Assignment03

zerofrom

2024-10-07

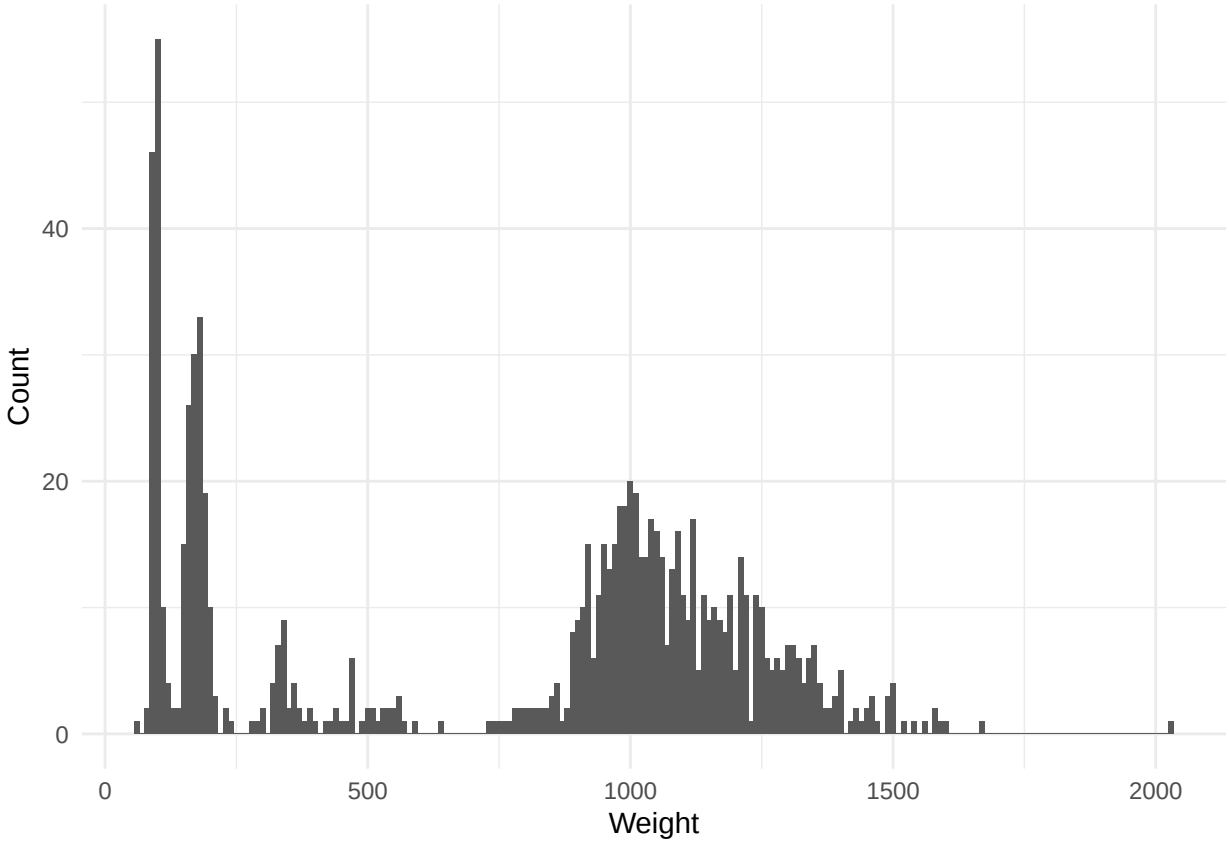
1. Data Wrangling

1 (Q1)

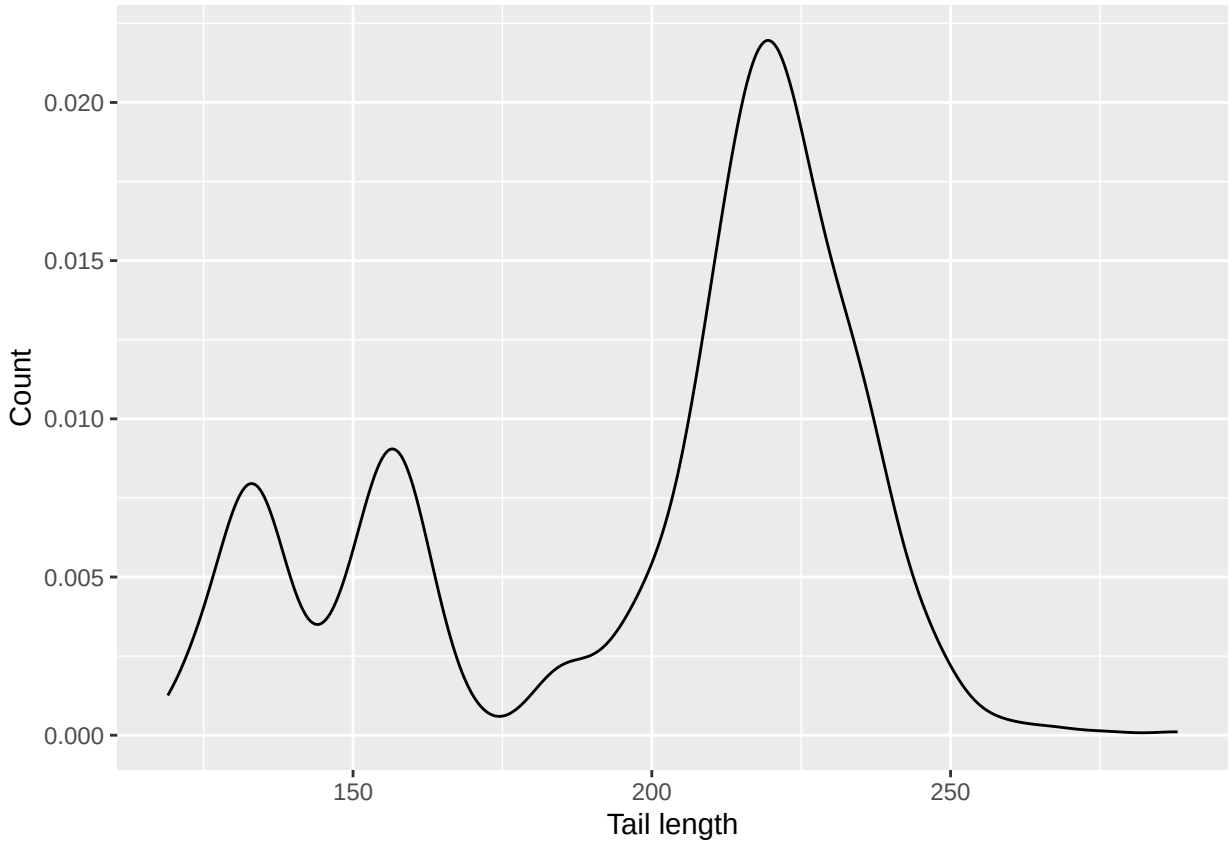
```
## [1] 897   9
```

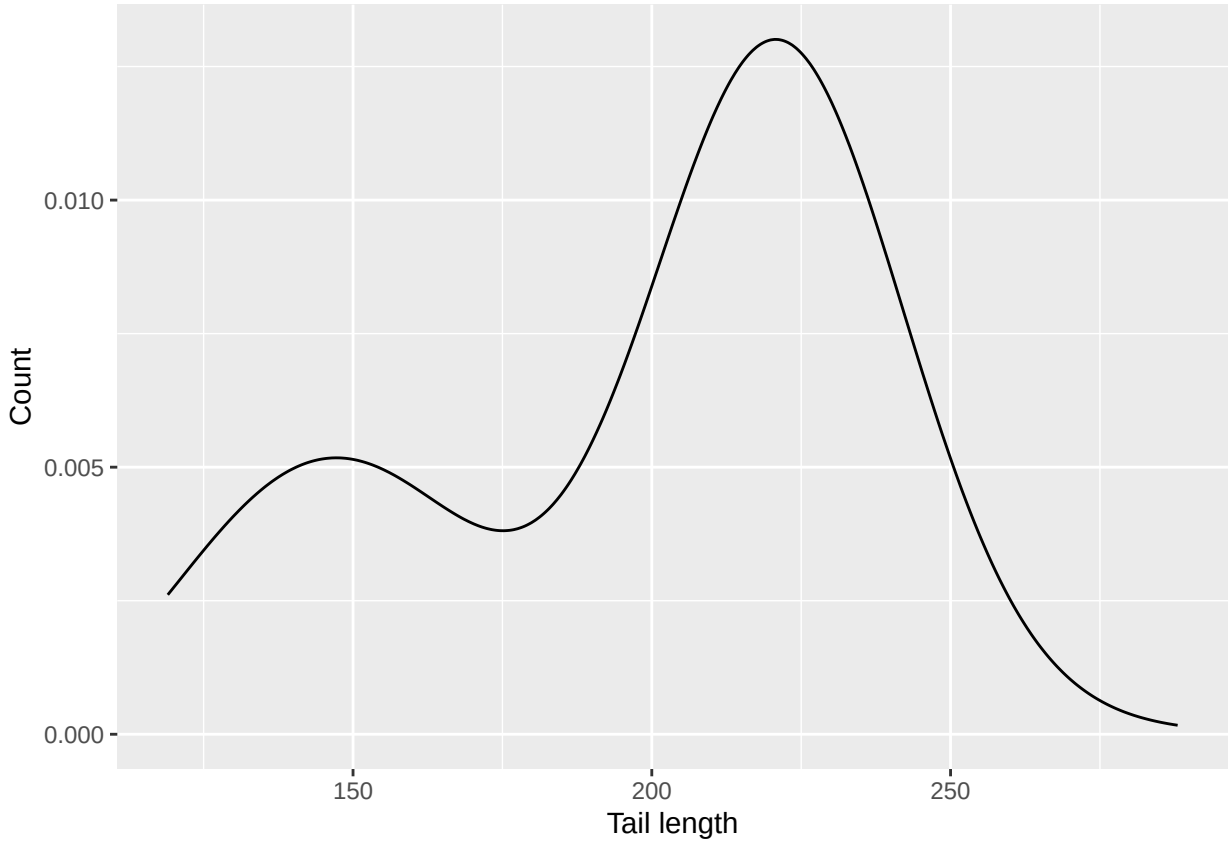
```
##   Age Day Month Year CaptureTime Species Wing Weight Tail
## 1   I  19     9 1992       13:30      RT  385    920  219
## 2   I  22     9 1992       10:30      RT  376    930  221
## 3   I  23     9 1992       12:45      RT  381    990  235
## 4   I  23     9 1992       10:50      CH  265    470  220
## 5   I  27     9 1992       11:15      SS  205    170  157
```

1 (Q2)

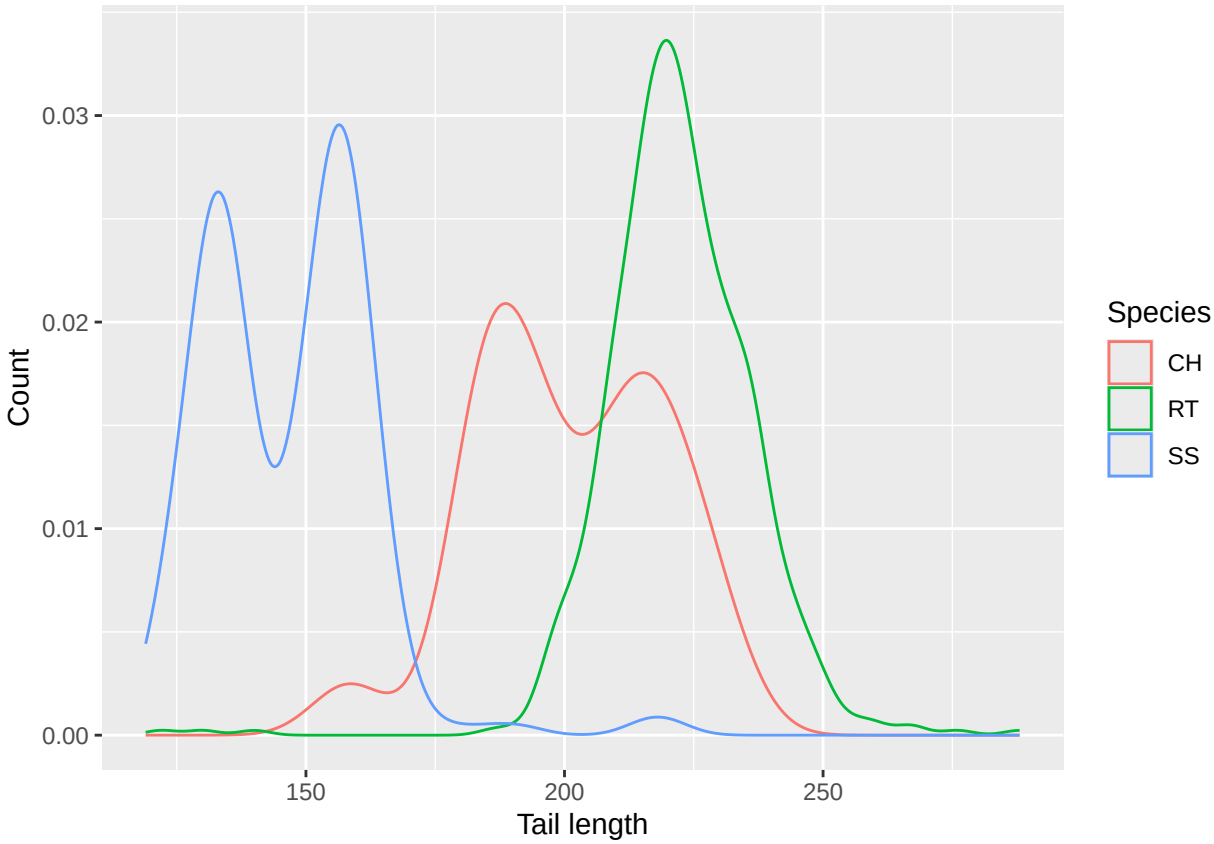


1 (Q3)

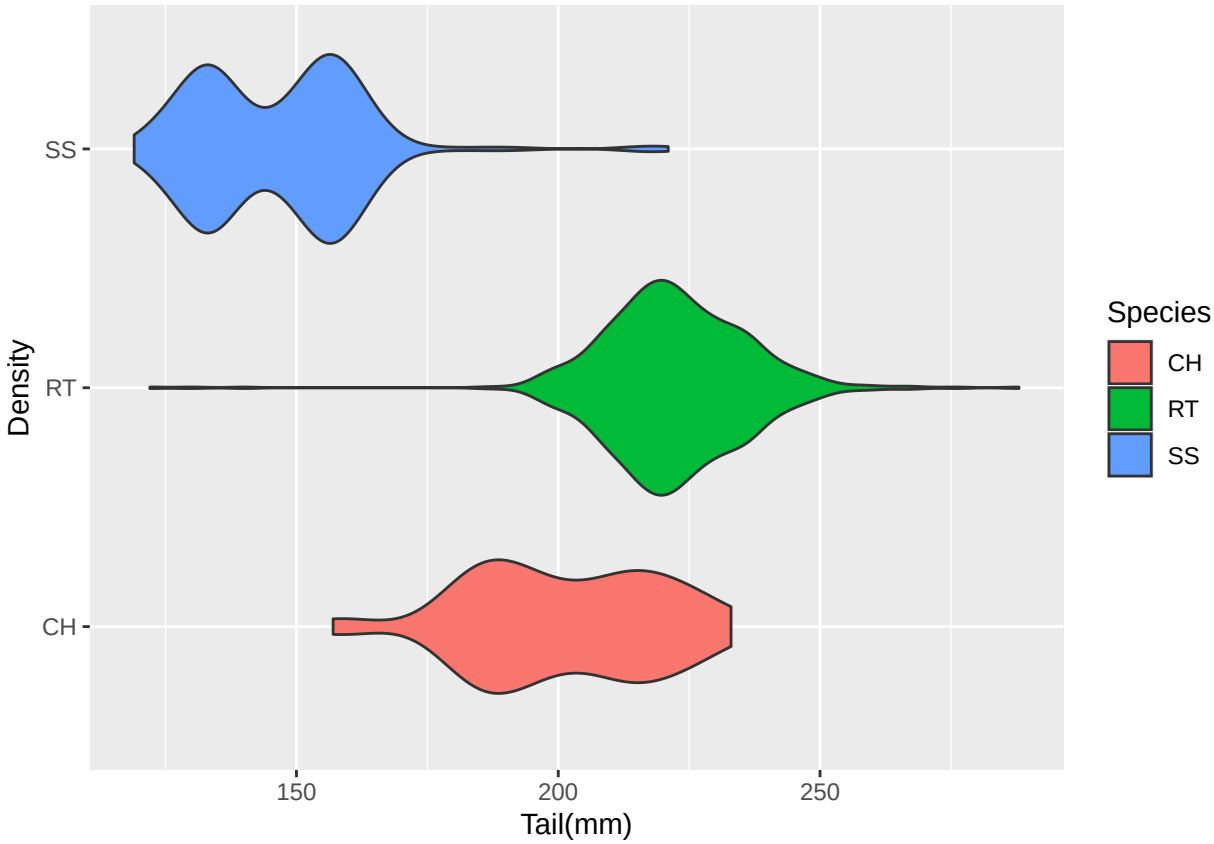

0.020
0.015
0.010
0.005
0.000
Count
150
200
250
Tail length


0.010
0.005
0.000
Count
150
200
250
Tail length

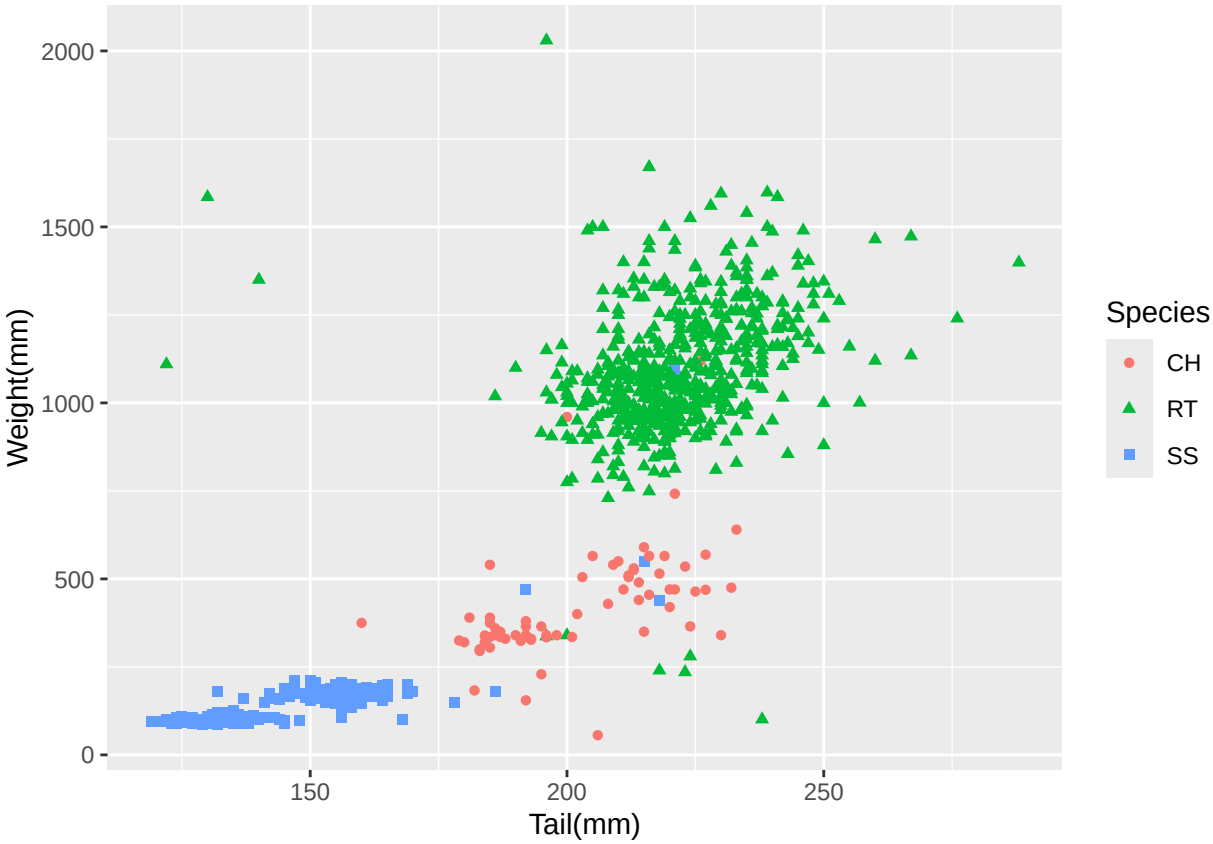
1 (Q4)



1 (Q5)



1 (Q6)



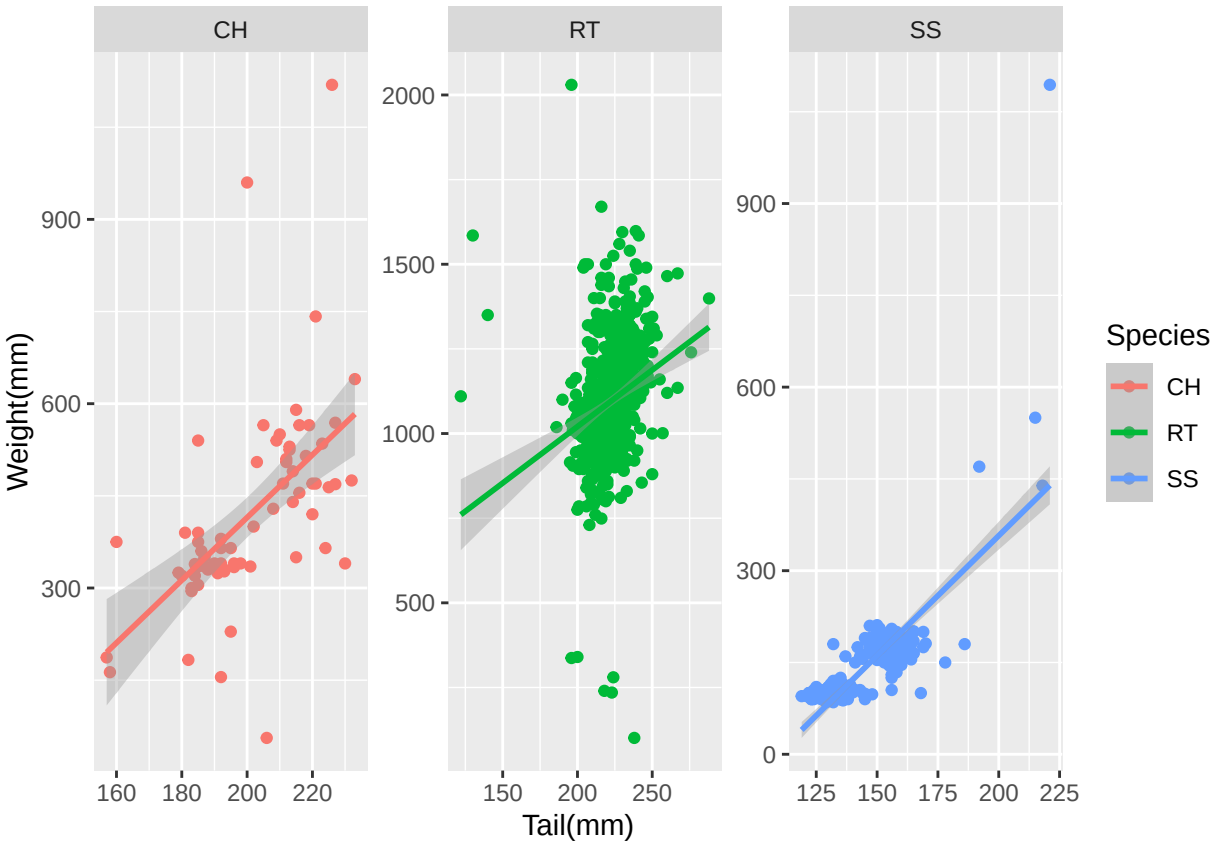
1. There are 4 aesthetics: x=Tail,y=Weight,color=Species,shape=Species.

2. Points.

3. Color: Different colors refers to different species.
   Shape: Different shapes refers to different species.
   Axes Labels: These provide context for what the x(Tail length) and y(Weight) axes represent.

1 (Q7)

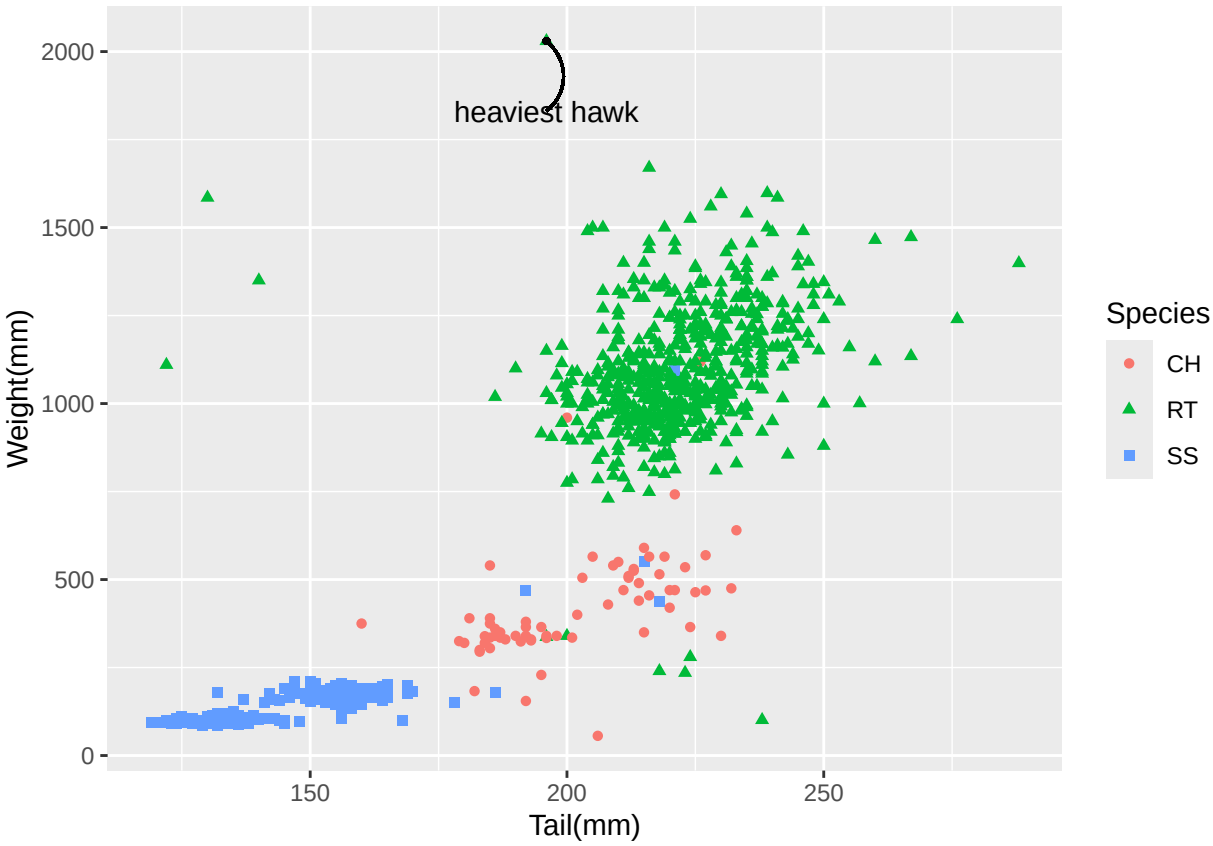
```
## `geom_smooth()` using formula = 'y ~ x'
```



1. Color: Used to differentiate between species.
   Line: A trend line indicates the relationship between Tail and Weight.
   3.The trend line shows a positive slope, which suggests positive correlation between the weight of the hawks and their tail lengths.

1 (Q8)

```
## Warning in geom_point(aes(x = heaviest_hawk$Tail, y = heaviest_hawk$Weight), : All aesthetics have le
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```



2. Finite probability spaces

- $n = 10$: number of balls.
- $r = 3$: number of red balls.
- $n - r = 7$: number of blue balls.
- $k = 22$: repeat times.
- $q = z$: red balls in repeat times.

$$P(X = z) = \binom{k}{q} \left(\frac{r}{n}\right)^{z} \left(\frac{n-r}{n}\right)^{k-z}$$

2 (Q1)
the probability that z out of the 22 selections were red spheres

$$P(X = z) = \binom{22}{z} \left(\frac{3}{10}\right)^{z} \left(\frac{7}{10}\right)^{22-z}$$

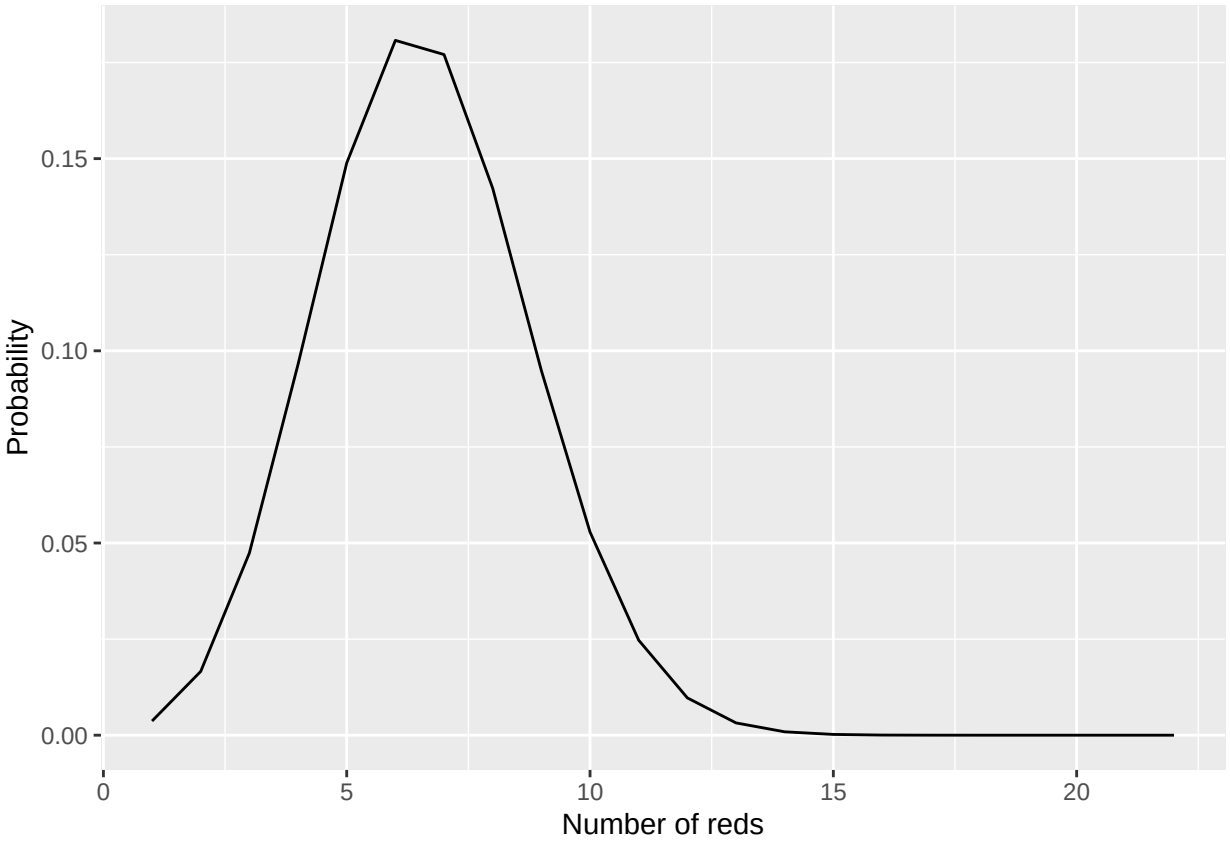
2 (Q2)

```
## [1] 0.05285129
```

2 (Q3)

```
##   num_reds        prob
## 1        1 0.003686403
## 2        2 0.016588812
## 3        3 0.047396606
```

2 (Q4)



2 (Q5)

2 (Q5.1)

```
## [1] 14 68 39  1 34
## [1] 87 43 14 82 59
## [1] 51 97 85 21 54
## [1] 74  7 73 79 85
## [1]  37  89 100  34  99
```

```
## [1] 68 39  1 34 87
```

```
## [1] 68 39  1 34 87
```

```
## [1] 68 39  1 34 87
```

```
## [1] 14 68 39  1 34
## [1] 87 43 14 82 59
## [1] 51 97 85 21 54
## [1] 74  7 73 79 85
## [1]  37  89 100  34  99
```

2 (Q5.2)

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

```
## [1] 2 4
##
## [[3]]
## [1] 3 9
```

```
## [1]  1  8 27
```

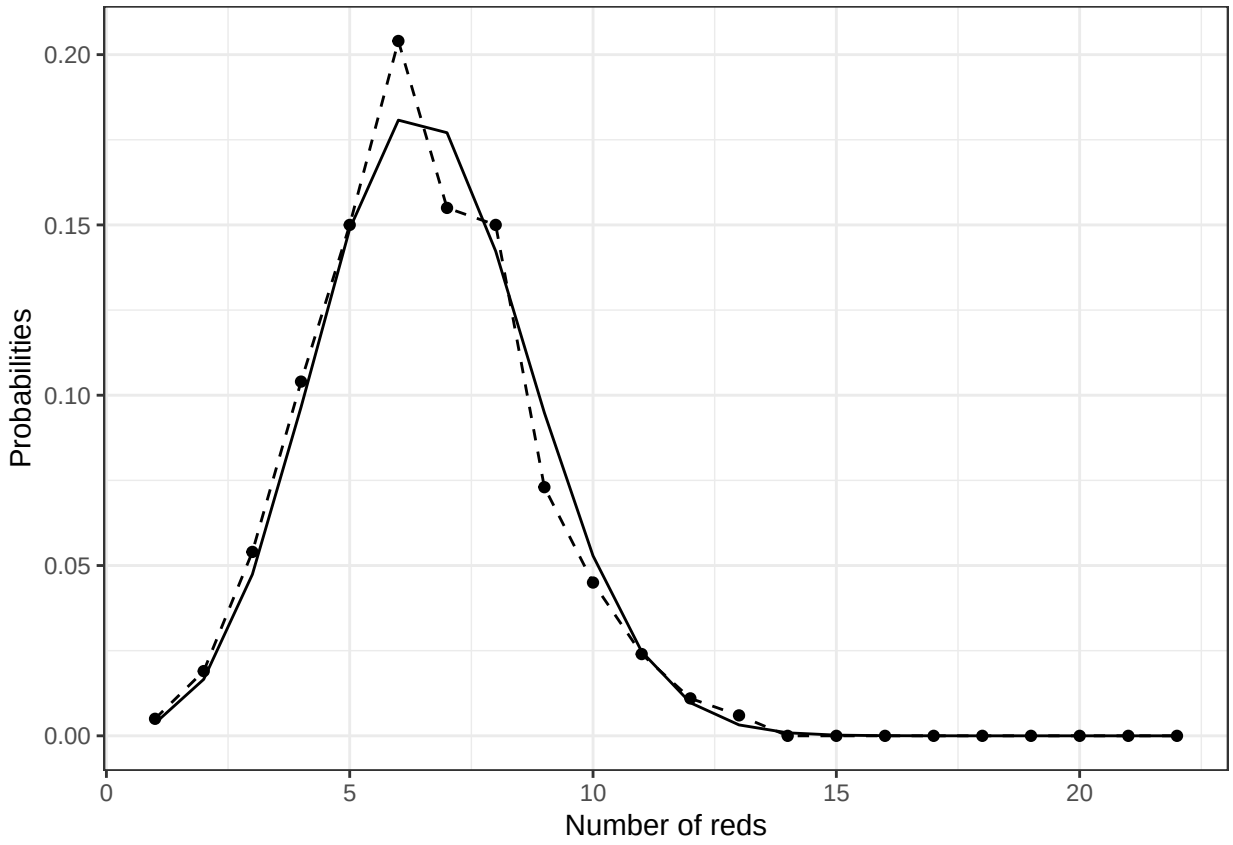
2 (Q5.3)

```
##   trial                                                          sample_balls
## 1     1   9, 4, 7, 1, 2, 7, 2, 3, 1, 5, 5, 10, 6, 10, 7, 9, 5, 5, 9, 9, 5, 5
## 2     2 2, 10, 9, 1, 4, 3, 6, 10, 10, 6, 4, 4, 10, 9, 7, 6, 9, 8, 9, 7, 8, 6
## 3     3   10, 7, 3, 10, 6, 8, 2, 2, 6, 6, 1, 3, 3, 8, 6, 7, 6, 8, 7, 1, 4, 8
##   num_reds
## 1        5
## 2        3
## 3        7
```

2 (Q6)

```
##   num_reds        prob predicted_prob
## 1        1 0.003686403          0.005
## 2        2 0.016588812          0.019
## 3        3 0.047396606          0.054
## 4        4 0.096485948          0.104
## 5        5 0.148864035          0.150
```

2 (Q7)



3

3 (Q1)

```
##   num_trail missing_proportion
## 1        10                0.1
## 2       100               0.14
## 3       500              0.128
## 4      1000              0.122
## 5      2000             0.1205
## 6      5000             0.1206
## 7     10000             0.1218
```

3 (Q2)

```
## [1] 0.1180318
```