Lab 3, Probability

Alice Ding

```
library(tidyverse)
library(openintro)
library(ggplot2)
data("kobe_basket")
```

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?

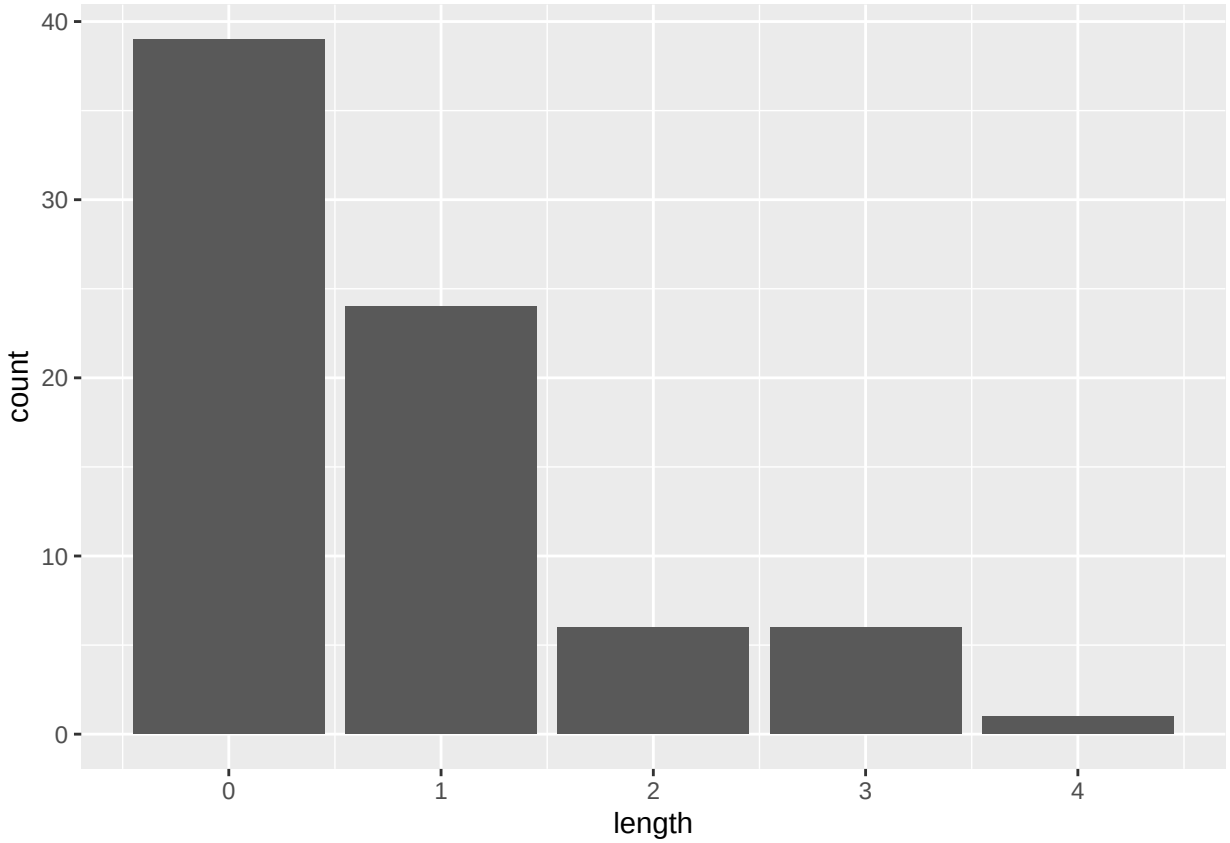
A stream length of one is 1 hit, 1 miss. A streak length of 0 is just 1 miss.

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.

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

```
ggplot(data = kobe_streak, aes(x = length)) +
  geom_bar()
```




```
kobe_streak_stats <- data.frame(kobe_streak) |> summarise(
  avg_streak = mean(as.numeric(unlist(kobe_streak))),
  median = median(as.numeric(unlist(kobe_streak)))
  )

kobe_streak_stats
```

```
##   avg_streak median
## 1  0.7631579      0
```

The distribution of this data leans more towards right in skewness as a majority of his streaks are 0 and 1. His typical streak length is either 0 or 1 – his average streak is 0.76 while the median is 0. I would say that typically, his streak would be 0 based on the fact that that is the median. His longest streak of baskets was 4.

Exericse 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.

```
set.seed(12345)

coin_outcomes <- c("heads", "tails")
```

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

```
## sim_fair_coin
## heads tails
##    46    54
```

```r
set.seed(12345)
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
##    20    80
```

20 flips came up heads.

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`.

```r
set.seed(12345)
shot_outcomes <- c("H", "M")
sim_basket <- sample(shot_outcomes, size = 1, replace = TRUE)
sim_basket <- sample(shot_outcomes, size = 133, replace = TRUE,
                          prob = c(0.45, 0.55))

table(sim_basket)
```

```
## sim_basket
##  H  M
## 65 68
```

Exercise 5

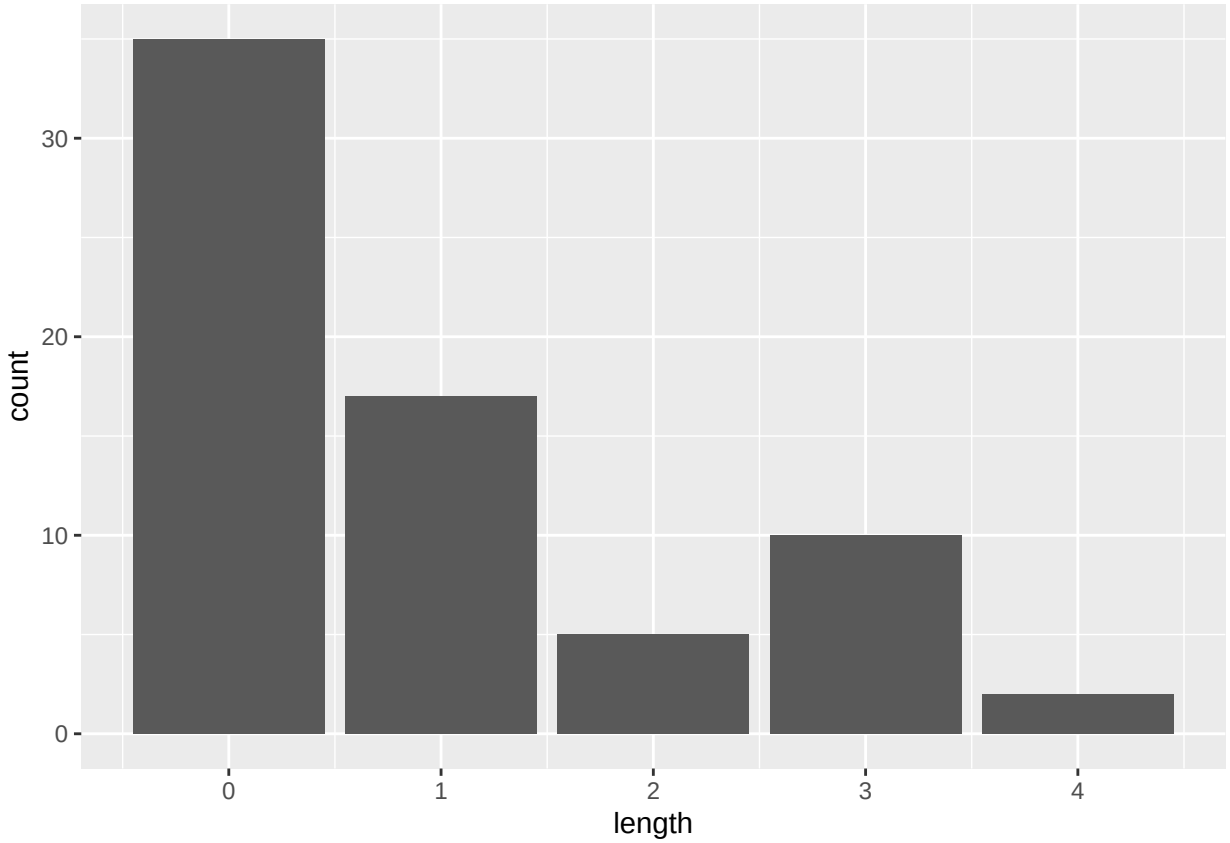
Using `calc_streak`, compute the streak lengths of `sim_basket`, and save the results in a data frame called `sim_streak`.

```r
sim_basket <- data.frame(sim_basket)
sim_streak <- calc_streak(sim_basket$sim_basket)
```

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.

```
ggplot(data = sim_streak, aes(x = length)) +
  geom_bar()
```




```
sim_streak |> data.frame(sim_streak) |> summarise(
  avg_streak = mean(as.numeric(unlist(sim_streak))),
  median = median(as.numeric(unlist(sim_streak)))
  )
```

```
##   avg_streak median
## 1   0.942029      0
```

The distribution of this data also leans more towards right in skewness as a majority of the streaks are 0 and 1, however there is a bit of a bump for a streak of 3. The typical streak length is still 0 or 1 – the average streak is .94 while the median is still 0. I would still say this simulation's typical streak is 0 and the longest streak of baskets is 4.

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.

> Assuming it's a different seed, I would expect it to be somewhat similar. Due to basic probability, it's very unlikely that the data would end up exactly the same, but due to the same weights in whether or not this independent shooter will land a shot, it'll follow the same trends and likely have similar looking data.

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.

> Kobe's distribution of streak lengths is pretty similar/comparable to this simulation, the only difference really is a few of the 1 streaks seemed to turn into 3 streaks interestingly enough. The skewness of the data though is very similar as they're both right leaning; the hot hand model doesn't seem to fit Kobe's shooting patterns since the independent shooter's data is similar to Kobe's and the independent shooter does *not* have a hot hand.