

FB friends Birthday

Hindam

January 13, 2019

```
# read csv file:
df <- read.csv("Friends' Birthdays.csv",header =T, check.names = F, row.names = NULL)
```

```
#View the data and take some insights
View(df)
dim(df)
```

```
## [1] 322 17
```

```
# remove the duplicated row
df <- df[-c(1), ]
View(df)
```

```
# remove the un_necessary columns
df<- df[-c(3:17)]
View(df)
```

```
# final dimension needed
dim(df)
```

```
## [1] 321 2
```

```
# adjust the title column by removing the ('s birthday)
df$title <- gsub("'s birthday","",df$title)
View(df)
```

Now time to extract the needed data of birthdate, first things first

```
# extract the day and month from the date
df$new_date <- strptime(df$date,"%m/%d/%Y")
typeof(df$date)
```

```
## [1] "integer"
```

```
typeof(df$new_date)
```

```
## [1] "list"
```

```
# import packages
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##      date
```

```
# set a new column for the month
df$month <- month(as.POSIXlt(df$new_date, format="%m/%d/%Y"))
View(df)
```

Same we will do for the day

```
# set a new column for the day
df$day <- day(as.POSIXlt(df$new_date, format="%m/%d/%Y"))
View(df)
```

```
# Let's zoom for the needed data
df<- df[-c(2,3)]
View(df)
```

Now to the fun part

1. How many people have the same birthday as you?

```
subset(df, df$month== 7 & df$day == 29)
```

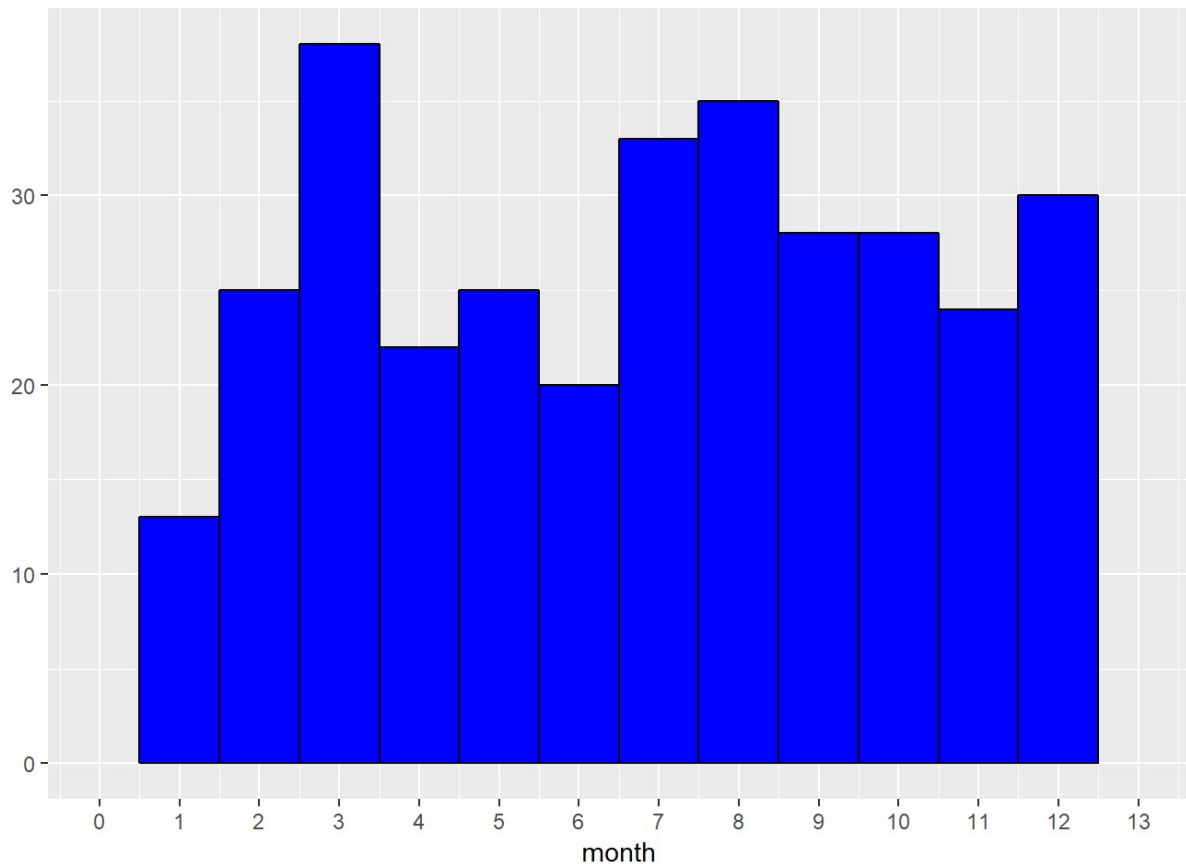
```
##              title month day
## 172   Biopharma Eg      7  29
## 173     Bazar Eta      7  29
## 174 Mohamed Hindam      7  29
```

Answer : is 3 , but the fun fact here that these other two are my own facebook accounts :)

2. Which month contains the most number of birthdays?

```
library(ggplot2)
qplot(x=month, data= df, binwidth=1,
      color=I("black"), fill=I("blue"))+
  scale_x_continuous(limits = c(0,13), breaks = seq(0,13,1))
```

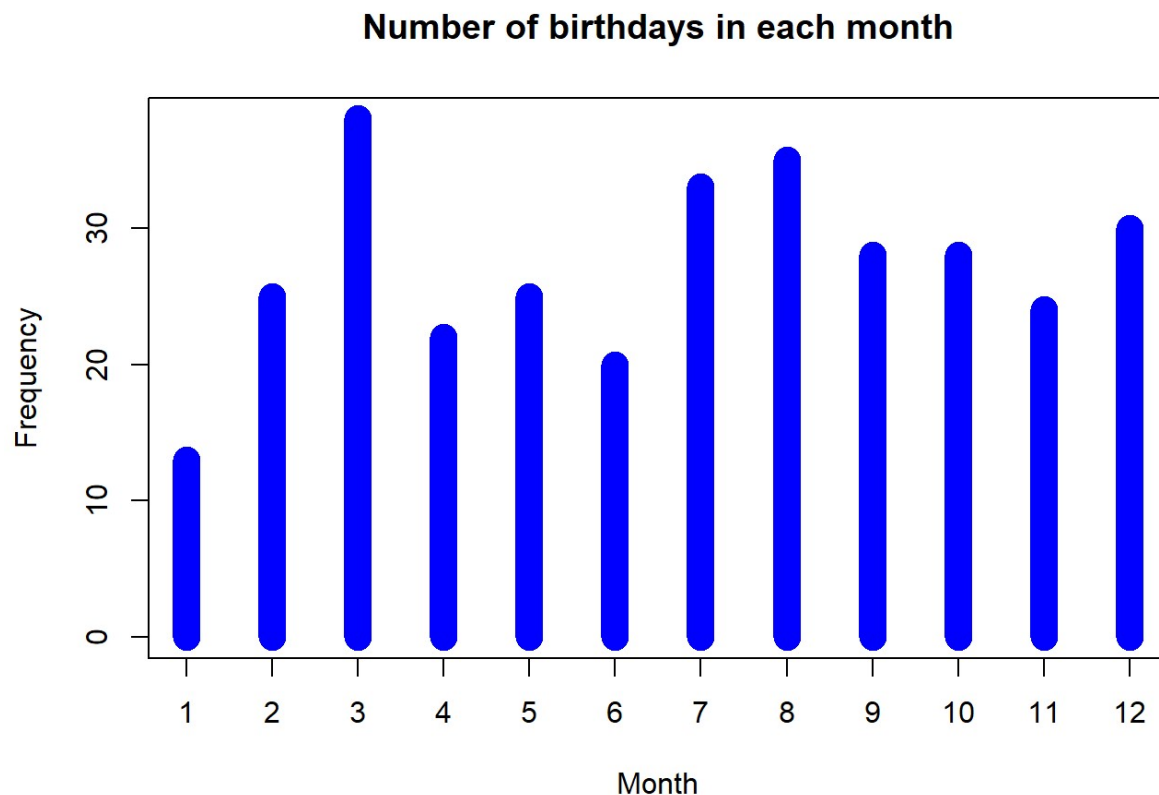
```
## Warning: Removed 2 rows containing missing values (geom_bar).
```



Answer : March

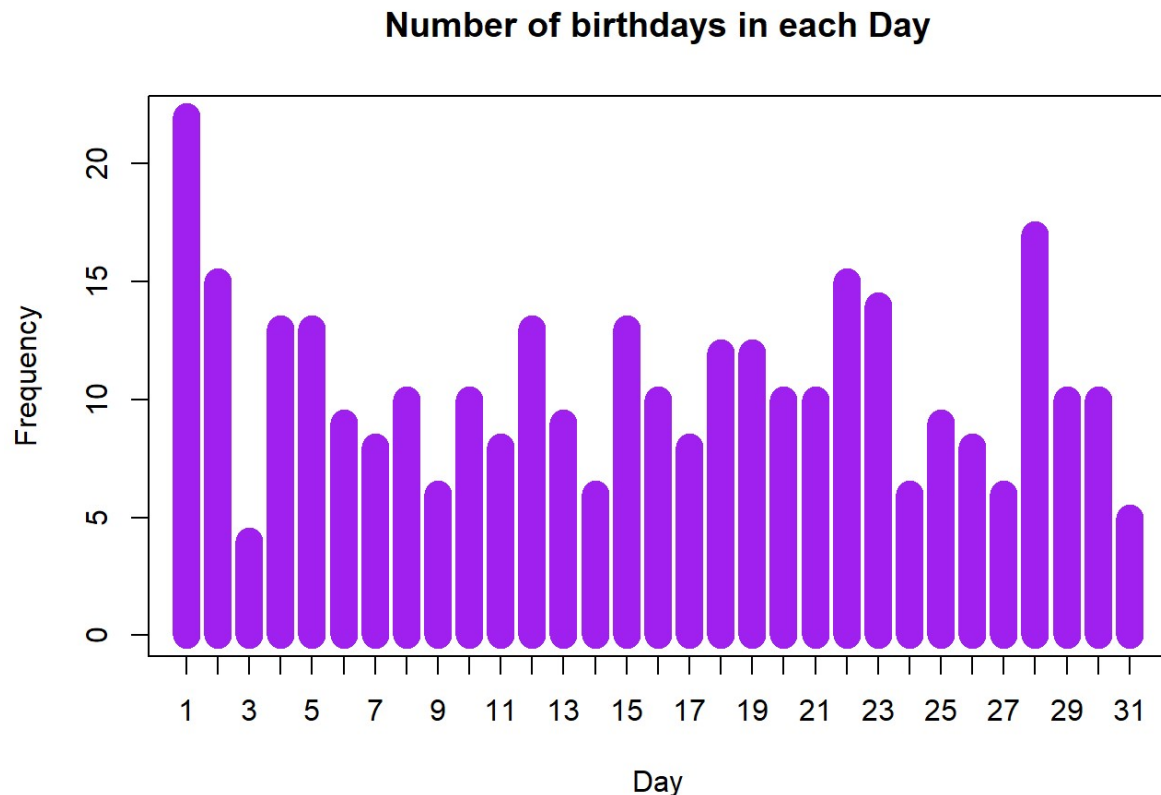
3. How many birthdays are in each month?

```
bpm <-table(df$month)
plot(bpm, type = "h", xlab = "Month", ylab = "Frequency",
     main = "Number of birthdays in each month", col="blue",lwd = 15)
```



4. Which day of the year has the most number of birthdays?

```
bpd <- table(df$day)
plot(bpd, type = "h", xlab = "Day", ylab = "Frequency",
     main = "Number of birthdays in each Day", col="purple",lwd = 15)
```



5. Do you have at least 365 friends that have birthdays on everyday of the year?

```
qplot(x= month, data= df, binwidth=1,
      color=I("white"), fill= I("orange"))+
  scale_x_continuous(limits = c(0,13), breaks = seq(1,12,1))+
  facet_wrap(day~.)
```

```
## Warning: Removed 62 rows containing missing values (geom_bar).
```

