Case Study with Data Applications

Professor Bradley Warner

26 May, 2020

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

```r
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 occurences as a bar chart.

d) Estimate the probability of exactly two people having the same birthday.

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

```
##      prob
## 1 0.3615
```

```r
(do(1000)*length(unique(sample(days,size=23,replace = TRUE)))) %>%
  gf_bar(~length)
```
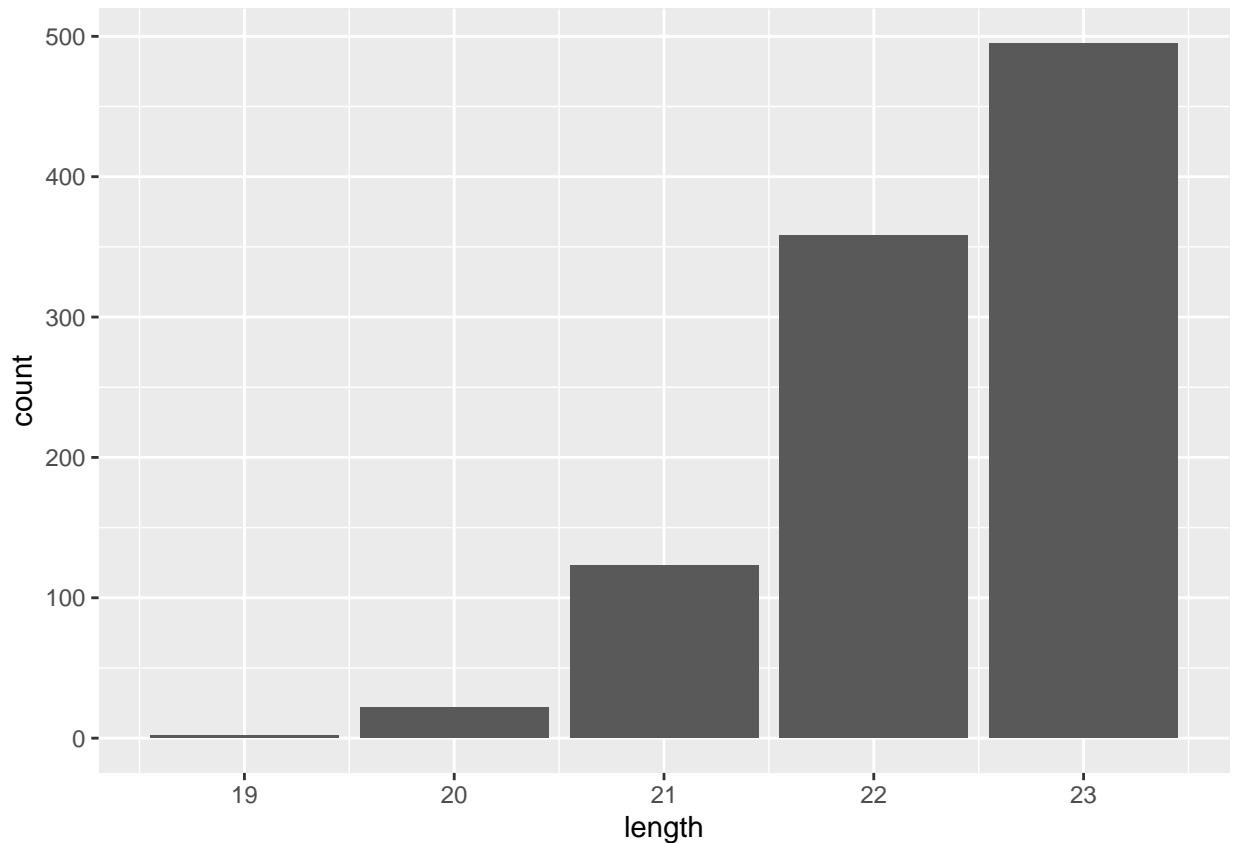
2. **Exactly 2 people with the same birthday - Mathematical**. Repeat problem 1 but do it mathematically. As a big hint, you will need to use the `choose()` function. The idea is that will 23 people we need to choose 2 of them to match. We thus need to multiply, the multiplication rule again, by `choose(23,2)`. If you are having trouble, work with 3 people in the room first.

   a) Find a formula for the exact probability of exactly 2 people in a room of 23 have the same birthday.

   b) Generalize your solution to any number `n` people and create a function.

   c) Vectorize the function.

   d) Plot the probability of exactly 2 people have the same birthday versus number of people in the room.

   e) Comment on the shape of the curve and explain it.

   f) `knit` and compile your report.

For two people we have

```
choose(23,2)*prod(365:344)/365^23
```
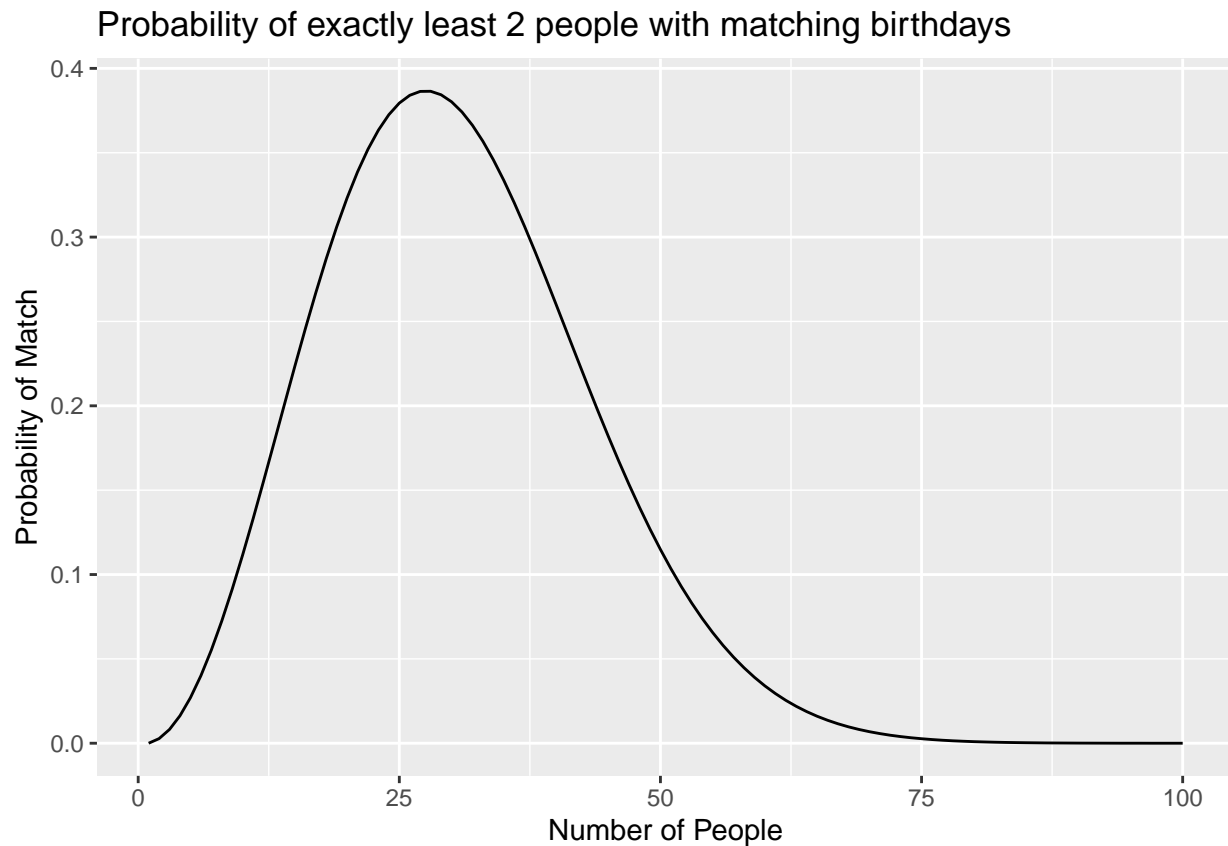
```
## [1] 0.3634222
```

```
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")
```

Probability of exactly least 2 people with matching birthdays

By the way, exactly three matches in simulation is hard. We have to table the data

```
set.seed(10)
temp <- table(sample(days,size=23,replace = TRUE))
temp
```

```
##
##  13  24  50  72  92 110 137 143 154 155 211 231 263 271 285 330 338 342 344 351
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   2   2   1   1   1   1
## 365
##   1
```

```
(sum(temp==2) == 2)+0
```

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

```
(do(10000)*((sum(table(sample(days,size=23,replace = TRUE)) == 3)==1)+0)) %>%
  summarise(prob=mean(result))
```

```
##     prob
## 1 0.0117
```

Two sets that have same but different birthday

```
(do(10000)*((sum(table(sample(days,size=23,replace = TRUE)) == 2)==2)+0)) %>%
  summarise(prob=mean(result))
```

```
##     prob
## 1 0.1139
```

```
(do(10000)*length(unique(sample(days,size=23,replace = TRUE)))) %>%
  mutate(match=if_else(length==21,1,0)) %>%
  summarise(prob=mean(match))
```

```
##     prob
## 1 0.1187
```

Mathematically exactly 3 is easy. Simulation seems to be off a little or the math formula is off.

```
choose(23,3)*prod(365:345)/365^23
```

```
## [1] 0.007395218
```