Case Study with Data Applications

Professor Bradley Warner

31 August, 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 and create a report. Yes, we know you could use this file but we want you to practice generating your own report.

b. Simulate having 23 people in the class with 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 from the Notes more than changing 18 to 23.

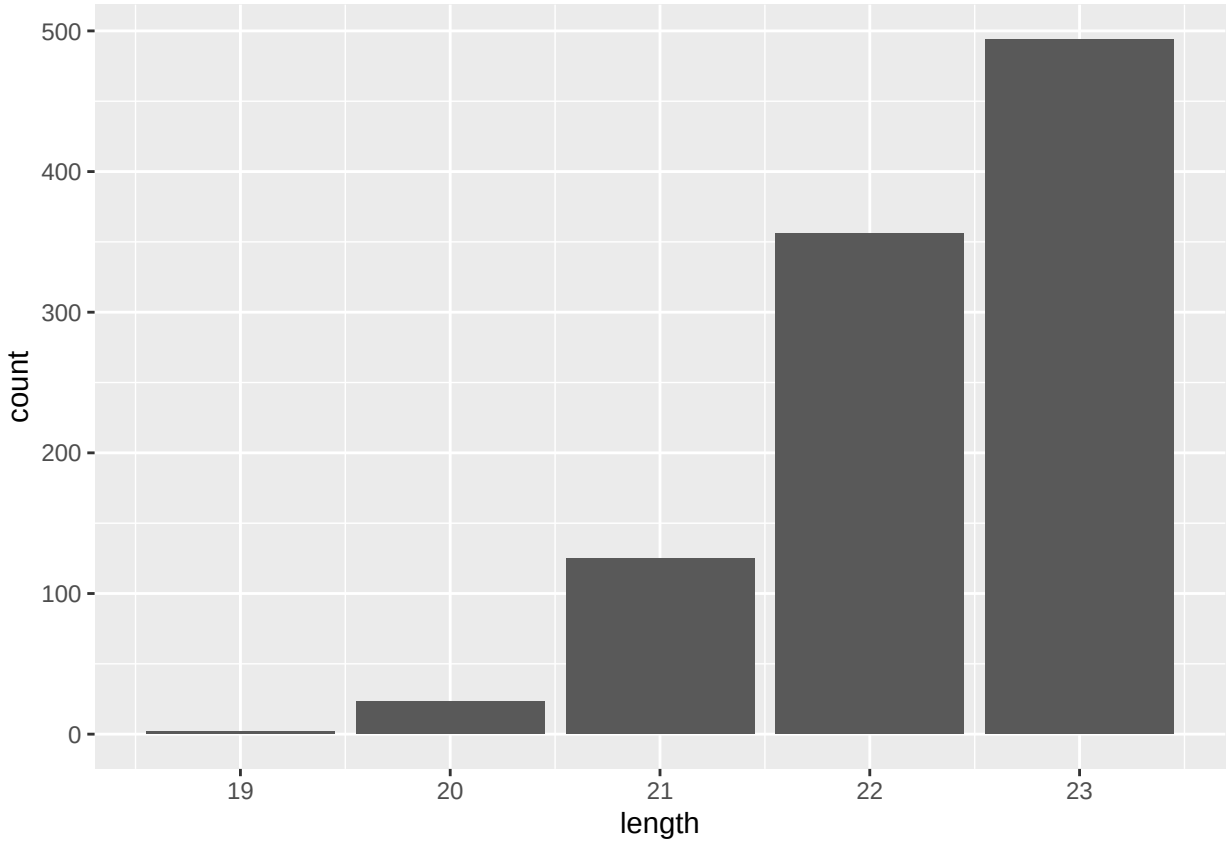
c. Plot the frequency of occurrences as a bar chart.

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 0.36
```

```
(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 a total of 3 people in the room first.

a. Find a formula to determine the exact probability of exactly 2 people in a room of 23 having the same birthday.

b. Generalize your solution to any number `n` people in the room and create a function.

c. Vectorize the function.

d. Plot the probability of exactly 2 people having 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
```

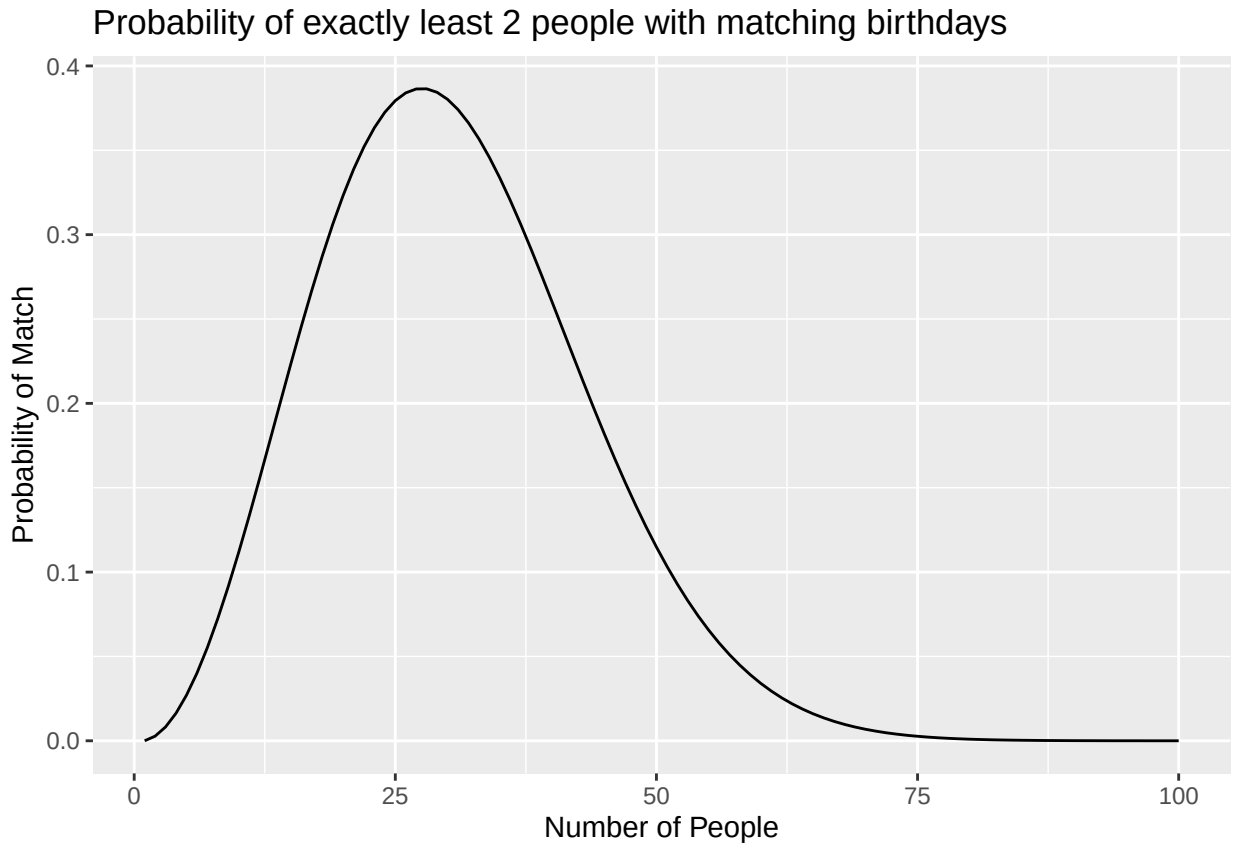
```
exactly_two <- function(n){
  choose(n,2)*prod(365:(365-(n-2)))/365^n
}
```

```
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")
```



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
```

File Creation Information

- File creation date: 2020-08-31
- R version 3.6.3 (2020-02-29)
- `mosaic` package version: 1.7.0
- `tidyverse` package version: 1.3.0