Quality Report

Jouke Profijt

September 30, 2018

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

TLDR;Summary

In my data cleaning i will be removing all samples with NA's becouse of the importance of these missing values. If the NA's were only in the smaller unimportant bones i would havee considered not removing them. I also took a look at the outliers which are the bigger birds, this data could be important but they are very diffrent from their respective means. I'm also not going to use bones with very little variation between them for classification, so the small bones, becouse: small variation makes for hard classification of bird groups.

NA's	Outliers
7	9

```
# Below is some summary data to genarate the above conclusion
```

Table 2: Table continues below

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	NA's :1	NA's :1	NA's :3

Table 3: Table continues below

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	NA

When we look at the summary we can see that every bone contains atleast 1 NA, And from our EDA we concluded that we dont want to use the small bones Femur & Tibiotarsus for classifiacation. So if the a sample only contains NA's in those two bones we can still use those samples.

Table 5: Table continues below

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw
161	160	76.43	4.11	86.79	3.84	NA	NA	67.13	2.48
205	204	63.76	4.74	NA	NA	57.33	4.88	75.67	4.33
208	207	98.08	7.77	113	5.76	82.04	7.17	107.5	6.65
343	342	NA	NA	NA	NA	32.54	2.65	55.06	2.81
379	378	20.1	1.86	NA	1.52	17.21	1.22	NA	NA
397	396	16.51	1.47	20.56	1.43	15.88	1.27	NA	1.19
405	404	20.36	1.87	22.19	1.6	NA	1.77	37.47	1.64

	tarl	tarw	type
161	41.65	2.1	W
205	60.19	3.82	R
208	NA	NA	R
343	38.94	2.25	SO
379	18.46	0.91	SO
397	17.63	1.02	SO
405	25.54	1.34	SO

we can see that there are no birds who only have NA's in the smaller bones so i suggest just not using the birds with NA's as shown above.

Table 7: Table continues below

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw
34	33	210	13.03	278	10.74	56.87	8.03	76.66	5.2
39	38	272	14.86	320	10.42	91.6	9.71	132	10.23
40	39	270	14.25	310	10.9	86.2	9.96	125.8	9.63
57	56	310	14.4	315	9.51	88.77	8.1	180	9.45
58	57	250	11.91	252	8.31	73.04	7.37	160	8.47
59	58	420	17.84	422	11.72	110.5	9.99	237	11.03
60	59	250	11.28	247.5	7.5	69.04	6.2	156	7.19
61	60	300	12.48	300	8.65	84.05	8.53	178	9.61
118	117	190	11.92	225	8.55	101.8	7.75	240	7.71

	tarl	tarw	type
34	22.54	7.16	SW
39	81.77	8.91	SW
40	79.18	10.05	SW
57	96.13	7.69	SW
58	82.46	7.04	SW
59	128.3	8.93	SW
60	83.36	6.13	SW
61	99.01	7.55	SW
118	175	7	W

As we can see most outliers come from the type Swimming Birds this could just mean that these birds are

quite large but dont have a lot of samples to back it up.

Table 9: Sample count outliers

Swimming Birds	Wading Birds
116	65

Table 10: Mean lenghts

Swimming Birds	Wading Birds
110.3	73.13

As we can see here, all outliers are quite far from their respective means so that makes me consider that we should remove all outliers as they are in the groups with a lot of samples.