

Evaluating environmental drivers of disease outbreaks from seasonal trends

K.H. Nguyen

7/10/2020

The following code and associated figures correspond to datasets obtained from the CDC NORS (<https://www.cdc.gov/norsdashboard/>) for the years 1998-2017. Questions can be sent to the corresponding author at [karena\(dot\)nguyen@gmail\(dot\)com](mailto:karena(dot)nguyen@gmail(dot)com).

```
## Set to your own working directory before beginning (example below)
setwd("C:/Users/khnguy3/Documents/2020-TIEE/Data")

# Use par(mfrow=c(1,3)) to view figures next to each other

## Create a string for months because all data sets have months labeled by number
months <- c("Jan", "Feb", "Mar", "April", "May", "June",
            "July", "Aug", "Sept", "Oct", "Nov", "Dec")

months_ordered <- factor(months, levels = c("Jan", "Feb", "Mar", "April", "May", "June",
            "July", "Aug", "Sept", "Oct", "Nov", "Dec"))
```

Cryptosporidium

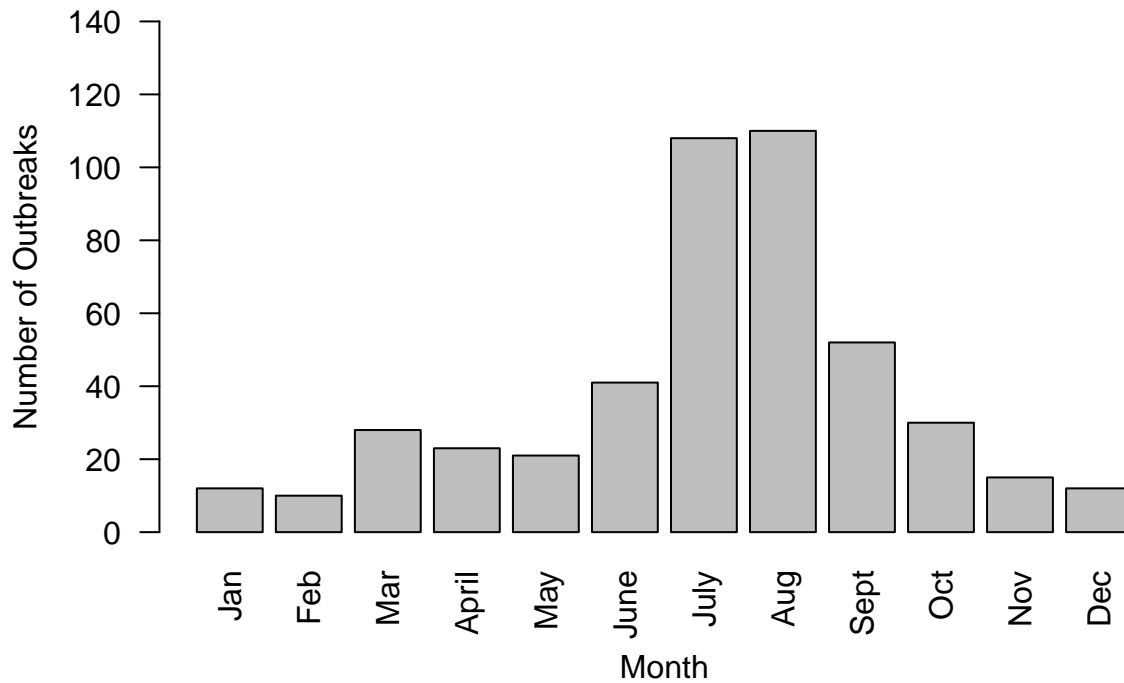
```
setwd("C:/Users/khnguy3/Documents/2020-TIEE/Data")
crypto <- read.csv(file = "Cryptosporidium.csv", header=TRUE)

# Calculate the number of outbreaks per month
# Use the function length to count the number of rows (i.e., outbreaks) across yrs per month
crypto_outbreaks <- aggregate(crypto, by = list(crypto$Month), FUN = length)

crypto_outbreaks$Group.1 <- months_ordered #Renames column to ordered months

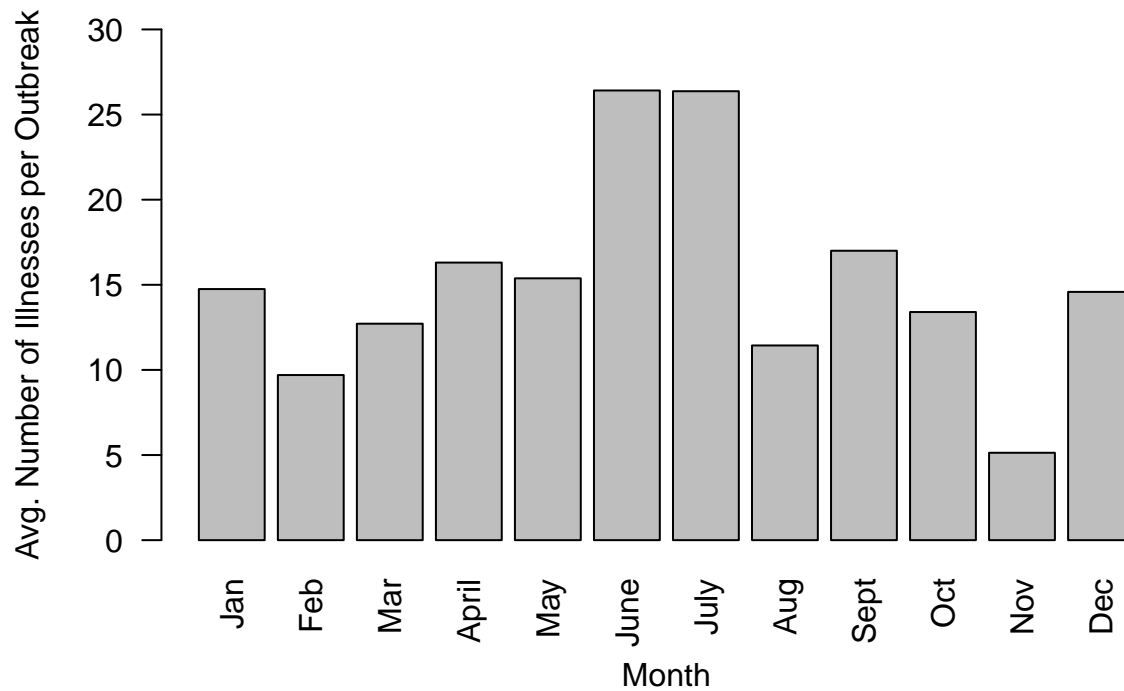
barplot(Month~Group.1, data= crypto_outbreaks,
        xlab = "Month", ylab = "Number of Outbreaks",
        main = "Cryptosporidium", ylim=c(0,140), las=2)
```

Cryptosporidium



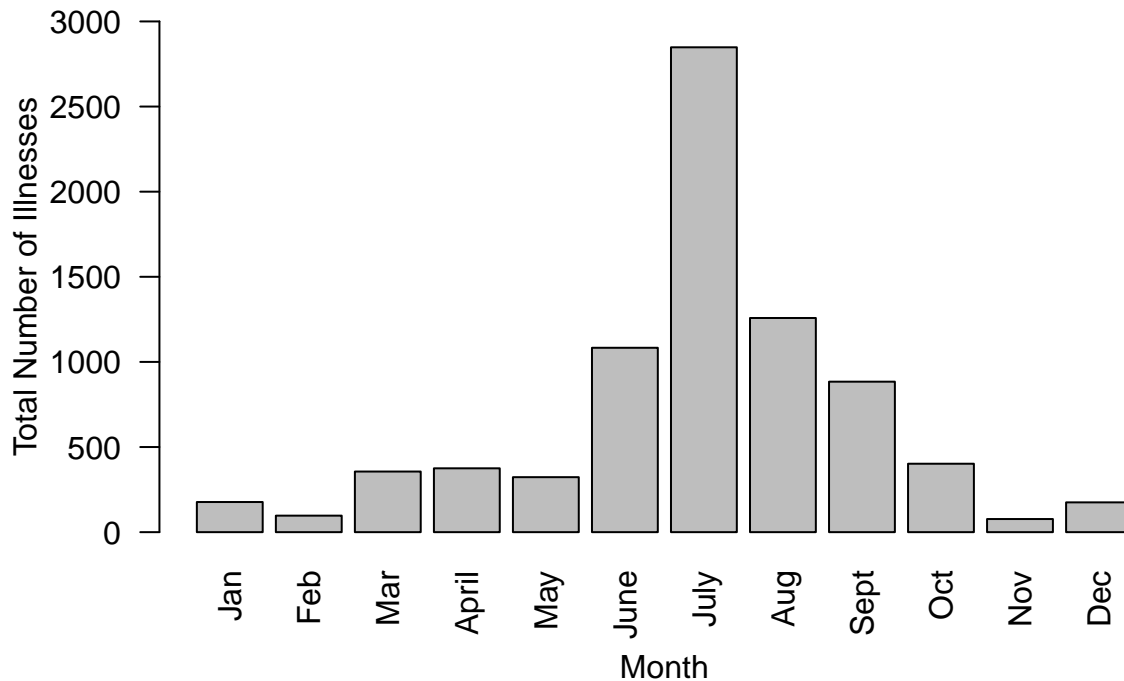
```
# Calculate the average size of outbreaks  
# The function mean will calculate the avg # of illnesses per outbreak per month  
crypto_size <- aggregate(Illnesses~Month, data= crypto, FUN = mean, na.rm = TRUE)  
  
crypto_size$Month <- months_ordered  
  
barplot(Illnesses~Month, data=crypto_size,  
        xlab = "Month", ylab = "Avg. Number of Illnesses per Outbreak",  
        main = "Cryptosporidium",  
        ylim=c(0,30), las=2)
```

Cryptosporidium



```
# Calculate the total number of illnesses per month  
# Use the function sum to add the total number of illnesses per month  
crypto_illnesses <- aggregate(Illnesses~Month, data= crypto, FUN = sum, na.rm = TRUE)  
  
crypto_illnesses$Month <- months_ordered  
  
barplot(Illnesses~Month, data=crypto_illnesses,  
        xlab = "Month", ylab = "Total Number of Illnesses", main = "Cryptosporidium",  
        ylim=c(0,3000), las=2)
```

Cryptosporidium



Cyclospora

```
setwd("C:/Users/khnguy3/Documents/2020-TIEE/Data")
cyclospora <- read.csv(file = "Cyclospora.csv", header=TRUE)

# Cyclospora does not have outbreaks documented for August - December
# we have to make a new data string
months2 <- c("Jan", "Feb", "Mar", "April", "May", "June", "July")

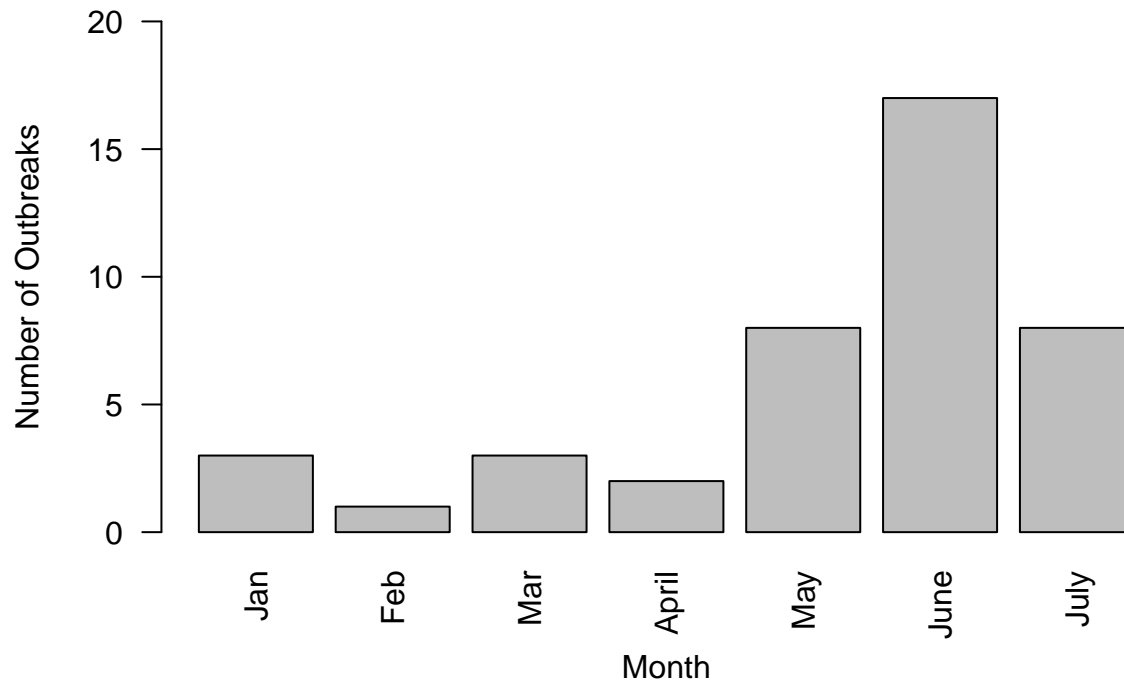
months_ordered2 <- factor(months2, levels = c("Jan", "Feb", "Mar", "April", "May", "June",
                                             "July"))

# Calculate the number of outbreaks per month
cyclo_outbreaks <- aggregate(cyclospora, by = list(cyclospora$Month), FUN = length)

cyclo_outbreaks$Group.1 <- months_ordered2

barplot(Month~Group.1, data= cyclo_outbreaks,
        xlab = "Month", ylab = "Number of Outbreaks",
        main = "Cyclospora", ylim=c(0,20), las=2)
```

Cyclospora

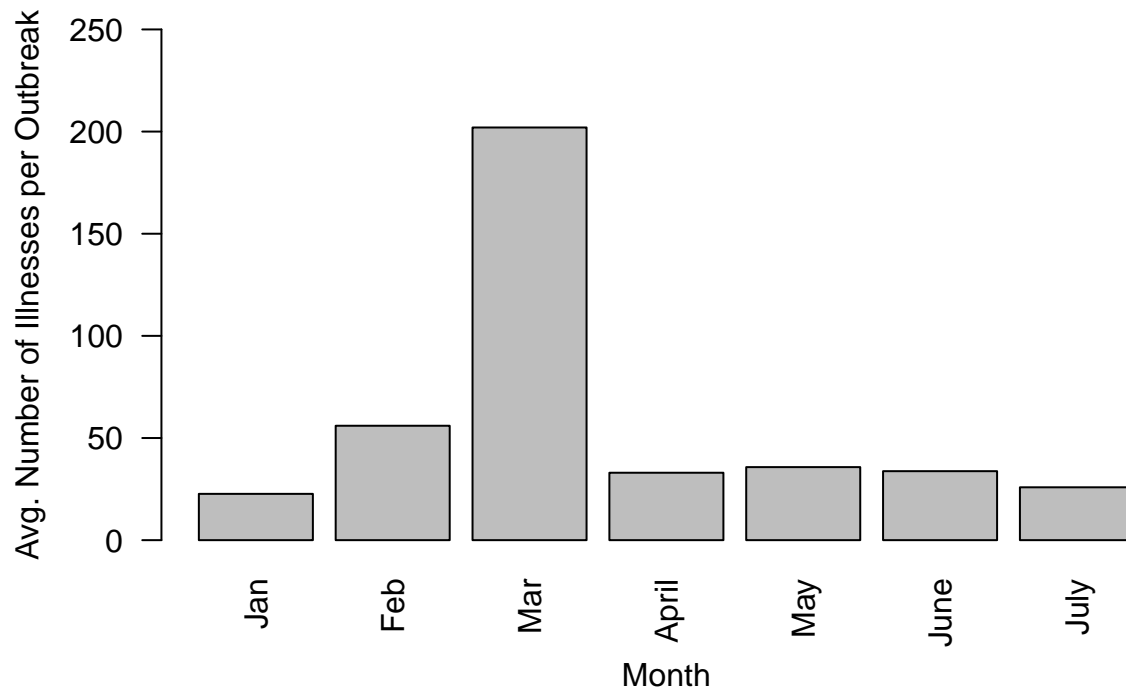


```
# Calculate the average size of outbreaks
cyclo_size <- aggregate(Illnesses~Month, data= cyclospora, FUN = mean, na.rm = TRUE)

cyclo_size$Month <- months_ordered2

barplot(Illnesses~Month, data=cyclo_size,
        xlab = "Month", ylab = "Avg. Number of Illnesses per Outbreak",
        main = "Cyclospora",
        ylim=c(0,250), las=2)
```

Cyclospora

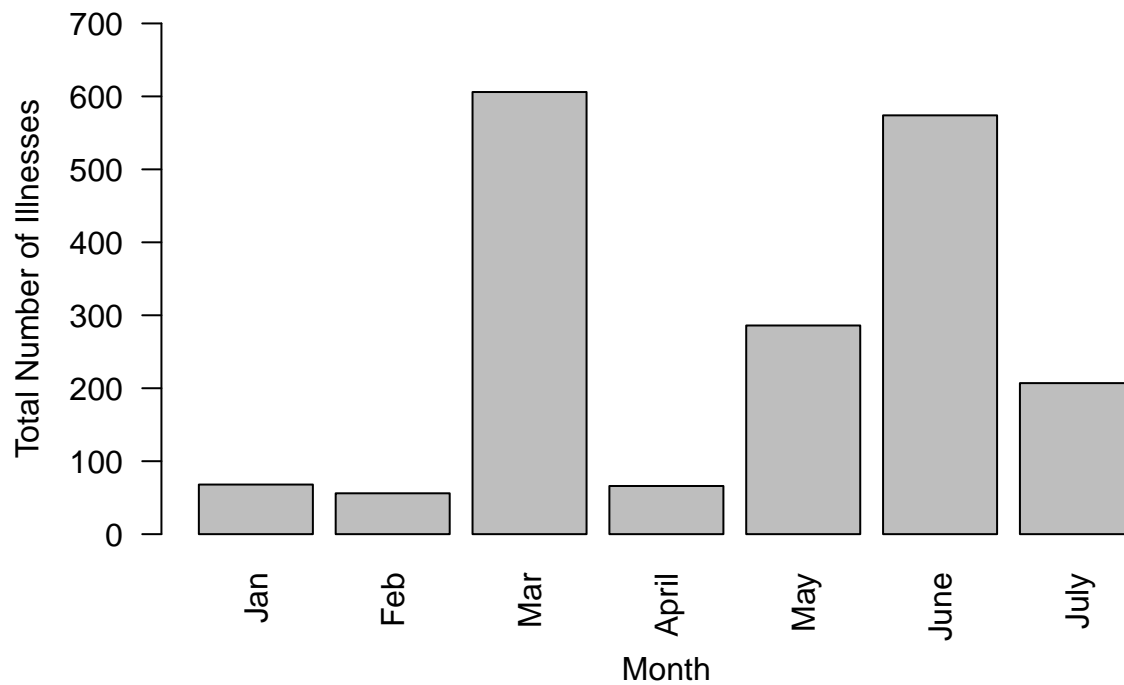


```
# Calculate the total number of illnesses per month
cyclo_illnesses <- aggregate(Illnesses~Month, data= cyclospora, FUN = sum, na.rm = TRUE)

cyclo_illnesses$Month <- months_ordered2

barplot(Illnesses~Month, data=cyclo_illnesses,
        xlab = "Month", ylab = "Total Number of Illnesses", main = "Cyclospora",
        ylim=c(0,700), las=2)
```

Cyclospora



Escherichia

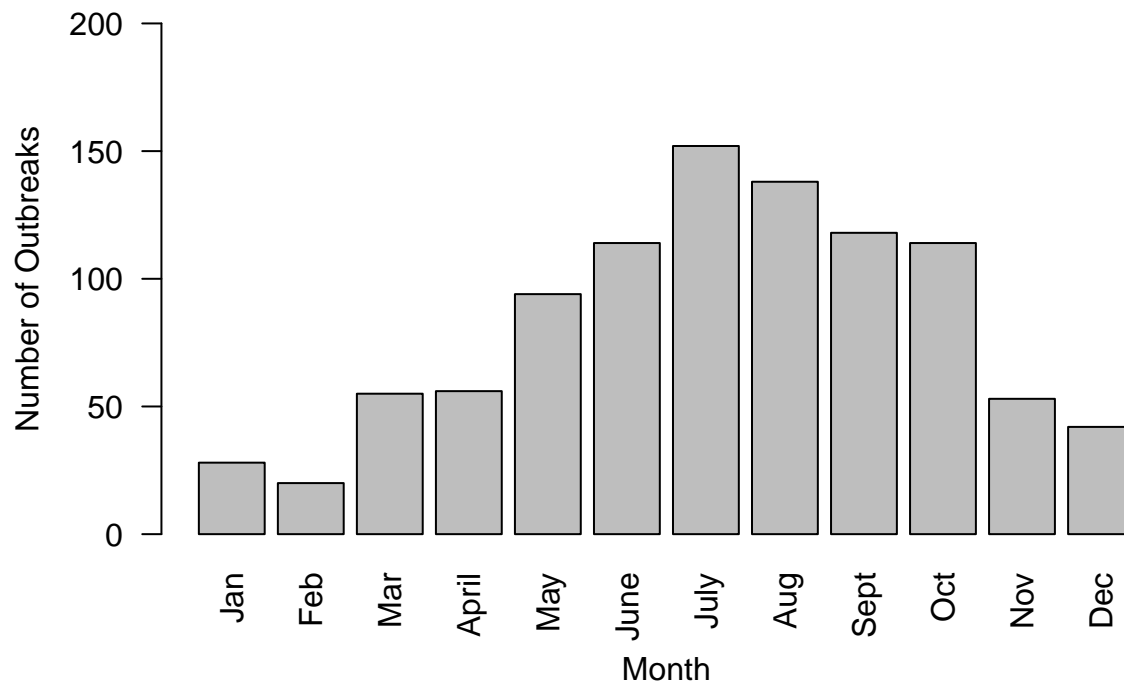
```
setwd("C:/Users/khnguy3/Documents/2020-TIEE/Data")
ecoli <- read.csv(file = "Escherichia.csv", header=TRUE)

# Calculate the number of outbreaks per month
ecoli_outbreaks <- aggregate(ecoli, by = list(ecoli$Month), FUN = length)

ecoli_outbreaks$Group.1 <- months_ordered

barplot(Month~Group.1, data= ecoli_outbreaks,
        xlab = "Month", ylab = "Number of Outbreaks",
        main = "Escherichia", ylim=c(0,200), las=2)
```

Escherichia

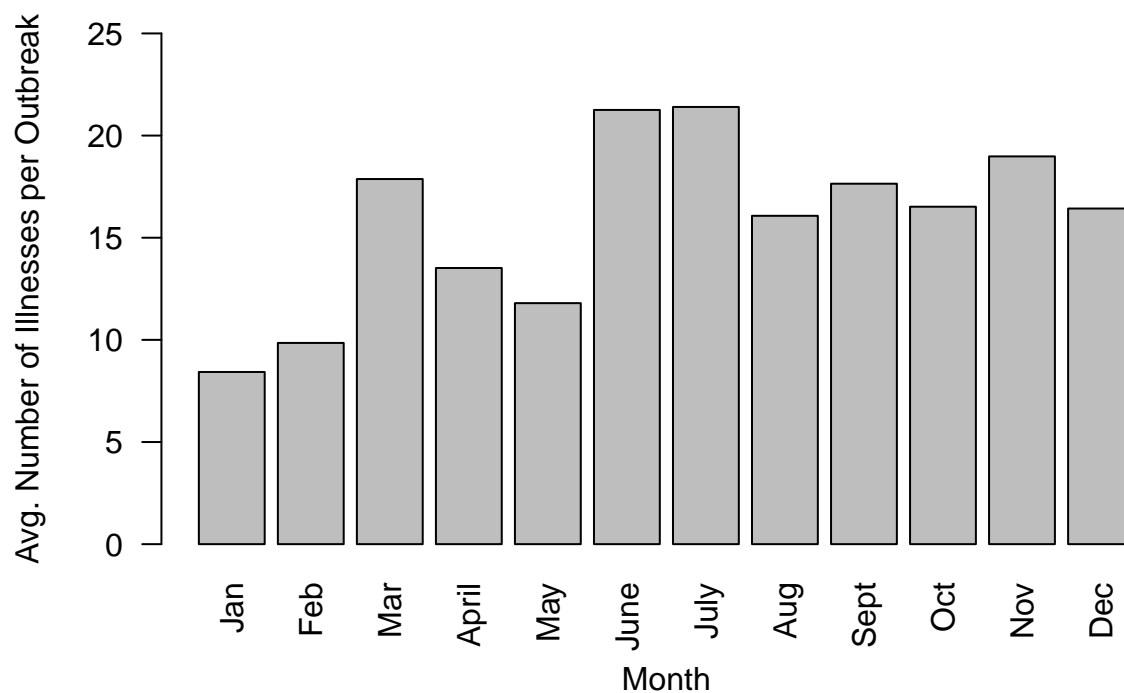


```
# Calculate the average size of outbreaks
ecoli_size <- aggregate(Illnesses~Month, data= ecoli, FUN = mean, na.rm = TRUE)

ecoli_size$Month <- months_ordered

barplot(Illnesses~Month, data=ecoli_size,
        xlab = "Month", ylab = "Avg. Number of Illnesses per Outbreak",
        main = "Escherichia", ylim=c(0,25), las=2)
```


Escherichia

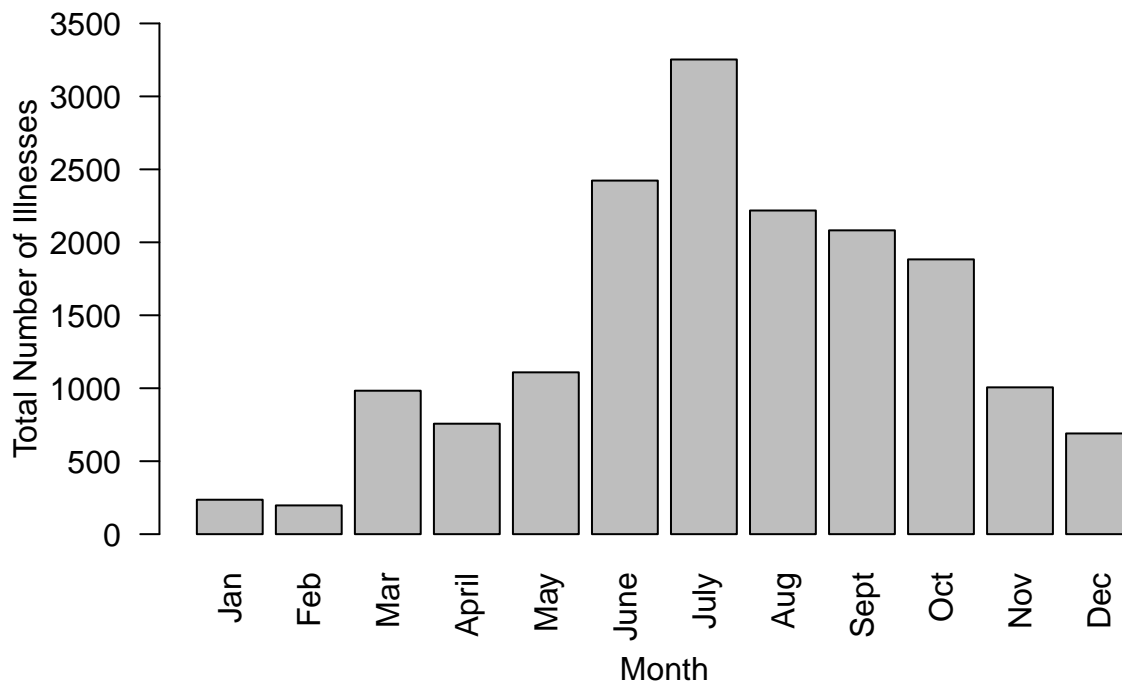


```
# Calculate the total number of illnesses per month
ecoli_illnesses <- aggregate(Illnesses~Month, data= ecoli, FUN = sum, na.rm = TRUE)

ecoli_illnesses$Month <- months_ordered

barplot(Illnesses~Month, data=ecoli_illnesses,
        xlab = "Month", ylab = "Total Number of Illnesses",
        main = "Escherichia", ylim=c(0,3500),
        las=2)
```

Escherichia



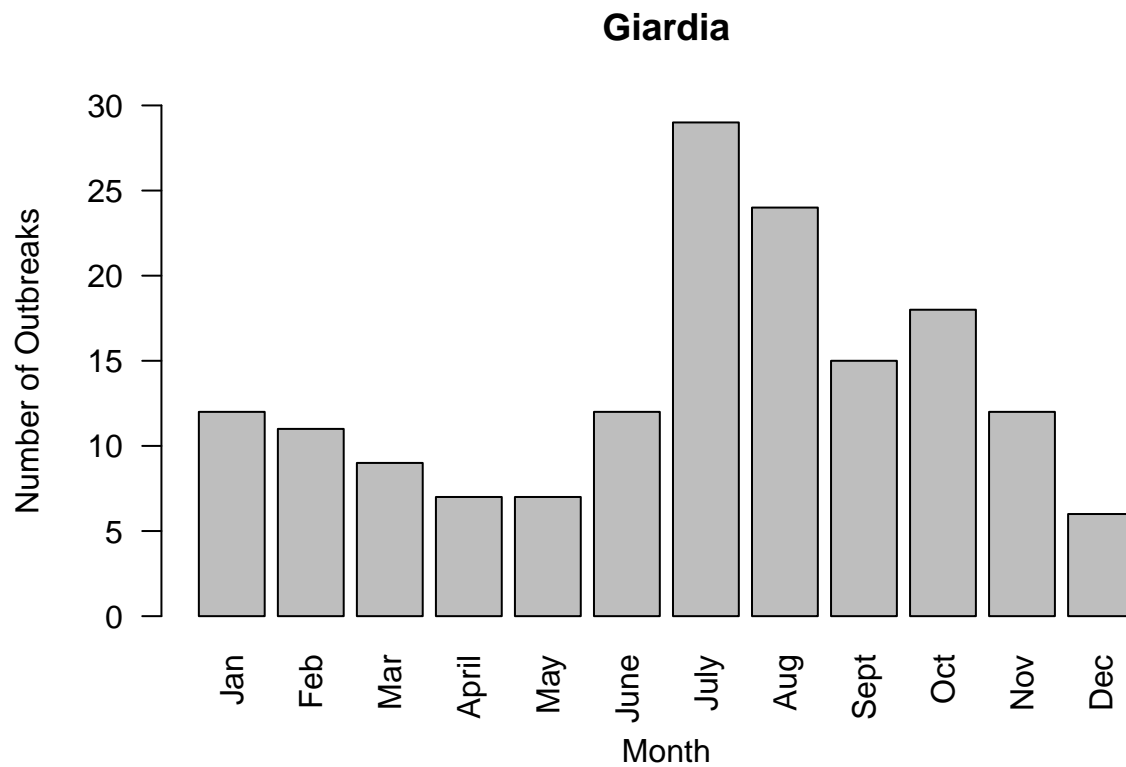
Giardia

```
setwd("C:/Users/khnguy3/Documents/2020-TIEE/Data")
giardia <- read.csv(file = "Giardia.csv", header=TRUE)

# Calculate the number of outbreaks per month
giardia_outbreaks <- aggregate(giardia, by = list(giardia$Month), FUN = length)

giardia_outbreaks$Group.1 <- months_ordered

barplot(Month~Group.1, data= giardia_outbreaks,
        xlab = "Month", ylab = "Number of Outbreaks",
        main = "Giardia", ylim=c(0,30), las=2)
```

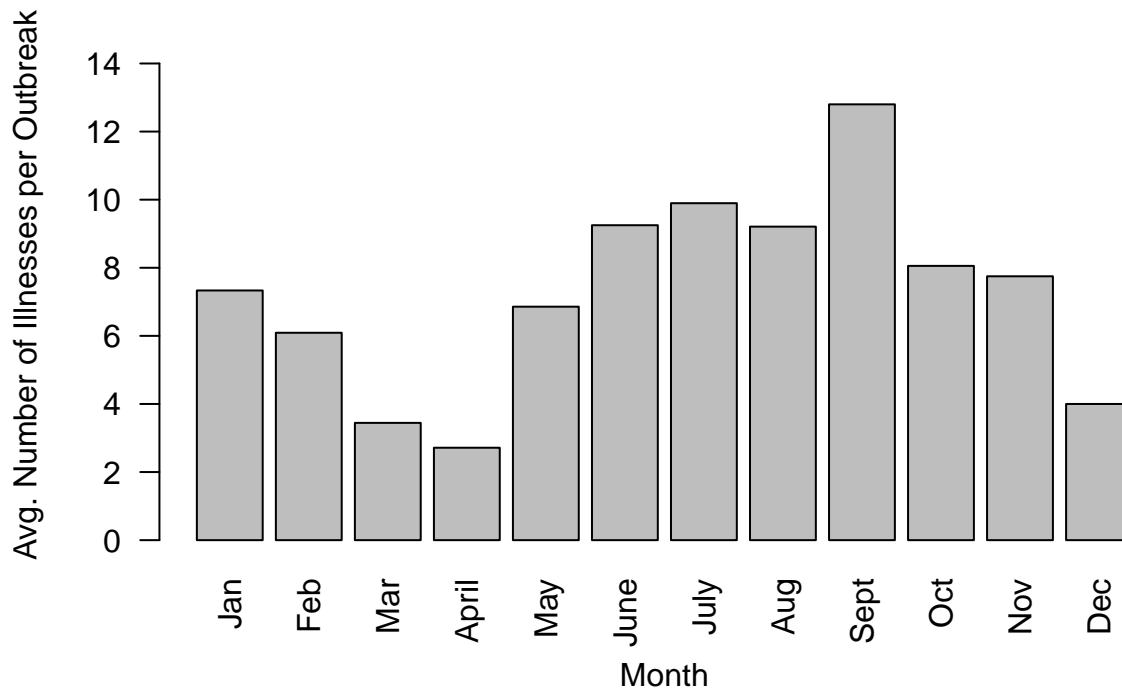


```
# Calculate the average size of outbreaks
giardia_size <- aggregate(Illnesses~Month, data= giardia, FUN = mean, na.rm = TRUE)

giardia_size$Month <- months_ordered

barplot(Illnesses~Month, data=giardia_size,
        xlab = "Month", ylab = "Avg. Number of Illnesses per Outbreak",
        main = "Giardia", ylim=c(0,15), las=2)
```

Giardia

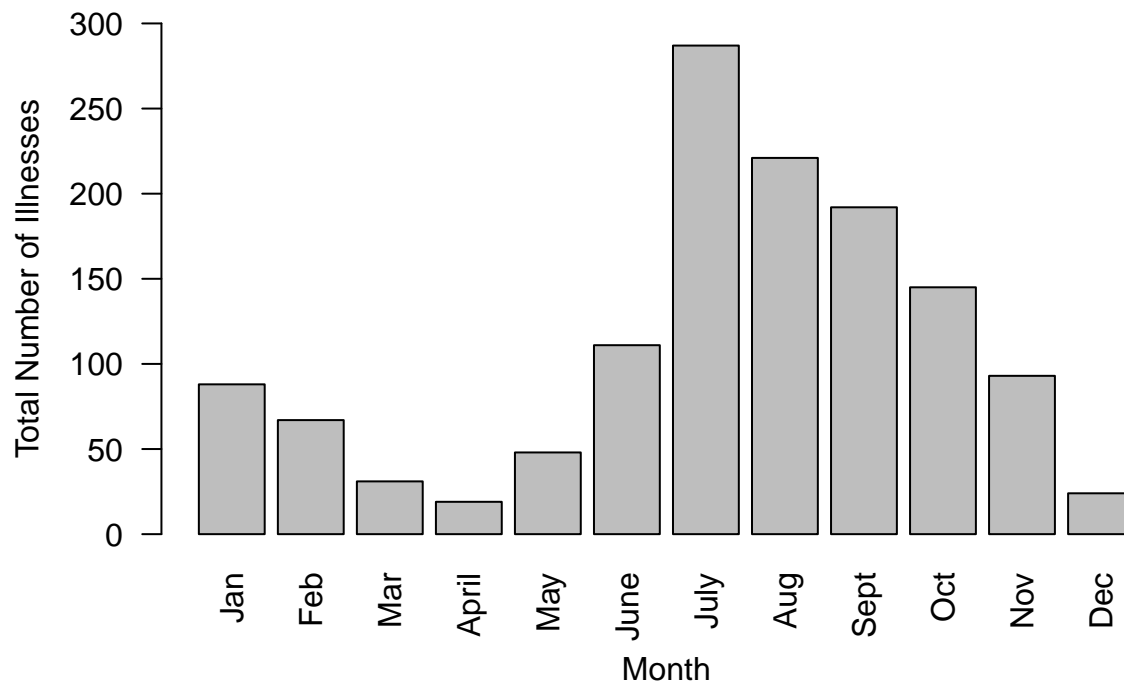


```
# Calculate the total number of illnesses per month
giardia_illnesses <- aggregate(Illnesses~Month, data= giardia, FUN = sum, na.rm = TRUE)

giardia_illnesses$Month <- months_ordered

barplot(Illnesses~Month, data=giardia_illnesses,
        xlab = "Month", ylab = "Total Number of Illnesses",
        main = "Giardia", ylim=c(0,300), las=2)
```

Giardia



Legionella

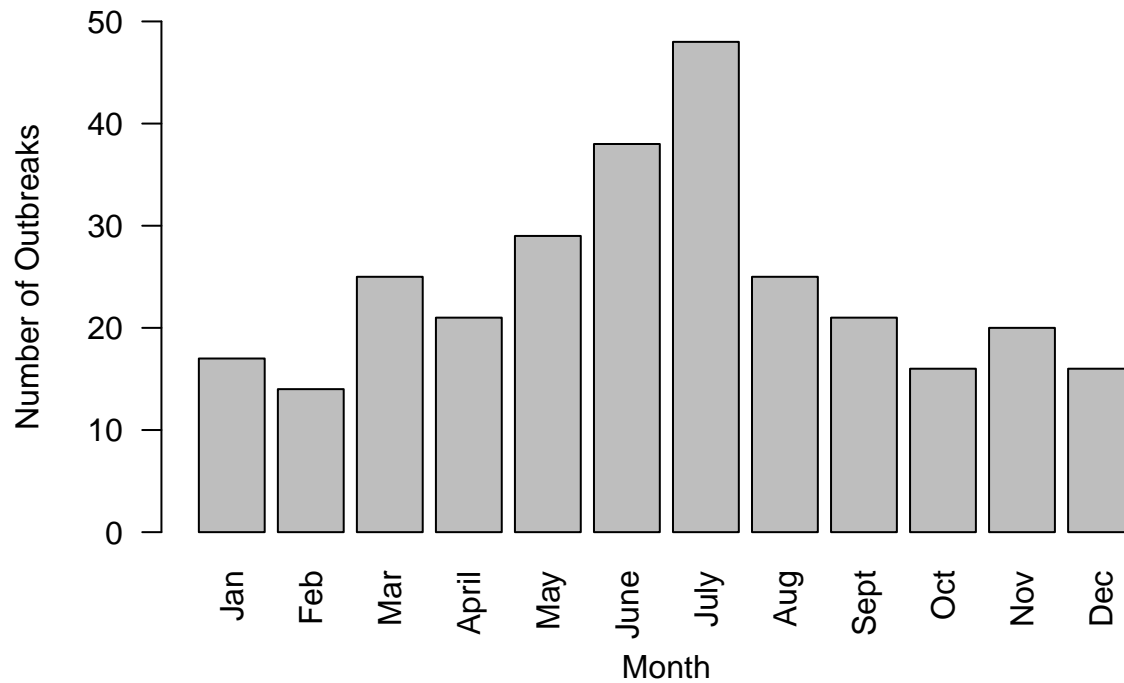
```
setwd("C:/Users/khnguy3/Documents/2020-TIEE/Data")
legionella <- read.csv(file = "Legionella.csv", header=TRUE)

# Calculate the number of outbreaks per month
legionella_outbreaks <- aggregate(legionella, by = list(legionella$Month), FUN = length)

legionella_outbreaks$Group.1 <- months_ordered

barplot(Month~Group.1, data= legionella_outbreaks,
        xlab = "Month", ylab = "Number of Outbreaks",
        main = "Legionella", ylim=c(0,50), las=2)
```

Legionella

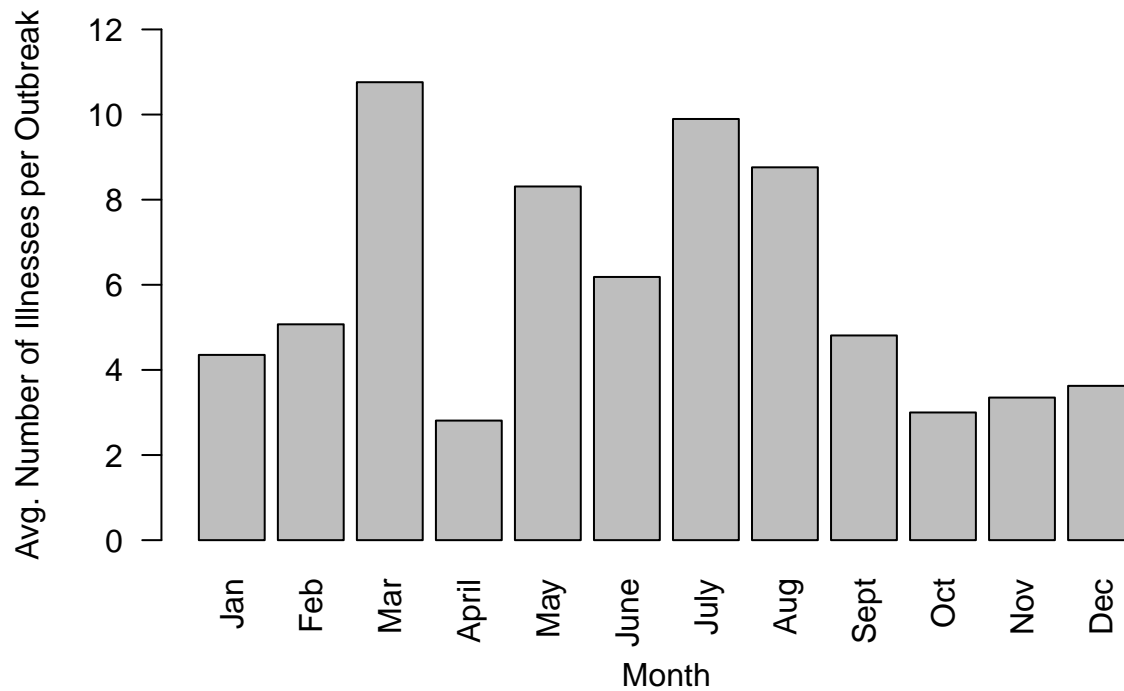


```
# Calculate the average size of outbreaks
legionella_size <- aggregate(Illnesses~Month, data= legionella, FUN = mean, na.rm = TRUE)

legionella_size$Month <- months_ordered

barplot(Illnesses~Month, data=legionella_size,
        xlab = "Month", ylab = "Avg. Number of Illnesses per Outbreak",
        main = "Legionella", ylim=c(0,12), las=2)
```

Legionella



```
# Calculate the total number of illnesses per month
legionella_illnesses <- aggregate(Illnesses~Month, data= legionella, FUN = sum, na.rm = TRUE)

legionella_illnesses$Month <- months_ordered

barplot(Illnesses~Month, data=legionella_illnesses,
        xlab = "Month", ylab = "Total Number of Illnesses",
        main = "Legionella", ylim=c(0,500), las=2)
```

Legionella

