

Case Study with Data Applications

Professor Bradley Warner

21 May, 2020

```
library(mosaic)
library(tidyverse)
```

```
days <- seq(1,365)
```

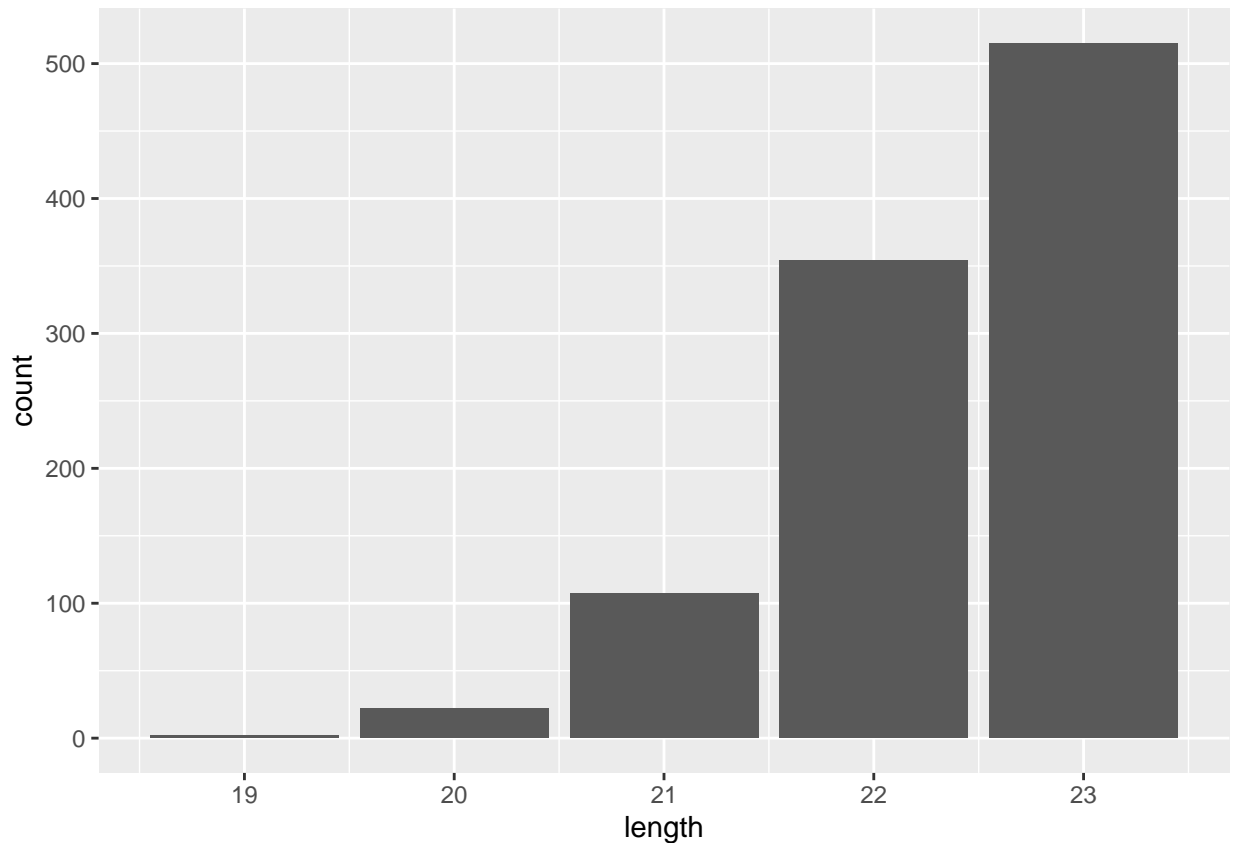
Exercises

1. **Exactly 2 people with the same birthday - Simulation.** Complete a similar analysis for case where exactly 2 people in a room of 23 people have the same birthday. In this exercise you will use a computational simulation.
 - a) Create a new R Markdown file.
 - b) Simulate having 23 people in the class which each day of the year equally likely. Find the cases where exactly 2 people have the same birthday, you will have to alter the code more than changing 18 to 23.
 - c) Plot the frequency of occurrences as a bar chart.
 - d) Estimate the probability of exactly two people having the same birthday.

```
(do(10000)*length(unique(sample(days,size=23,replace = TRUE)))) %>%
  mutate(match=if_else(length==22,1,0)) %>%
  summarise(prob=mean(match))
```

```
##      prob
## 1 0.3695
```

```
(do(1000)*length(unique(sample(days,size=23,replace = TRUE)))) %>%
  gf_bar(~length)
```



2. **Exactly 2 people with the same birthday - Mathematical.** Repeat problem 1 but do it mathematically. As a big hint, you will need to use the `choose()` function. The idea is that with 23 people we need to choose 2 of them to match. We thus need to multiply, the multiplication rule again, by `choose(23,2)`. If you are having trouble, work with 3 people in the room first.

- Find a formula for the exact probability of exactly 2 people in a room of 23 have the same birthday.
- Generalize your solution to any number n people and create a function.
- Vectorize the function.
- Plot the probability of exactly 2 people have the same birthday versus number of people in the room.
- Comment on the shape of the curve and explain it.
- `knit` and compile your report.

For two people we have

```
choose(23,2)*prod(365:344)/365^23
```

```
## [1] 0.3634222
```

```
exactly_two <- function(n){
  choose(n,2)*prod(365:(365-(n-2)))/365^n
}
```

```
exactly_two(23)
```

```
## [1] 0.3634222
```

```
exactly_two <- Vectorize(exactly_two)
```

```
gf_line(exactly_two(1:100)~ seq(1,100),  
  xlab="Number of People",  
  ylab="Probability of Match",  
  title="Probability of exactly least 2 people with matching birthdays")
```

