

데이터과학기초 1차과제

컴퓨터학부 2017110762 노태현

기본사항

6 종류의 생태학적 범주로 나눌 수 있는 새들의 뼈의 길이와 지름의 크기에 대한 데이터

데이터 출처: <https://www.kaggle.com/zhangjuefei/birds-bones-and-living-habits>

Attributes: 11개

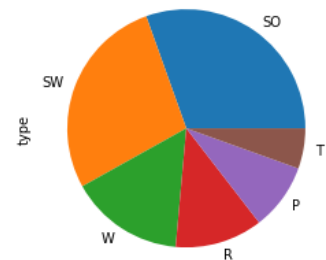
- 생태학적 범주: Swimming Birds(물새, **SW**), Wading Birds(습금류, **W**), Terrestrial Birds(타조목, **T**), Raptors(맹금류, **R**), Scansorial Birds(딱따구리목, **P**), Singing Birds(참새목, **SO**)
- **humerus**(상완골) length, diameter, **ulna**(자뼈) length, diameter, **femur**(대퇴골) length, diameter, **tibiotarsus**(경골) length, diameter, **tarsometatarsus**(부척골) length, diameter

Instances: 420개

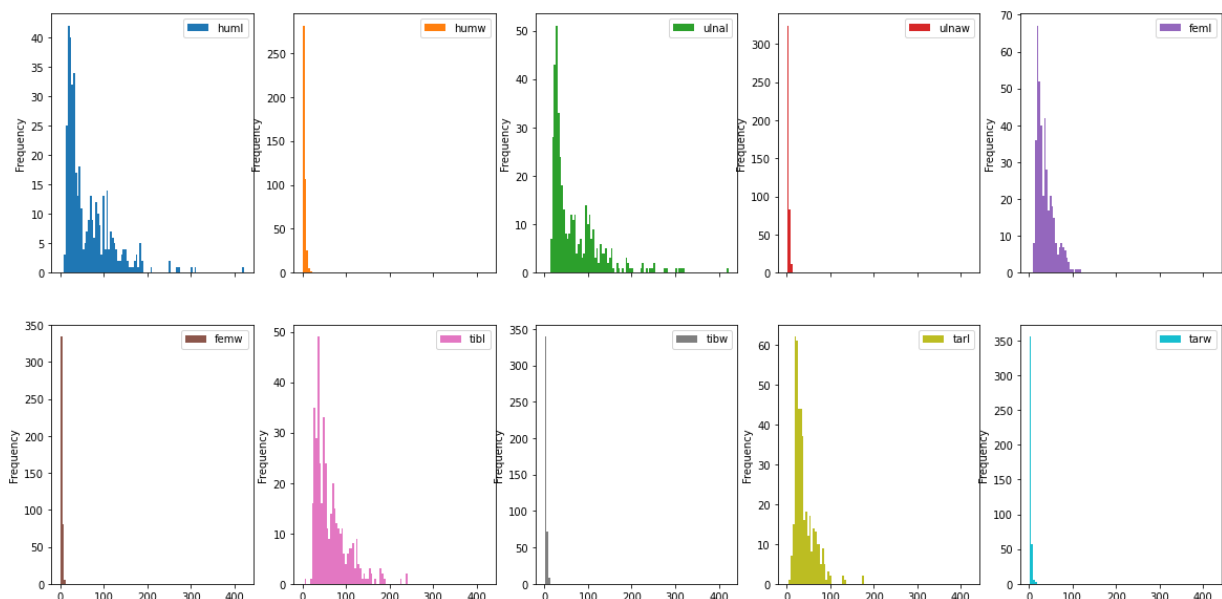
- type별 분류: SW(116), W(65), T(23), R(50), P(38), SO(128)

특성분석 및 시각화: Pie plot

- 물새와 참새목의 새들이 데이터의 절반 이상을 차지함
- 타조목과 딱따구리목의 새들은 표본의 수가 비교적 적음



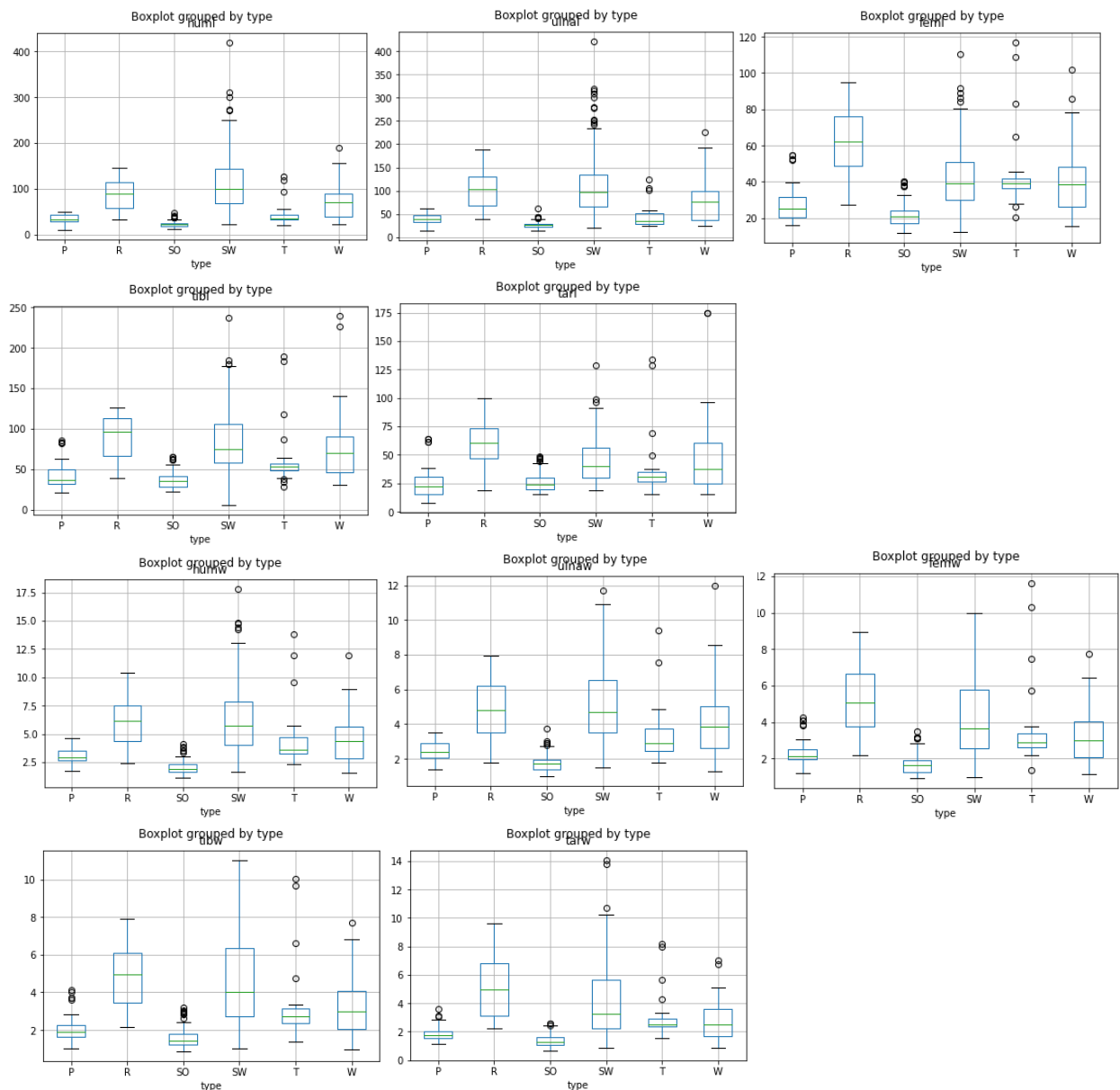
특성분석 및 시각화: Histogram(단위: mm)



- huml, ulnal의 분포 모양이 서로 닮아 있으므로 종을 고려하지 않고 뼈의 길이만 고려한다면, humerus, ulna의 길이에는 서로 상관관계가 있을 것으로 추정됨

- feml, tibl, tarl의 분포 모양이 서로 닮아 있으므로 뼈의 길이만 고려한다면, femur, tibiotarsus, tarsometatarsus의 길이에 서로 상관관계가 있을 것으로 추정됨
- 데이터에서 제공된 뼈의 길이들이 200mm 이하인 새들이 대부분이므로 히스토그램에서 나온 빈도수만 보고 생태학적 범주에 따라 종류가 나뉜다고 생각하기는 어려움

특성분석 및 시각화: Box Plot(단위: mm)



- Raptor, Swimming Birds, Wading Birds는 Scansorial Birds, Terrestrial Birds, Singing Birds에 비해 굵고 긴 뼈들을 가지고 있으며, 비교적 큰 분산을 가짐
- Scansorial Birds는 femur, tibiotarsus, tarsometatarsus의 길이, 지름에 대해서만 큰 이상치를 가진 새들이 존재하므로, 히스토그램에서 분포 모양을 본 것처럼 상관관계가 보임
- Raptor는 뼈의 길이, 지름에 대해 이상치의 값이 없으나, 박스 크기는 작지 않음

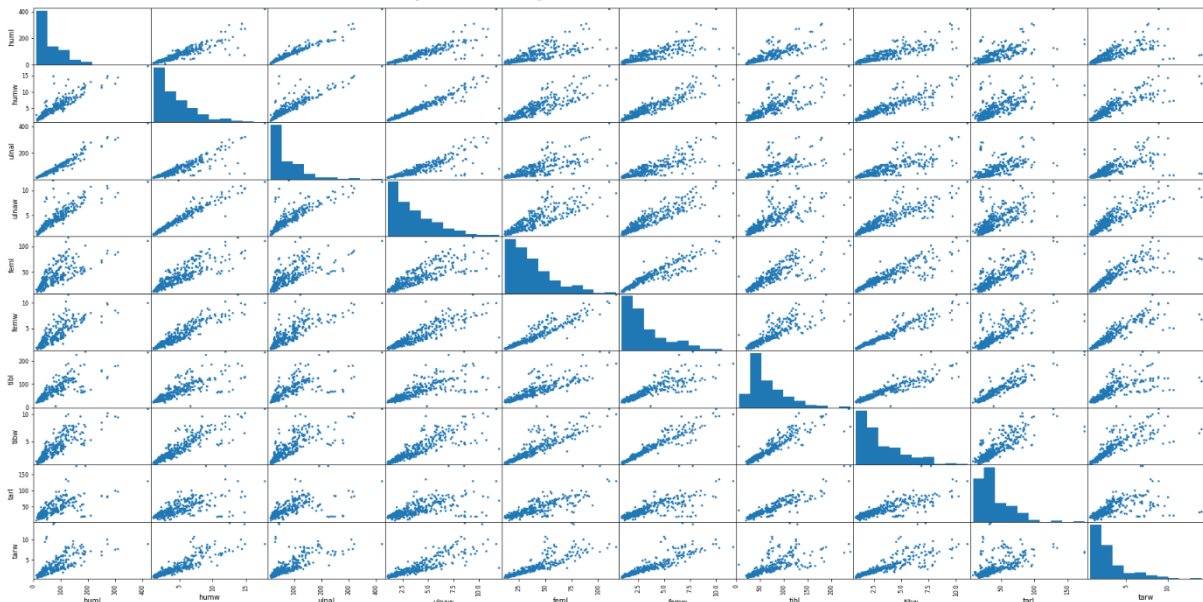
- Singing Birds는 뼈의 길이, 지름에 대해 큰 이상치의 값은 많고 작은 이상치의 값은 없으며, 이상치가 많기에 1분위수와 3분위수의 차이가 작음
- Swimming Birds는 뼈의 길이에 대해서는 모두 큰 이상치의 값을 여럿 가지지만, 뼈의 지름에 대해서는 지름에 비해 큰 이상치의 값이 몇몇 없으므로 Swimming Birds의 데이터에서의 뼈들은 길이가 길다고 무조건적으로 지름도 굵어지지는 않는다고 분석이 가능함
- Terrestrial Birds는 이상치들 간의 격차가 다른 새들에 비해서 크고, femur의 길이와 지름에 대해서는 다른 새들에게는 존재하지 않는 데이터인 낮은 이상치 값이 존재함
- Swimming Birds를 제외하고는 뼈의 길이에서 이상치가 존재하면, 그에 대응하는 뼈의 지름에서도 대부분 이상치가 존재하므로 뼈의 길이와 지름에 대한 상관관계가 유추 가능함

특성분석 및 시각화: 종에 따른 새들의 뼈들의 길이와 지름의 평균값 비교(단위: mm)

huml	64.650501	huml	34.423947	huml	86.934400	huml	22.361102	huml	110.251121	huml	45.699565	huml	73.133077
humw	4.370573	humw	3.039211	humw	6.065600	humw	2.029764	humw	6.423966	humw	4.786087	humw	4.607077
ulnal	69.115372	ulnal	39.177368	ulnal	100.342449	ulnal	26.398095	ulnal	111.755517	ulnal	45.617391	ulnal	78.101385
ulnaw	3.597249	ulnaw	2.476316	ulnaw	4.813265	ulnaw	1.743858	ulnaw	5.221638	ulnaw	3.471739	ulnaw	4.102615
feml	36.872416	feml	28.217368	feml	62.020000	feml	21.383465	feml	42.169655	feml	46.277391	feml	40.119531
femw	3.220883	femw	2.307895	femw	5.266800	femw	1.680313	femw	4.275776	femw	3.819565	femw	3.118594
tibl	64.662823	tibl	41.881316	tibl	89.877400	tibl	36.326032	tibl	85.317069	tibl	66.090870	tibl	76.150000
tibw	3.182339	tibw	2.095789	tibw	4.949200	tibw	1.548740	tibw	4.513534	tibw	3.451739	tibw	3.179231
tarl	39.229976	tarl	25.787368	tarl	59.166327	tarl	25.842656	tarl	45.156207	tarl	40.086522	tarl	47.543385
tarw	2.930024	tarw	1.902632	tarw	5.073878	tarw	1.349375	tarw	4.149914	tarw	3.183913	tarw	2.760308
dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64	dtype: float64
mean All	mean P	mean R	mean SO	mean SW	mean T	mean W							

- 날개의 두 뼈인 humerus, ulna가 평균적으로 가장 길고 굵은 새들은 Swimming Birds였고, 평균적으로 그 뼈들이 가장 짧고 얇은 새들은 Singing Birds로 측정됨
- 다리의 세 뼈인 femur, tibiotarsus, tarsometatarsus가 평균적으로 가장 길고 굵은 새들은 Raptor였고, 평균적으로 그 뼈들이 가장 짧고 얇은 새들은 Singing Birds로 측정됨
- 다리에 있는 세 뼈들에 비해 날개에 있는 두 뼈들의 길이 차이가 더 작은 것으로 분석됨

특성분석 및 시각화: Scatter Plot(단위: mm)



- humerus, ulna, femur, tibiotarsus의 길이와 지름을 비교하면 대부분 길이가 길어질수록 지름도 커지는 경향이 보이므로 길이와 지름 간의 상관관계가 있다고 할 수 있음
- 그러나 tarsometatarsus의 경우에는 길이가 길어질수록 지름도 커진다고 하기 어려우므로 길이와 지름 간의 상관관계를 규정하기 어려움
- 새의 날개에 있는 두 뼈인 humerus와 ulna 사이에는 길이와 지름이 커질수록 서로 커지는 경향이 깔끔하게 보이므로 히스토그램에서 추정하 것처럼 두 뼈들 사이에는 다른 뼈들에 비해 강한 상관관계가 있다고 볼 수 있음
- 새의 다리에 있는 두 뼈인 femur와 tibiotarsus 사이에도 길이와 지름이 커질수록 서로 커지는 경향이 깔끔하게 보이므로 히스토그램에서 추정하 것처럼 두 뼈들 사이에는 상관관계가 있다고 할 수 있음