

# Leniwce: analiza danych

Jakub Pryc

## 1. Opis Danych

Źródło: [Sloth Species \(Kaggle\)](#)

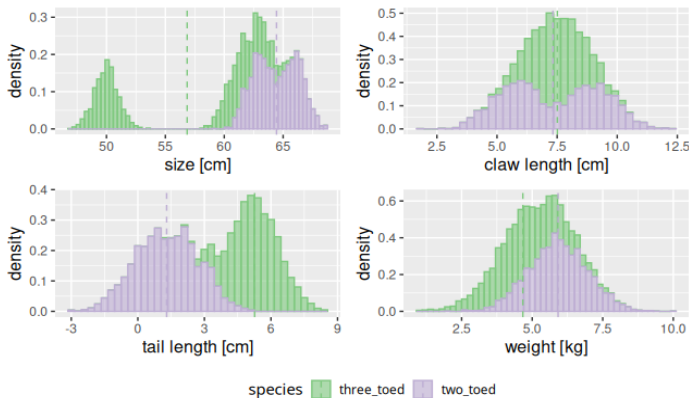
Po obróbce analizowane dane zawierają 5000

rekordów oraz 7 kolumn opisujących parametry:

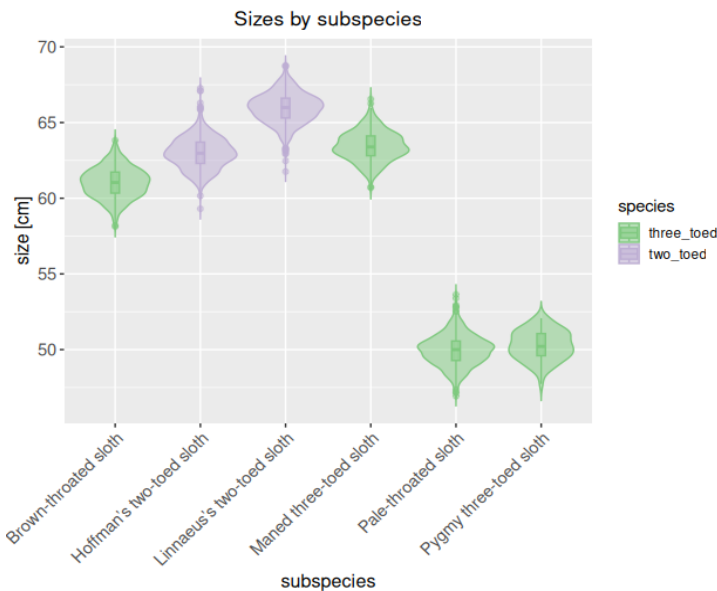
- endangered - stopień zagrożenia wyginięciem
- species - rodzina leniwca
- sub\_species - gatunek leniwca
- size\_cm - długość leniwca w cm
- weight\_kg - waga leniwca w kg
- claw\_length\_cm - średnia długość pazurów w cm
- tail\_length\_cm - długość ogona w cm

## 2. Podstawowa analiza

Histogramy cech leniwców aproksymujące gęstość z podziałem na rodziny, gdzie pionowa linia oznacza średnią:



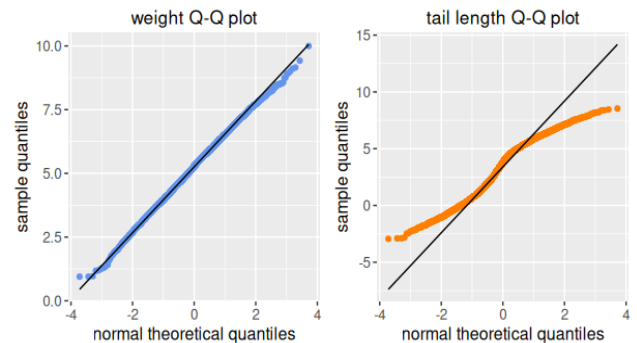
Najbardziej nietypowy kształt ma histogram dotyczący długości leniwca, więc w celu dokładniejszej analizy przyjrzymy się bardziej szczegółowo długości leniwców, teraz z podziałem na gatunki oraz z pomocą wykresów skrzypcowych i pudełkowych:



Przyczyna zjawiska, które wcześniej zaobserwowaliśmy szybko stała się jasna, niektóre gatunki z rodziny leniuchowców (dwupalczastych) są znacznie mniejsze od innych. W obrębie gatunku rozkłady ukazane nam przez wykresy skrzypcowe i pudełkowe są bliższe rozkładowi normalnemu.

## 3. Normalność rozkładów

Niektóre z histogramów wyglądają, jakby dane przez nie przedstawiane pochodziły z rozkładu normalnego, lecz tak postawioną tezę należy odpowiednio poprzeć. W tym celu przeanalizujemy wykresy Q-Q oraz wyniki testu Kołmogorowa-Smirnowa.



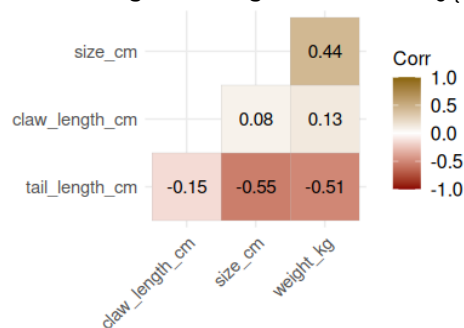
Wykresy te pokazują, jak kwantyle naszych danych odnoszą się do kwantyli rozkładu normalnego. Im bliżej naszym punktom do prostej nachylonej pod kątem 45 stopni, tym bliżej naszemu rozkładowi do rozkładu normalnego. Na podstawie powyższych wykresów można przypuszczać, że wagi leniwców pochodzą z rozkładu normalnego, lecz długości ich ogonów już nie. W celu weryfikacji tych przypuszczeń przeprowadzimy test Kołmogorowa-Smirnowa. Daje on następujące wyniki:

	Długość leniwca	Długość ogona
D	0.015203	0.093202
p-value	0.1981	< 2.2e-16

Ponieważ wartość p-value dla długości ogona jest mniejsza od 0.05 możemy założyć, że jej rozkład jest znacząco różny od rozkładu normalnego. W przypadku długości leniwca p-value jest większe od 0.05, więc rozkład ten nie różni się znacząco od normalnego. Możemy zatem oszacować parametry  $\mu$  oraz  $\sigma^2$  jako średnią oraz odchylenie standardowe kwadrat. W związku z tym możemy stwierdzić, że wartości długości leniwców pochodzą z rozkładu normalnego  $N(60.39985, 35.16452)$ .

## 4. Zależności pomiędzy danymi

W celu zbadania zależności pomiędzy cechami liczbowymi leniwców posłużymy się macierzą korelacji oraz jej reprezentacją graficzną. Możemy z niej odczytać, że długość i waga leniwców mają współczynnik korelacji



równy 0.44, co oznacza, że mają między sobą umiarkowanie silny dodatni związek. Długość ogona natomiast jest skorelowana z długością i wagą leniwca ze współczynnikami -0.55 oraz -0.51, co oznacza, że są one silnie ujemnie skorelowane.

## 5. Wnioski

Analiza statystyczna naszych danych dostarczyła wielu interesujących informacji na temat rozważanych gatunków leniwców. Pokazuje to jak przydatnym narzędziem jest analiza danych i jak pozwala nam dowiadywać się nowych rzeczy o naszym świecie, które przy zwykłym spojrzeniu pozostałyby niewidoczne, a jedynie zestawione z licznymi innymi danymi dają pełny obraz rzeczywistości.