

# Predicting bird classification by bone lengths

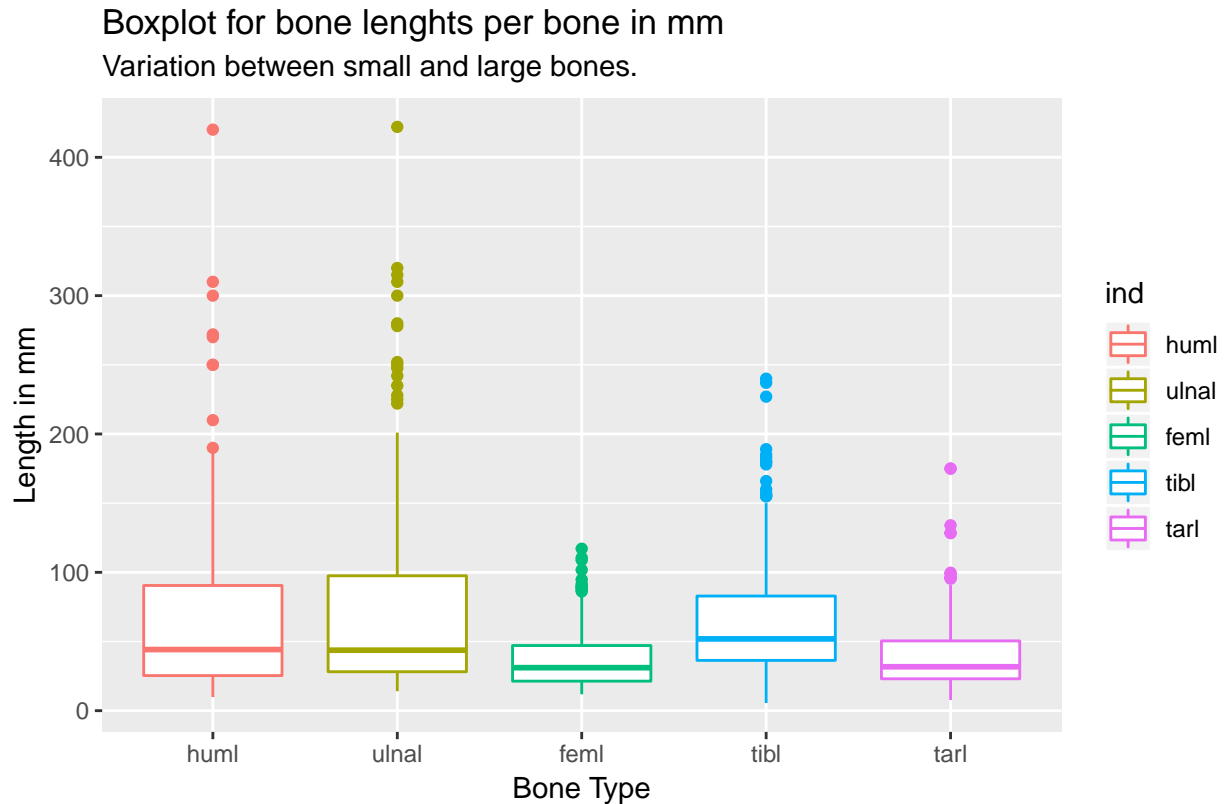
*Jouke Profijt*

*October 8, 2018*

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

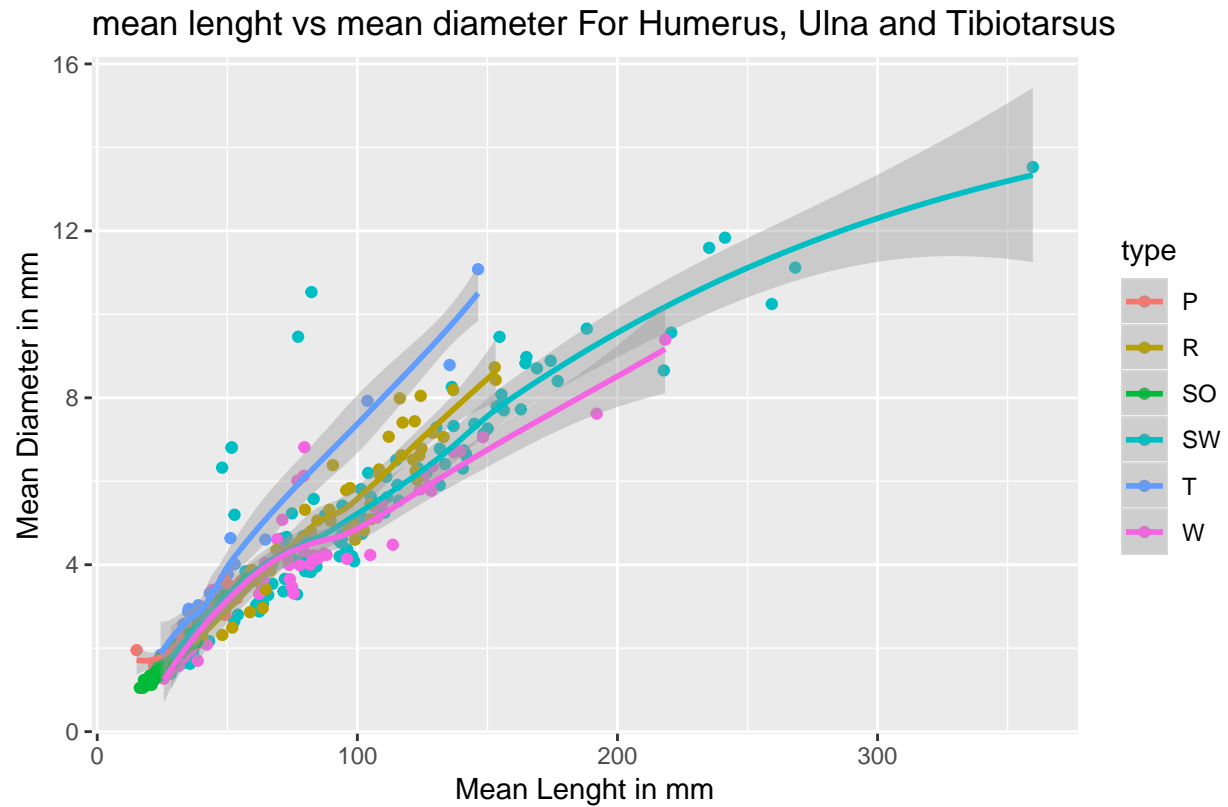
## Results

### EDA



Graph 1: Bone lengths

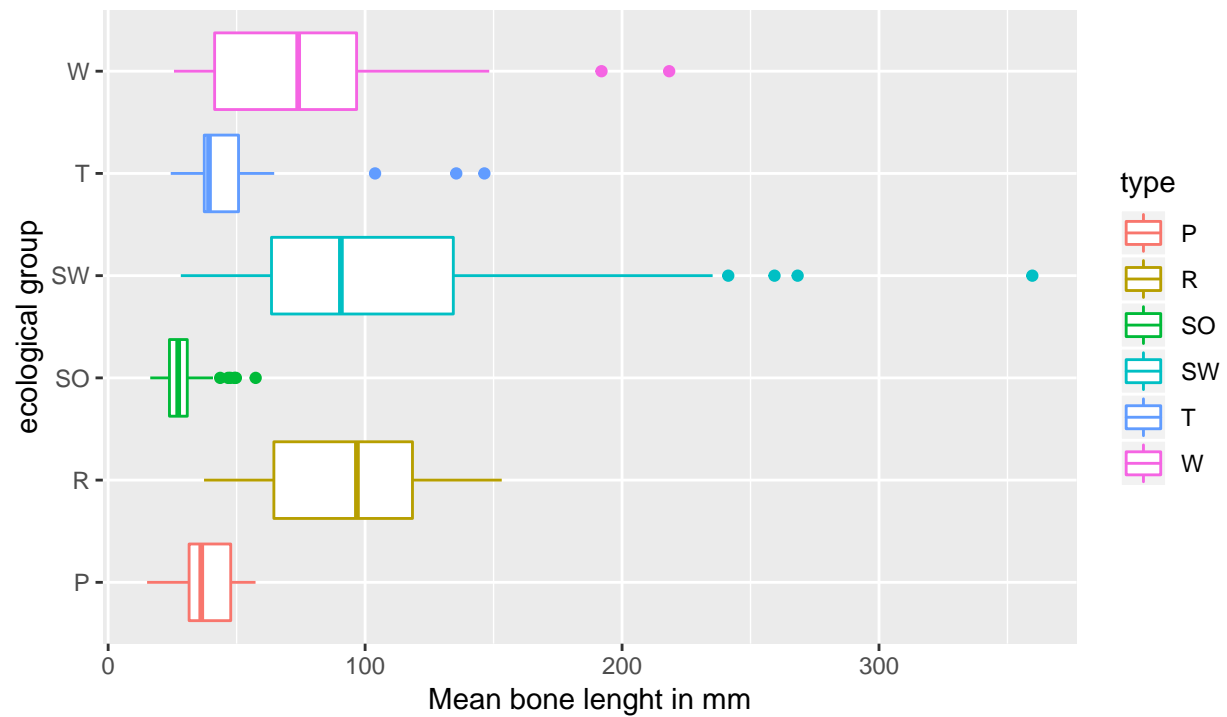
The Femur as we can see in Graph 1 is the one of the smallest bones in all bird groups, 2nd to the Tarsometatarsus. Because these 2 bones are so small there is not a lot of variation measurable between the different bird groups and thus not that important for classification, that is why in this experiment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

Boxplot for each ecological group's mean bone lenght  
For Humerus, Ulna and Tibiotarsus



graph 3: mean bone lenghts

In graph 3 again we see the bigger birds separated between their bonelenghts, and can quite easily distinguish each group. Although their is some overlap between lenghts we can conlude that classification via lenght the birds bone lenght is probably the best way of 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.

## Zero R

Table 1: Zero R Confusion matrix

	SW	W	T	R	P	SO
<b>SW</b>	0	0	0	0	0	108
<b>W</b>	0	0	0	0	0	63
<b>T</b>	0	0	0	0	0	23
<b>R</b>	0	0	0	0	0	48
<b>P</b>	0	0	0	0	0	38
<b>SO</b>	0	0	0	0	0	124

Table 2: Zero R classification results

	Correct	Incorrect
<b>Instances</b>	124	280
<b>Percentage</b>	30.6931 %	69.3069 %

Zero R uses zero rules as the name suggests, because of this it chooses 1 group that is most likely to be classified as seen in table 1.

## One R

Table 3: One R Confusion matrix

	SW	W	T	R	P	SO
<b>SW</b>	69	11	1	10	8	9
<b>W</b>	35	2	4	4	10	8
<b>T</b>	4	1	4	2	8	4
<b>R</b>	29	5	1	6	6	1
<b>P</b>	5	0	7	0	16	10
<b>SO</b>	2	0	3	0	10	109

Table 4: One R classification results

	Correct	Incorrect
<b>Instances</b>	208	196
<b>Percentage</b>	51.4851 %	48.5149 %

Chosen classifier: **Random.Forest**

Table 5: Random.Forest Confusion matrix as chosen classifier

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

Table 6: Random.Forest as chosen classifier

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