

```
bats_data = read.csv('bats.csv')
head(bats_data, n=5)
```

```
##   Bat Gene.1 Gene.2 Gene.3 Gene.4 Gene.5 Ebola
## 1   1  FALSE  FALSE   TRUE   TRUE   TRUE  TRUE
## 2   2   TRUE  FALSE   TRUE  FALSE  FALSE FALSE
## 3   3   TRUE  FALSE   TRUE   TRUE   TRUE FALSE
## 4   4  FALSE   TRUE   TRUE   TRUE   TRUE  TRUE
## 5   5  FALSE  FALSE  FALSE   TRUE  FALSE FALSE
```

a. What is the chance of a random bat carrying the Ebola virus?

```
mean(bats_data['Ebola']==TRUE) * 100
```

```
## [1] 30.0793
```

b. For each gene, calculate the likelihood that it is expressed in a random bat.

```
gene_likelihoods = colMeans(bats_data[, 2:6])
gene_likelihoods * 100
```

```
##   Gene.1   Gene.2   Gene.3   Gene.4   Gene.5
## 70.22770 30.07630 50.08950 80.16180 32.70533
```

c. Is the presence or absence of any of the genes indicative of a random bat potentially carrying the Ebola virus?

```
ebola_true_rows = bats_data[bats_data$Ebola == TRUE, ]
colMeans(ebola_true_rows[, 2:6]) * 100
```

```
##   Gene.1   Gene.2   Gene.3   Gene.4   Gene.5
## 70.51431 30.21709 97.11759 98.74996 97.85565
```

```
ebola_false_rows = bats_data[bats_data$Ebola == FALSE, ]
colMeans(ebola_false_rows[, 2:6]) * 100
```

```
##   Gene.1   Gene.2   Gene.3   Gene.4   Gene.5
## 70.104405 30.015732 29.858410 72.165332 4.678204
```

Based on the above data, among all the rows that have ebola, 97% of them have presence of gene-5 in them and among the rows that do not have ebola, the presence of gene-5 is also quiet low that is 4%.

so, presence of gene-5 could be a indicator of Ebola