

# Predicting bird classification by bone lengths

*Jouke Profijt*

*October 8, 2018*

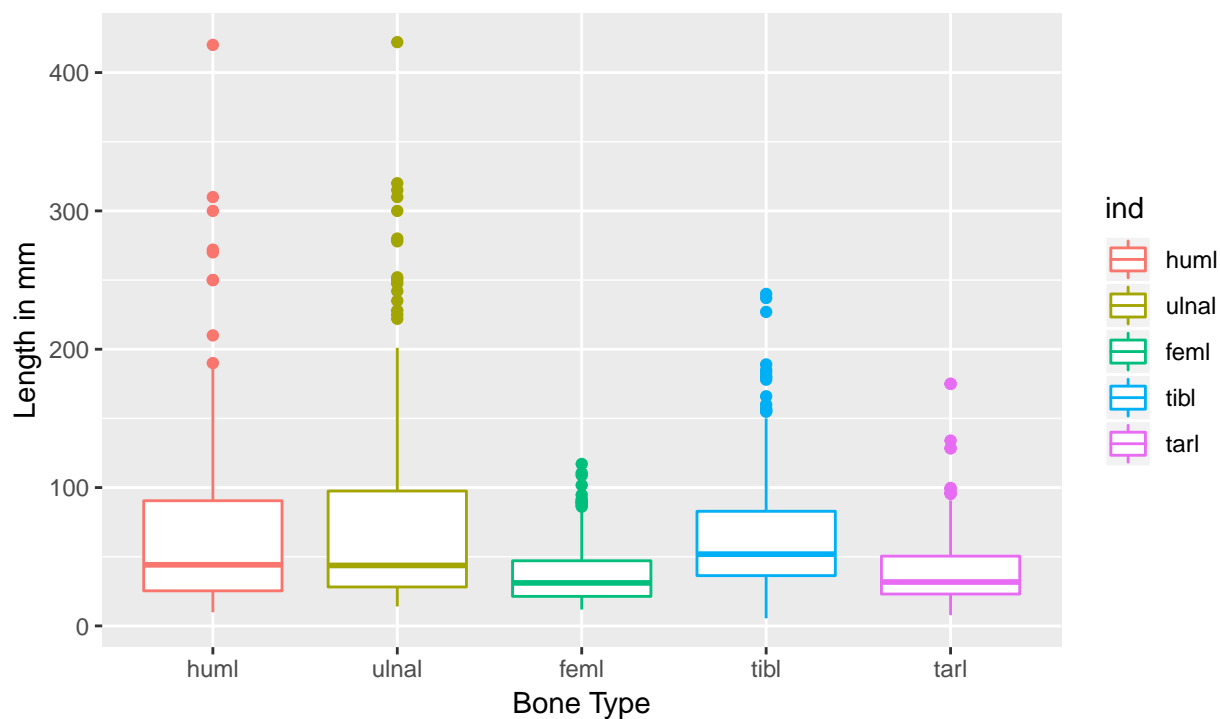
*#Copyright (c) 2018 Jouke Profijt.  
#Licensed under GPLv3. See LICENSE*

## Results

### EDA

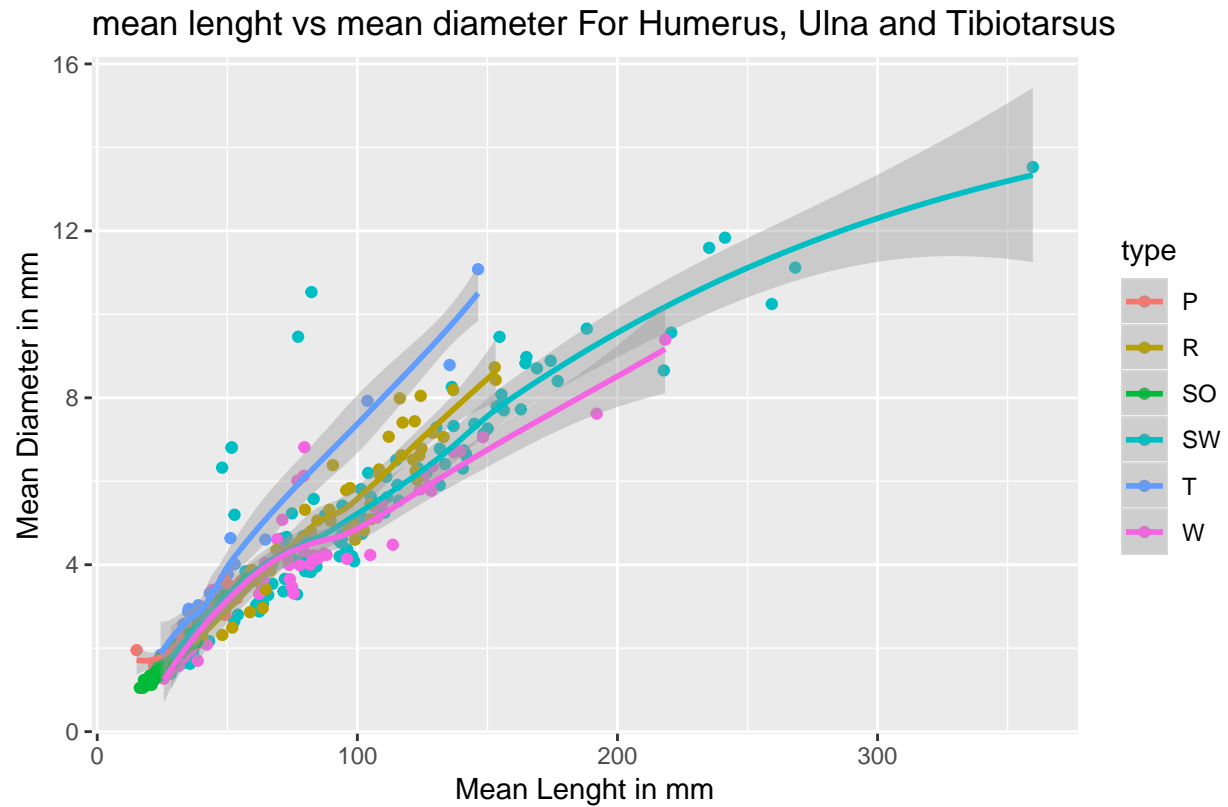
#### Boxplot for bone lengths per bone in mm

Variation between small and large bones.



Graph 1: Bone lengths

The Femur as we can see in Graph 1 is the one of the smallest bones in all bird groups, 2nd to the Tarsometatarsus. Because these 2 bones are so small there is not a lot of variation measurable between the different bird groups and thus might not be that important for classification.

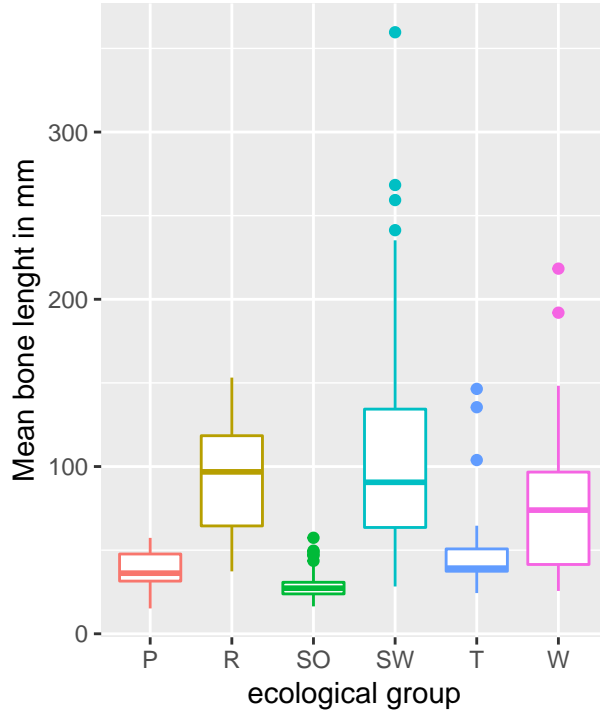


graph 2: lenght vs diameter

graph 2 shows us that the total bone lenght for the Swimming Birds is quite a bit larger than the other birds. Also we can see that the Terrestrial birds have thicker bones than the rest of the birds but are still quite small. the singing birds are overall very small and should be easy to classify. The raptors and Wading birds are a bit average and could become difficult to classify but the raptors are a bit thicker and the Wading birds are a bit smaller.

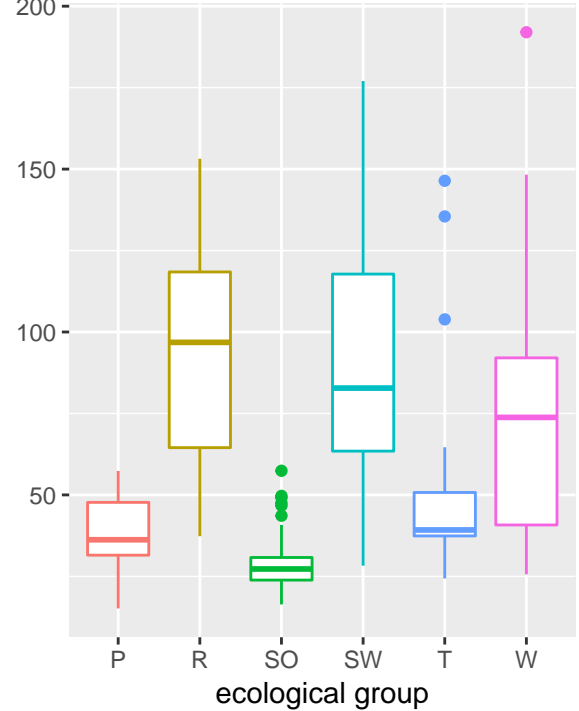
## Data Cleaning

Boxplot for each ecological group:  
For Humerus, Ulna and Tibiotarsus



graph 3: mean bone lengths

Boxplot for each ecological group:  
For Humerus, Ulna and Tibiotarsus



graph 4: Cleaned data

For our cleaning of data we removed all rows with NA's which cost us 7 datapoints. After that we calculated the 1st and 3rd quartile outliers from the humerus which can be seen in graph 3 & 4, and removed 9 rows. After all cleaning we left with 404 datapoints to use for Weka Analyses and classification.

## Weka

In weka the goal was to find a classification algorithm that has the highest possible accuracy because classifying a birds heritage it does not matter that much if there is a False positive, the goal was to keep false negatives low and keep true positives high.

**Chosen classifier: Random.Forest**

Table 1: Random.Forest Confusion matrix as chosen classifier

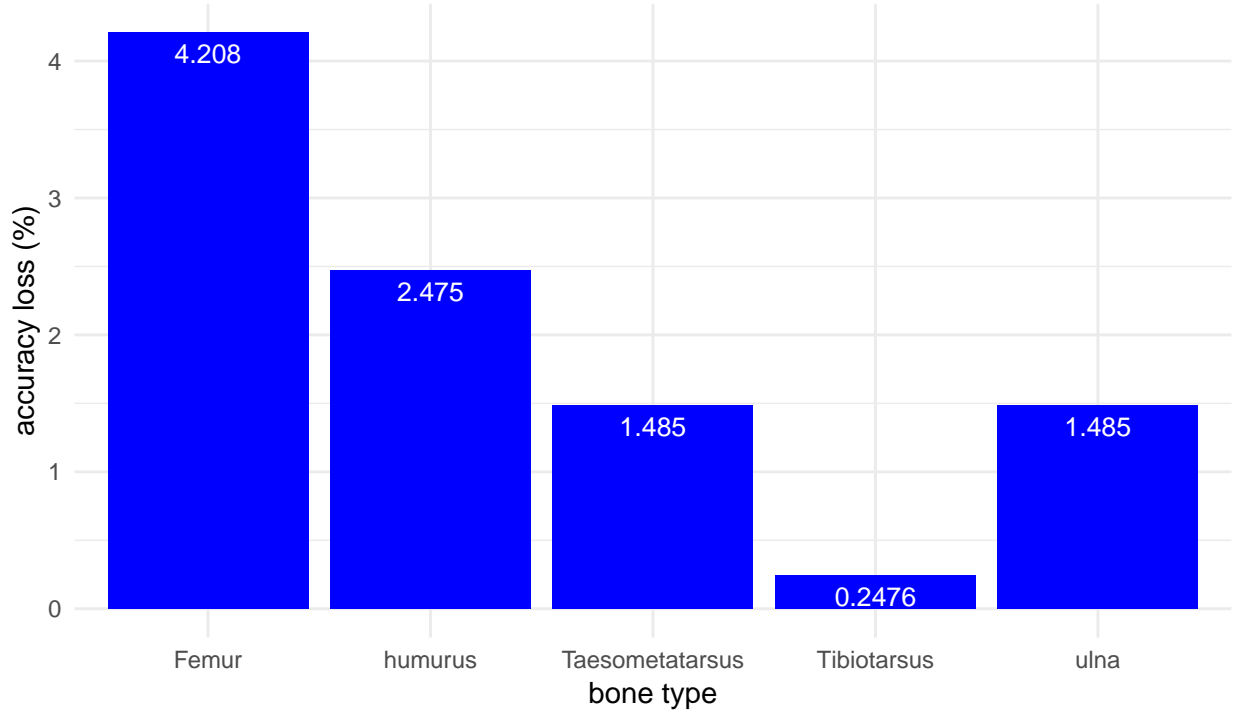
	SW	W	T	R	P	SO
SW	87	11	0	4	0	6
W	17	34	0	2	4	6
T	2	0	12	2	4	3
R	7	1	0	37	3	0
P	0	0	1	1	30	6
SO	0	0	1	1	1	121

Table 2: Random.Forest as chosen classifier

	Correct	Incorrect
<b>Instances</b>	321	83
<b>Percentage</b>	79.4554 %	20.5446 %

### Classification loss if certain bones are removed

For Random.forest



graph 5: accuracy loss per bone

in graph 5 we can see what the importance is from certain bones if we are talking about classification. the differences we can see are because of their importance for the functionality of the bird groups.

## Conclusion & discussion

What is the most important bone for each ecological group their function? we can conclude that (At least for classification) the Femur is the most important as when we remove this from our classification algorithms the accuracy loss is great. we can see this in graph 5 where the loss per bone is displayed. while researching the subject we made the assumption that the Femur and Taesometatarsus were the least important for classification. Because of this the first classification algorithms were done only using the longer bones but further in reverted this decision. The classification algorithm that was chosen however wasn't changed. In the future we should first collect some data about the importance for some bones before making such rational decisions.

We do want to think about removing some bones as when we have less data to be put in the more unknown fossils with missing bones could possibly be classified