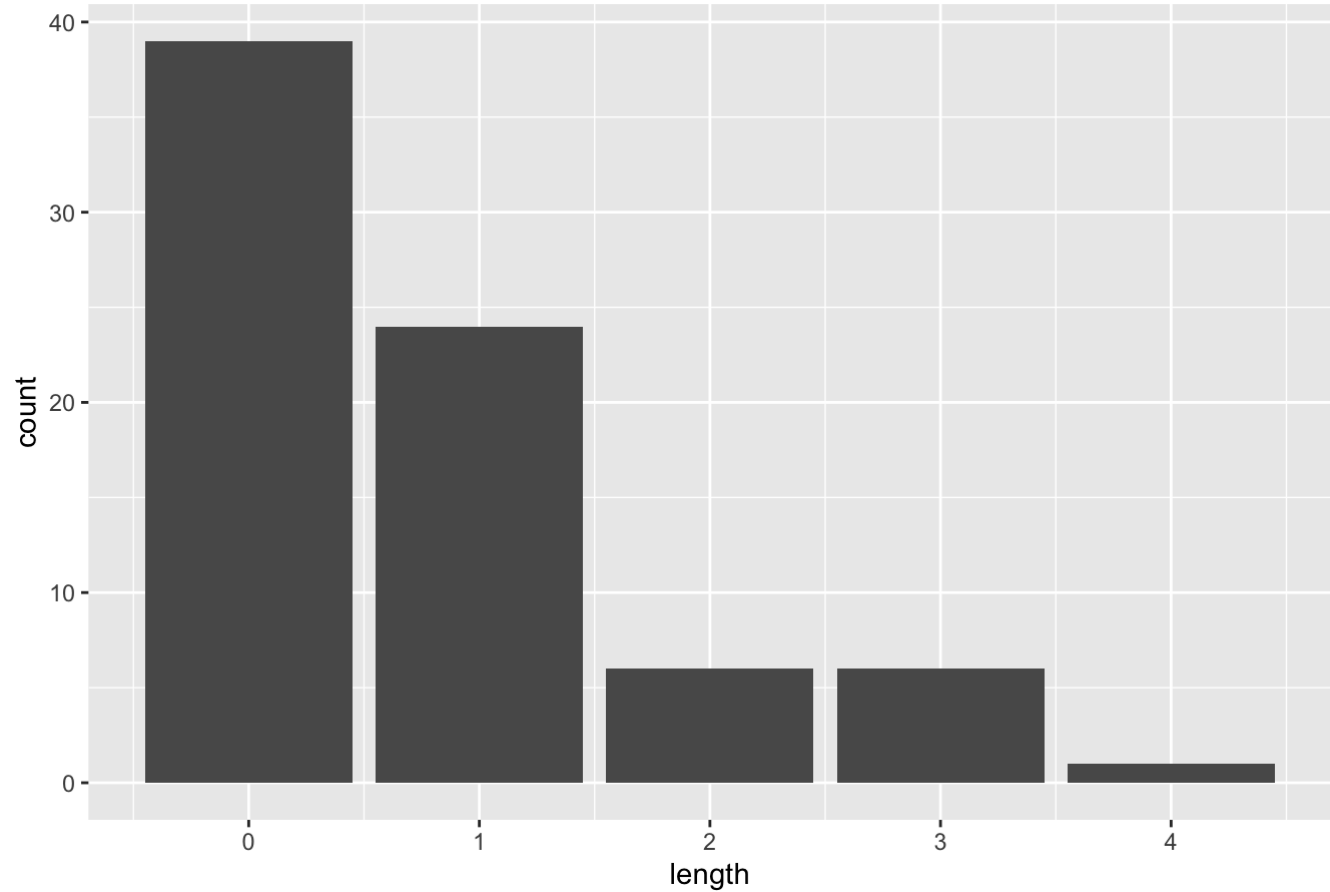
Independent Shooter's Streaks



Hide

```
ggplot(data = kobe_streak, aes(x = length)) + geom_bar() + ggtitle("Kobe's Streaks")
```

Kobe's Streaks



...