

Assignment 2

Guanjie Liang

2024-09-27

1. Data Wrangling

1.1 (Q1)

```
hSF <- Hawks %>% filter(Species=="RT", Weight>=1000) %>% select("Wing", "Weight", "Tail")
print(head(hSF, 10))
```

```
##      Wing Weight Tail
## 1     412   1090   230
## 2     412   1210   210
## 3     405   1120   238
## 4     393   1010   222
## 5     371   1010   217
## 6     390   1120   213
## 7     416   1170   243
## 8     436   1390   232
## 9     418   1150   238
## 10    396   1010   227
```

1.1 (Q2)

there are 3 variables in hSF, show it how many columns it has

There are 4 examples `nrow()`, `dim()`, `length(rownames())`, `summary()` or `str()`

obervations: `print(nrow(hSF))` 398

`dim(hSF)` 398 3 `dim(hSF)[1]` 398 `dim(hSF)[2]` 3

`length(rownames(hSF))` 398

`summary(hSF)` overview of the data, `str(hSF)` all the data and the number of rows

1.2 (Q1)

```
sorthSF <- hSF %>% arrange(Wing)
print(sorthSF %>% head(10))
```

```
##      Wing Weight Tail
## 1    37.2    1180  210
## 2   111.0    1340  226
## 3   199.0    1290  222
## 4   241.0    1320  235
## 5   262.0    1020  200
## 6   277.0    1500  207
## 7   330.0    1000  220
## 8   331.0    1055  210
## 9   345.0    1000  200
## 10  350.0    1115  199
```

1.3 (Q1)

```
species_code <- c("CH", "RT", "SS")
species_name_full <- c("Cooper's", "Red-tailed", "Sharp-shinned")
hawkSpeciesNameCodes <- data.frame(species_code, species_name_full)
print(hawkSpeciesNameCodes)
```

```
##   species_code species_name_full
## 1           CH      Cooper's
## 2           RT      Red-tailed
## 3           SS      Sharp-shinned
```

1.3 (Q2)

```
hawksFullName <- hawkSpeciesNameCodes %>% rename(Species = species_code)
hawksFullName <- left_join(Hawks, hawksFullName) %>% select(-Species) %>% rename(Species = species_name_full)
```

```
## Joining with `by = join_by(Species)`
```

```
print(hawksFullName %>% head(5))
```

```
##   Month Day Year CaptureTime ReleaseTime BandNumber Age Sex Wing Weight Culmen
## 1     9  19 1992      13:30           877-76317   I    385   920   25.7
## 2     9  22 1992      10:30           877-76318   I    376   930    NA
## 3     9  23 1992      12:45           877-76319   I    381   990   26.7
## 4     9  23 1992      10:50           745-49508   I    F  265   470   18.7
## 5     9  27 1992      11:15          1253-98801   I    F  205   170   12.5
##   Hallux Tail StandardTail Tarsus WingPitFat KeelFat Crop      Species
## 1   30.1  219           NA     NA           NA     NA   NA   Red-tailed
## 2    NA  221           NA     NA           NA     NA   NA   Red-tailed
## 3   31.3  235           NA     NA           NA     NA   NA   Red-tailed
## 4   23.5  220           NA     NA           NA     NA   NA   Cooper's
## 5   14.3  157           NA     NA           NA     NA   NA   Sharp-shinned
```

1.3 (Q3)

```
print(hawksFullName %>% select(Species, Wing, Weight) %>% head(7))
```

```
##      Species Wing Weight
## 1   Red-tailed  385    920
## 2   Red-tailed  376    930
## 3   Red-tailed  381    990
## 4   Cooper's    265    470
## 5 Sharp-shinned 205    170
## 6   Red-tailed  412   1090
## 7   Red-tailed  370    960
```

1.4 (Q1)

```
hawksWithBMI <- Hawks %>% mutate(bird_BMI = 1000*Weight/Wing^2) %>% select(Species, bird_BMI) %
>% arrange(desc(bird_BMI))
print(hawksWithBMI %>% head(8))
```

```
##   Species bird_BMI
## 1      RT 852.69973
## 2      RT 108.75741
## 3      RT  32.57493
## 4      RT  22.72688
## 5      CH  22.40818
## 6      RT  19.54932
## 7      CH  15.21998
## 8      RT  14.85927
```

1.5 (Q1)

```
hawksFullNameSum <- hawksFullName %>% group_by(Species) %>% summarize(num_rows=n(),
                                                                    mn_wing=mean(Wing, na.rm=
TRUE),
                                                                    nd_wing=median(Wing, na.rm
=TRUE),
                                                                    t_mn_wing=mean(Wing, min
(0.1), na.rm=TRUE),
                                                                    b_wt_ratio=max(Wing/Tail,
na.rm=TRUE))
print(hawksFullNameSum)
```

```
## # A tibble: 3 × 6
##   Species      num_rows mn_wing nd_wing t_mn_wing b_wt_ratio
##   <chr>          <int>   <dbl>   <dbl>   <dbl>     <dbl>
## 1 Cooper's         70    244.    240    243.     1.67
## 2 Red-tailed      577    383.    384    385.     3.16
## 3 Sharp-shinned   261    185.    191    184.     1.67
```

1.5 (Q2)

```
hawksFullNameNA <- hawksFullName %>% group_by(Species) %>% summarize(Wing=sum(is.na(Wing)),
                                                                    Weight=sum(is.na(Weight)),
                                                                    Culmen=sum(is.na(Culmen)),
                                                                    Hallux=sum(is.na(Hallux)),
                                                                    Tail=sum(is.na(Tail)),
                                                                    StandardTail=sum(is.na(StandardTail)),
                                                                    Tarsus=sum(is.na(Tarsus)))

print(hawksFullNameNA)
```

```
## # A tibble: 3 × 8
##   Species      Wing Weight Culmen Hallux Tail StandardTail Tarsus
##   <chr>      <int>  <int>  <int>  <int> <int>      <int>  <int>
## 1 Cooper's      1     0     0     0     0         19     62
## 2 Red-tailed    0     5     4     3     0        250    538
## 3 Sharp-shinned 0     5     3     3     0         68    233
```

2. Random experiments, events and sample spaces, and the set theory

2.1 (Q1)

A Random experiment is a procedure (real or imagined) which: 1. has a well-defined set of possible outcomes 2. could (at least in principle) be repeated arbitrary many times

An event is a set of possible outcomes of an experiments An event is any subset of the sample space, including the empty set and the sample space itself.

A sample space is the set of all possible outcomes of interest for a random experiment

2.1 (Q2)

event: {1,2}

sample space: {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),...,(6,6)}

total number of different events: $2^{6 \times 6} = 2^{36}$

Yes, the empty set is considered an event. It represents the impossible event

2.2 (Q1)

1. $A \cup B = \{1, 2, 3, 4, 6\}$ $A \cup B = \{1, 2, 3, 4, 5, 6\}$

2. $A \cap B = \{2\}$ $A \cap B = \{\}$

3. $A \setminus B = \{1, 3\}$ $A \setminus B = \{1, 2, 3\}$

4. A and B are not disjoint, A and C are disjoint

5. Yes, B and $A \setminus B$ are disjoint

6. two sets: $\{\{1, 2, 3\}, \{4, 5, 6\}\}$

three sets: $\{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$

2.2 (Q2)

1. A

2. empty

3. $A^c = \Omega \setminus A, B^c = \Omega \setminus B$ because $A \subseteq B$, then $B^c \subseteq A^c$

4. $\cup_{k=1}^K A_k^c$

5. $(A \cup B)^c = \Omega \setminus (A \cup B) = (\Omega \setminus A) \cap (\Omega \setminus B) = A^c \cap B^c$

6. $\cap_{k=1}^K A_k^c$

2.2 (Q3)

$$|E| = 2^K$$

2.2 (Q4)

1. empty set: \emptyset

2. $S_1 \cup S_2 \cup S_3 \cup S_4 = A_1 \cup A_2 \cup A_3 \cup A_4$

$$S_1 \cap S_2 = \emptyset \quad S_2 \cap S_3 = \emptyset$$

$$S_1 \cap S_3 = \emptyset \quad S_2 \cap S_4 = \emptyset$$

$$S_1 \cap S_4 = \emptyset \quad S_3 \cap S_4 = \emptyset$$

So S_1, S_2, S_3, S_4 form a partition of $A_1 \cup A_2 \cup A_3 \cup A_4$

2.2 (Q5)

1. $1_{A^c}(w) = 1 - 1_A(w)$

2. Ω

3.

Step 1: $1_{(A \cap B)^c} = 1 - 1_{(A \cap B)} = 1 - 1_A \cdot 1_B$

Step 2: $1_{A^c \cup B^c} = 1_{A^c} + 1_{B^c} - 1_{A^c} \cdot 1_{B^c} = (1 - 1_A) + (1 - 1_B) - (1 - 1_A)(1 - 1_B) = 1 - 1_A \cdot 1_B$

So $(A \cap B)^c = A^c \cup B^c$

2.2 (Q6)

the real number between 0 and 1 is infinite

3. Probability theory

3.1 (Q1)

$$P(x) = \begin{cases} 0, & A = \emptyset \\ 0.5, & A = a \\ 0.1, & A = b \\ 0.4, & A = c \\ 0.6, & A = a, b \\ 0.9, & A = a, c \\ 0.5, & A = b, c \\ 1, & A = a, b, c \end{cases}$$

3.1 (Q2)

1. $\mathbb{P}(A) \geq 0$ for any event A

$$\mathbb{P}(\emptyset) = 0, \mathbb{P}(0) = 1 - q, \mathbb{P}(0, 1) = 1$$

$$2. \mathbb{P}(\Omega) = \mathbb{P}(0, 1) = 1$$

$$3. \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \mathbb{P}(0) + \mathbb{P}(1) = 1$$

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \mathbb{P}(0) + \mathbb{P}(1) = 1$$

3.2 (Q1)

$$\mathbb{P}(\cup_{i=1}^n A_i) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$$

$$\sum_{i=1}^n \mathbb{P}(A_i) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$$

3.2 (Q2)

$$\mathbb{P}(S) \cup \mathbb{P}(S^c) = \Omega, \quad \mathbb{P}(S) \cap \mathbb{P}(S^c) = \emptyset$$

$$\text{So } \mathbb{P}(S^c) = 1 - \mathbb{P}(S)$$

3.2 (Q3)

$$S_1 = \{1, 2\}, S_2 = \{2, 3\}, S_3 = \{3, 4\}$$

3.2 (Q4)

Draw a diagram