

```
#Copyright (c) 2018 Jouke Profijt.  
#Licensed under GPLv3. See gpl.md
```

```
BirdBones <- read.csv("../data/bird.csv",header = T, sep = ",")
```

Introduction

Research Question

What bone or group of bones that most birds have in common, is the most significant for the function in the different ecological groups?

Data

Data recieved from:

Birds' Bones and Living Habits, Kaggle dataset

Bone measurements were measured from a skeleton collection of Natural History Museum of Los Angeles County provided by Dr. D. Liu of Beijing Museum of Natural History

Exploratory Data Analyses

The data contains 420 bird samples where the bone lengths and diameters have been measured. The birds are separated in 6 different groups:

- Swimming Birds, SW
- Wading Birds, W
- Terrestrial Birds, T
- Raptors, R
- Scansorial Birds, P
- Singing Birds, SO

Most samples have data for:

- Length and Diameter of the Humerus
- Length and Diameter of the Ulna
- Length and Diameter of the Femur
- Length and Diameter of the Tibiotarsus
- Length and Diameter of the Mesometatarsus

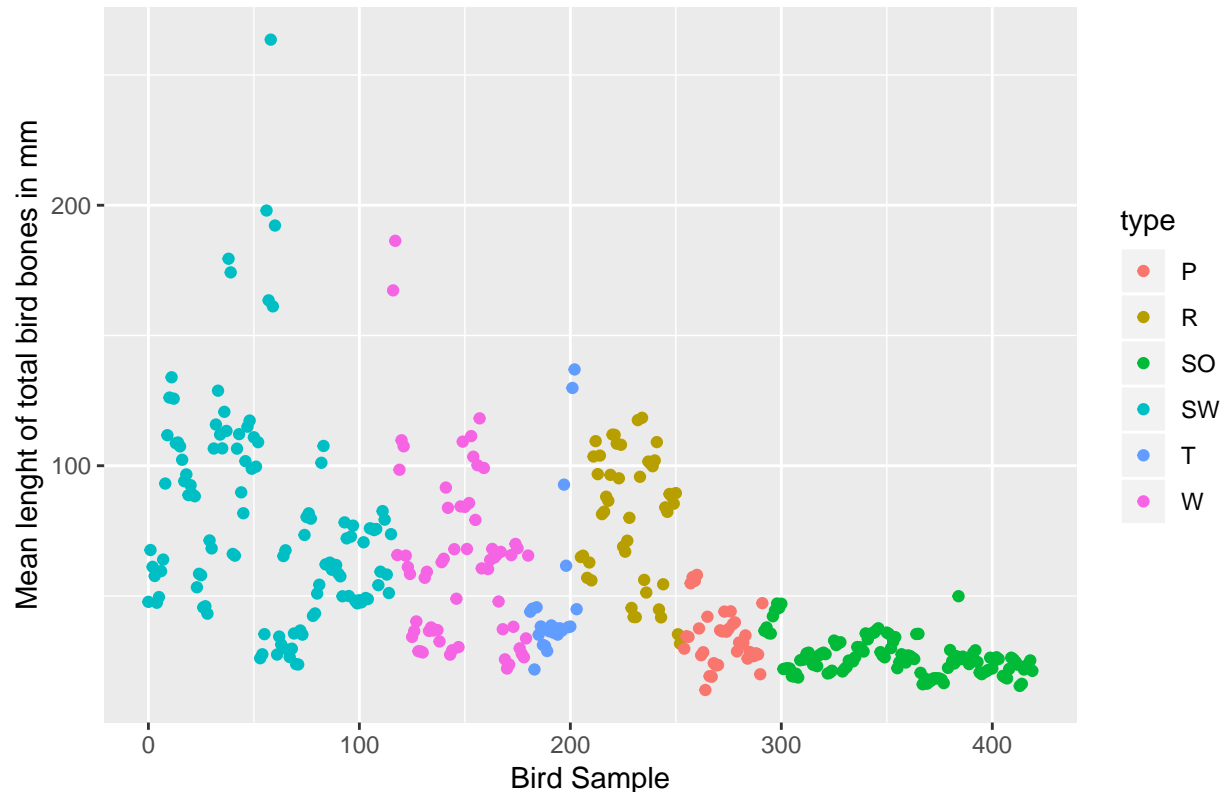
I'm creating a graph which displays the bonelengths on y axis and the Id on x colorcoded by their ecological group. By evaluating this we can see if some groups have overall larger or smaller bones and we see if there are big outliers.

```
library(ggplot2)  
library(reshape)  
source("../scripts/BoneMeans.R")  
BirdBones <- BoneMeans(BirdBones)  
ggplot(data = BirdBones, aes(id, length.mean, colour = type)) +  
  ggtitle("Bone lengths per Ecological group") +  
  ylab("Mean length of total bird bones in mm") +
```

```
xlab("Bird Sample")+
geom_point()
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

Bone lengths per Ecological group



As seen above swimming birds have the biggest bones, but also shown is that there are a lot more samples in that group where there is a lot of variation. I can look into cleaning up the data and removing the biggest outliers in this group. Singing birds also have a lot of samples but there is much less variation and so more certainty.

For the rest of the birds there are not a lot of sample so maby we could try and normalizing the data so there is an even amount of samples per group.

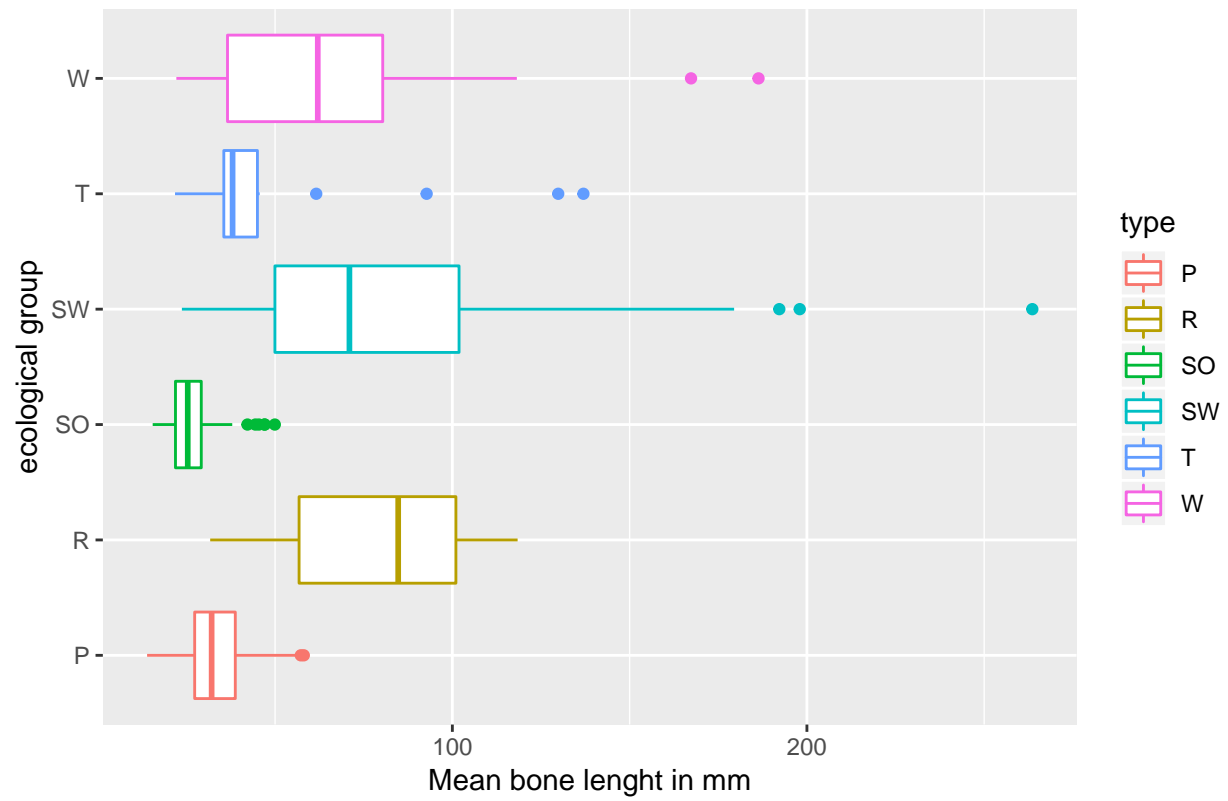
There are also 7 samples that contain missing values, we could just straight out not use these samples because 4 of these are part of the biggest group of samples. and the others are not part of the smallest groups.

```
library(ggplot2)
```

```
ggplot(BirdBones, aes(x = type, y = length.mean, color = type)) +
  geom_boxplot()+
  coord_flip()+
  ggtitle("Boxplot for each ecological group's mean bone lenght")+
  ylab("Mean bone lenght in mm")+
  xlab("ecological group")
```

```
## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
```

Boxplot for each ecological group's mean bone lenght



As you can see there are quite a few outliers in all groups except in group R ()