MATH2270 Assignment 2

Visualising Open Data

Student Details

- Alistair Grevis-James (s3644119)

https://www.kaggle.com/zhangjuefei/birds-bones-and-living-habits (https://www.kaggle.com/zhangjuefei/birds-bones-and-living-habits)

About

According to their living environments and living habits, birds are classified into different ecological groups. There are 8 ecological groups of birds:

- Swimming Birds
- Wading Birds
- Terrestrial Birds
- Raptors
- Scansorial Birds
- Singing Birds
- Cursorial Birds (not included in dataset)
- Marine Birds (not included in dataset)

First 6 groups are main and are covered by this dataset.

There are 420 birds contained in this dataset. Each bird is represented by 10 measurements (features):

- Length and Diameter of Humerus
- Length and Diameter of Ulna
- Length and Diameter of Femur
- Length and Diameter of Tibiotarsus
- Length and Diameter of Tarsometatarsus

```
# Load your packages
library(tidyr)
library(dplyr)
library(knitr)
library(GGally)
library(ggplot2)
library(ggridges)
```

Data

```
# Load your data and prepare for visualisation
birdDS <- read.csv("bird.csv")
```

Exploring the Data

```
# Load your data and prepare for visualisation
head(birdDS)
```

	id <int>	huml <dbl>	humw <dbl>	ulnal <dbl>	ulnaw <dbl>	feml <dbl>	femw <dbl>	tibl <dbl>	tibw <dbl>
1	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03
2	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51
3	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04
4	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40
5	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96
6	5	61.92	4.78	50.46	3.47	49.52	4.41	56.95	2.73

6 rows | 1-10 of 13 columns

```
colnames(birdDS)
```

```
##  [1] "id"    "huml"  "humw"  "ulnal" "ulnaw" "feml"  "femw"  "tibl"
##  [9] "tibw"  "tarl"  "tarw"  "type"
```

```
colNamesFull <- c("ID", "HumerusLength", "HumerusDiameter", "UlnaLength", "UlnaDiamet
er",
                  "FemurLength", "FemurDiameter", "TibiotarusLength", "TibiotarusDiam
eter",
                  "TarsometatarusLength", "TarsometatarusDiameter", "EcologicalBirdTy
pe")
```

```
# Load your data and prepare for visualisation
summary(birdDS)
```

```
##        id              huml             humw            ulnal      
##  Min.   :  0.0   Min.   :  9.85   Min.   : 1.140   Min.   : 14.09  
##  1st Qu.:104.8   1st Qu.: 25.17   1st Qu.: 2.190   1st Qu.: 28.05  
##  Median :209.5   Median : 44.18   Median : 3.500   Median : 43.71  
##  Mean   :209.5   Mean   : 64.65   Mean   : 4.371   Mean   : 69.12  
##  3rd Qu.:314.2   3rd Qu.: 90.31   3rd Qu.: 5.810   3rd Qu.: 97.52  
##  Max.   :419.0   Max.   :420.00   Max.   :17.840   Max.   :422.00  
##                  NA's   :1        NA's   :1        NA's   :3       
##      ulnaw             feml             femw             tibl       
##  Min.   : 1.000   Min.   : 11.83   Min.   : 0.930   Min.   :  5.50  
##  1st Qu.: 1.870   1st Qu.: 21.30   1st Qu.: 1.715   1st Qu.: 36.42  
##  Median : 2.945   Median : 31.13   Median : 2.520   Median : 52.12  
##  Mean   : 3.597   Mean   : 36.87   Mean   : 3.221   Mean   : 64.66  
##  3rd Qu.: 4.770   3rd Qu.: 47.12   3rd Qu.: 4.135   3rd Qu.: 82.87  
##  Max.   :12.000   Max.   :117.07   Max.   :11.640   Max.   :240.00  
##  NA's   :2        NA's   :2        NA's   :1        NA's   :2       
##       tibw             tarl             tarw        type   
##  Min.   : 0.870   Min.   :  7.77   Min.   : 0.660   P : 38  
##  1st Qu.: 1.565   1st Qu.: 23.04   1st Qu.: 1.425   R : 50  
##  Median : 2.490   Median : 31.74   Median : 2.230   SO:128  
##  Mean   : 3.182   Mean   : 39.23   Mean   : 2.930   SW:116  
##  3rd Qu.: 4.255   3rd Qu.: 50.25   3rd Qu.: 3.500   T : 23  
##  Max.   :11.030   Max.   :175.00   Max.   :14.090   W : 65  
##  NA's   :1        NA's   :1        NA's   :1
```

```
# Load your data and prepare for visualisation
str(birdDS)
```

```
## 'data.frame':    420 obs. of  12 variables:
##  $ id   : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ huml : num  80.8 88.9 80 77.7 62.8 ...
##  $ humw : num  6.68 6.63 6.37 5.7 4.84 ...
##  $ ulnal: num  72 80.5 69.3 65.8 52.1 ...
##  $ ulnaw: num  4.88 5.59 5.28 4.77 3.73 3.47 4.5 4.55 6.13 7.05 ...
##  $ feml : num  41.8 47 43.1 40 34 ...
##  $ femw : num  3.7 4.3 3.9 3.52 2.72 4.41 3.41 3.78 5.45 7.44 ...
##  $ tibl : num  5.5 80.2 75.3 69.2 56.3 ...
##  $ tibw : num  4.03 4.51 4.04 3.4 2.96 2.73 3.56 3.81 5.58 7.31 ...
##  $ tarl : num  38.7 41.5 38.3 35.8 31.9 ...
##  $ tarw : num  3.84 4.01 3.34 3.41 3.13 2.83 3.64 3.81 4.37 6.34 ...
##  $ type : Factor w/ 6 levels "P","R","SO","SW",..: 4 4 4 4 4 4 4 4 4 4 ...
```

```
# Load your data and prepare for visualisation
head(birdDS)
```

	id <int>	huml <dbl>	humw <dbl>	ulnal <dbl>	ulnaw <dbl>	feml <dbl>	femw <dbl>	tibl <dbl>	tibw <dbl>
1	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03
2	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51
3	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04
4	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
5	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96
6	5	61.92	4.78	50.46	3.47	49.52	4.41	56.95	2.73

6 rows | 1-10 of 13 columns

```
# Check the species ID
unique(birdDS$type)
```

```
## [1] SW W  T  R  P  SO
## Levels: P R SO SW T W
```

Determine the mean values

```
# SW - Swimming Birds
SWbirds <- subset(birdDS, birdDS$type == "SW") # SW - Swimming Birds
Wbirds <- subset(birdDS, birdDS$type == "W") # W - Wading Birds
Tbirds <- subset(birdDS, birdDS$type == "T") # T - Terrestrial Birds
Rbirds <- subset(birdDS, birdDS$type == "R") # R - Raptors
Pbirds <- subset(birdDS, birdDS$type == "P") # P - Scansorial Birds
SObirds <- subset(birdDS, birdDS$type == "SO") # SO - Singing Birds
```

```
# Filling NA - Rbirds
Wbirds$feml[is.na(Wbirds$feml)] <- mean(Wbirds$feml, na.rm = TRUE)
Wbirds$femw[is.na(Wbirds$femw)] <- mean(Wbirds$femw, na.rm = TRUE)
```

```
# Filling NA - Rbirds
Rbirds$ulnal[is.na(Rbirds$ulnal)] <- mean(Rbirds$ulnal, na.rm = TRUE)
Rbirds$ulnaw[is.na(Rbirds$ulnaw)] <- mean(Rbirds$ulnaw, na.rm = TRUE)
Rbirds$tarl[is.na(Rbirds$tarl)] <- mean(Rbirds$tarl, na.rm = TRUE)
Rbirds$tarw[is.na(Rbirds$tarw)] <- mean(Rbirds$tarw, na.rm = TRUE)
```

```
# Filling NA - SObirds
SObirds$huml[is.na(SObirds$huml)] <- mean(SObirds$huml, na.rm = TRUE)
SObirds$humw[is.na(SObirds$humw)] <- mean(SObirds$humw, na.rm = TRUE)
SObirds$ulnal[is.na(SObirds$ulnal)] <- mean(SObirds$ulnal, na.rm = TRUE)
SObirds$ulnaw[is.na(SObirds$ulnaw)] <- mean(SObirds$ulnaw, na.rm = TRUE)
SObirds$feml[is.na(SObirds$feml)] <- mean(SObirds$feml, na.rm = TRUE)
SObirds$tibl[is.na(SObirds$tibl)] <- mean(SObirds$tibl, na.rm = TRUE)
SObirds$tibw[is.na(SObirds$tibw)] <- mean(SObirds$tibw, na.rm = TRUE)
```

```
birdDS_clean <- rbind(SWbirds, Wbirds, Tbirds, Rbirds, Pbirds, SObirds)
```

```
summary(birdDS_clean)
```

```
##        id              huml             humw            ulnal      
##  Min.   :  0.0   Min.   :  9.85   Min.   : 1.140   Min.   : 14.09  
##  1st Qu.:104.8   1st Qu.: 25.04   1st Qu.: 2.188   1st Qu.: 28.00  
##  Median :209.5   Median : 44.08   Median : 3.495   Median : 43.51  
##  Mean   :209.5   Mean   : 64.55   Mean   : 4.365   Mean   : 68.99  
##  3rd Qu.:314.2   3rd Qu.: 90.22   3rd Qu.: 5.805   3rd Qu.: 97.56  
##  Max.   :419.0   Max.   :420.00   Max.   :17.840   Max.   :422.00  
##      ulnaw             feml             femw             tibl       
##  Min.   : 1.000   Min.   : 11.83   Min.   : 0.930   Min.   :  5.50  
##  1st Qu.: 1.867   1st Qu.: 21.33   1st Qu.: 1.718   1st Qu.: 36.34  
##  Median : 2.945   Median : 31.13   Median : 2.525   Median : 51.94  
##  Mean   : 3.596   Mean   : 36.84   Mean   : 3.221   Mean   : 64.53  
##  3rd Qu.: 4.772   3rd Qu.: 47.10   3rd Qu.: 4.122   3rd Qu.: 82.87  
##  Max.   :12.000   Max.   :117.07   Max.   :11.640   Max.   :240.00  
##       tibw             tarl             tarw         type    
##  Min.   : 0.870   Min.   :  7.77   Min.   : 0.660   P : 38  
##  1st Qu.: 1.560   1st Qu.: 23.04   1st Qu.: 1.427   R : 50  
##  Median : 2.490   Median : 31.81   Median : 2.235   SO:128  
##  Mean   : 3.178   Mean   : 39.28   Mean   : 2.935   SW:116  
##  3rd Qu.: 4.253   3rd Qu.: 50.46   3rd Qu.: 3.522   T : 23  
##  Max.   :11.030   Max.   :175.00   Max.   :14.090   W : 65
```

Visualisation

```
colnames(birdDS_clean)
```

```
##  [1] "id"    "huml"  "humw"  "ulnal" "ulnaw" "feml"  "femw"  "tibl"
##  [9] "tibw"  "tarl"  "tarw"  "type"
```

```
allMeasurements <- c("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",  "tibl",  "ti
bw",  "tarl",  "tarw")
lengthMeasurements <- c("huml", "ulnal", "feml",  "tibl",  "tarl")
widthMeasurements <- c("humw", "ulnaw",  "femw", "tibw", "tarw")
```

```
test_reorder <- birdDS_clean[,c(12,2,3,4,5,6,7,8,9,10,11,1)]
```

```
test <- birdDS_clean %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
"tibl",  "tibw",  "tarl",  "tarw",
                  key = "id", value = "measurement")
```

```
SWbirdsLong <- SWbirds %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
 "tibl",  "tibw",  "tarl",  "tarw",
                  key = "type", value = "measurement")
SWbirdsLong$id <- 'SW'
WbirdsLong <- Wbirds %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
"tibl",  "tibw",  "tarl",  "tarw",
                  key = "type", value = "measurement")
WbirdsLong$id <- 'W'
TbirdsLong <- Tbirds %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
"tibl",  "tibw",  "tarl",  "tarw",
                  key = "type", value = "measurement")
TbirdsLong$id <- 'T'
RbirdsLong <- Rbirds %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
"tibl",  "tibw",  "tarl",  "tarw",
                  key = "type", value = "measurement")
RbirdsLong$id <- 'R'
PbirdsLong <- Pbirds %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
"tibl",  "tibw",  "tarl",  "tarw",
                  key = "type", value = "measurement")
PbirdsLong$id <- 'P'
SObirdsLong <- SObirds %>% gather("huml", "humw", "ulnal", "ulnaw", "feml",  "femw",
 "tibl",  "tibw",  "tarl",  "tarw",
                  key = "type", value = "measurement")
SObirdsLong$id <- 'SO'
```

```
birdDS_Long <- rbind(SWbirdsLong, WbirdsLong, TbirdsLong, RbirdsLong, PbirdsLong, SOb
irdsLong)
```

```
summary(birdDS_Long)
```

```
##       id                type            measurement
##  Length:4200        Length:4200        Min.   :  0.66
##  Class :character   Class :character   1st Qu.:  2.67
##  Mode  :character   Mode  :character   Median : 12.47
##                                        Mean   : 29.15
##                                        3rd Qu.: 38.50
##                                        Max.   :422.00
```

```
colnames(birdDS_Long) <- c("birdType", "bone", "measurement")
```

```
birdDS_Long$measurementType <- sapply(strsplit(as.character(birdDS_Long$bone), ""), t
ail, 1)
birdDS_Long$boneAlone = substr(birdDS_Long$bone,1,nchar(birdDS_Long$bone)-1)
```

```
birdDS_Long$boneAlone <- factor(birdDS_Long$boneAlone)
levels(birdDS_Long$boneAlone) <- c("Femur", "Humerus", "Tarsometatarus", "Tibiotarsu
s", "Ulna")
levels(birdDS_Long$boneAlone)
```

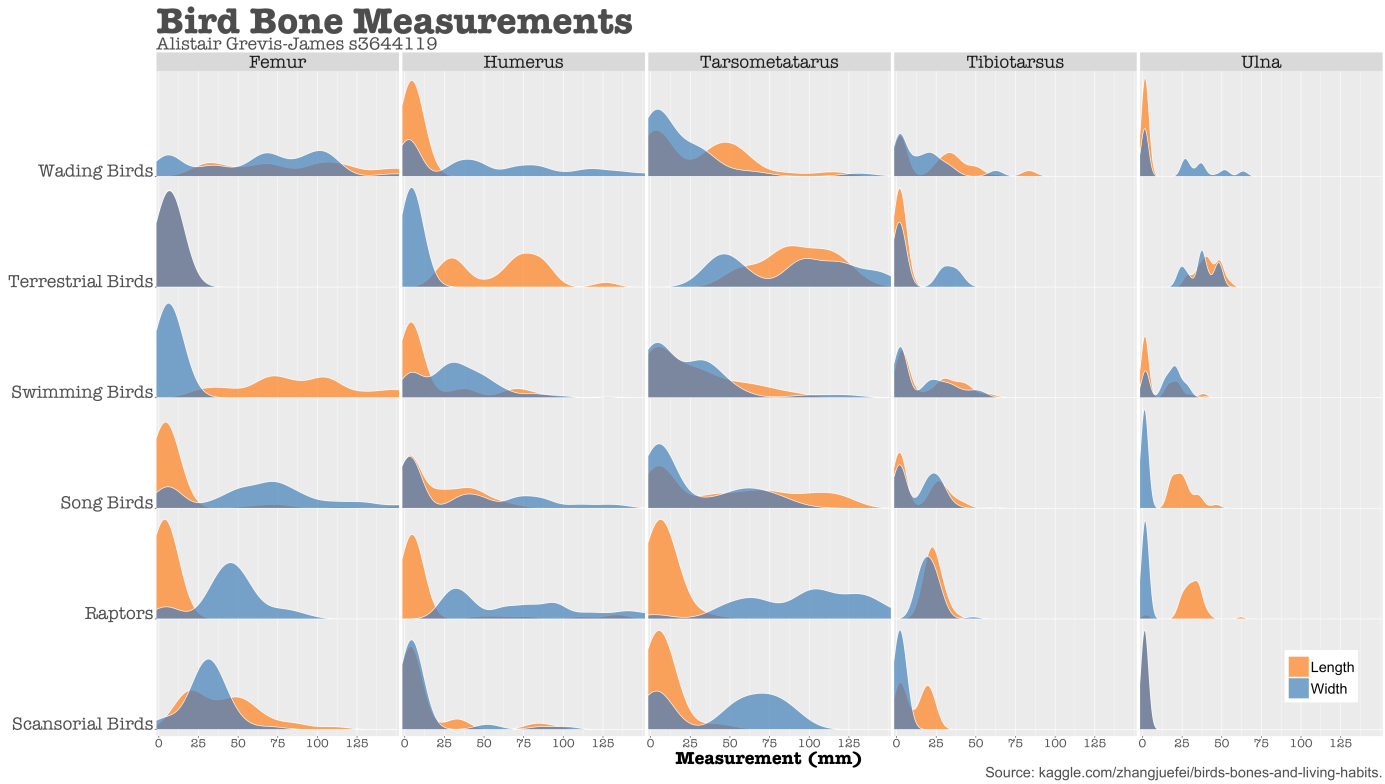
```
## [1] "Femur"          "Humerus"        "Tarsometatarus" "Tibiotarsus"
## [5] "Ulna"
```

```
png('birds.png', units="in", width=32, height=18, res=300)
birdsPlot <- ggplot(data = birdDS_Long,
                    aes(x=birdDS_Long$measurement,
                        y = birdType,
                        fill = measurementType))
birdsPlot <- birdsPlot + geom_density_ridges2(scale = 0.9, alpha=0.7, colour = 'whit
e', size = 0.5) +
  labs(y = NULL, x = "Measurement (mm)",
       title = "Bird Bone Measurements",
       subtitle = "Alistair Grevis-James s3644119",
       caption="Source: kaggle.com/zhangjuefei/birds-bones-and-living-habits.") +
  theme(plot.title=element_text(size=62,
                                face="bold",
                                family="American Typewriter",
                                color="grey30"),
        plot.subtitle=element_text(size=30,
                                family="American Typewriter",
                                color="grey30"),
        plot.caption = element_text(size=24,
                               color="grey30"),
        axis.title.x=element_text(vjust=0,
                              face = "bold",
                              family="American Typewriter",
                              size=30),
        axis.text.y = element_text(vjust = 0,
                                family="American Typewriter",
                                size = 30),
        axis.text.x = element_text(size = 20,
                                family="American Typewriter"),
        strip.text.x = element_text(size = 30,
                                family="American Typewriter"),
        legend.text = element_text(size=24)) +
  guides(fill = guide_legend(keywidth = 2.5, keyheight = 2.5)) +
  scale_x_continuous(expand = c(0.01, 0), breaks = c(0, 25, 50, 75, 100, 125)) +
  coord_cartesian(xlim = c(0, 150)) +
  scale_y_discrete(expand = c(0.01, 0), labels=c("Scansorial Birds", "Raptors",
                                                 "Song Birds", "Swimming Birds",
                                                  "Terrestrial Birds", "Wading Bird
s")) +
  facet_grid(.~boneAlone) +
  theme(legend.title = element_blank()) +
  theme(legend.position = c(0.95, 0.09)) +
  scale_fill_manual(values = c("#ff7f00", "#377eb8"), labels=c("Length", "Width"))

birdsPlot
dev.off()
```

```
## quartz_off_screen
##                 2
```

```
birdsPlot
```



The figure shows a series of faceted density plots of bird bone measurements. The measurements are of either length in orange or diameter (width) in blue – all measurements are in mm. The plots are faceted by bone type, and each row represents a type of bird. The visualisation is highly effective at showing the relation between length and diameter of a particular bone, within a particular bird type. The visualisation also allows for distribution comparisons between bird groups and for distribution comparisons between bone types. Starting in the top left-hand corner, we see for example that femur measurements of wading birds are very widely distributed. If we look to terrestrial birds, we see the femur length and width are extremely highly correlated. This correlation can then be compared with the other bones measured.