Exploratory Data Analysis

Wisse Schuuring

9/21/2021

Data origin

"Crab body metrics" is the title of the dataset sourced from "A Multivariate Study of Variation in Two Species of Rock Crab of the Genus Leptograpsus", an article published in 1974 by N. A. Campbell and R. J. Mahon.

Data

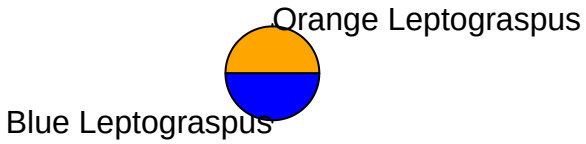
This dataset contains 200 rows and 8 columns, describing 5 morphological measurements on 50 crabs. Each crab of two colour forms and both sexes, of the species Leptograpsus variegatus. These crabs were collected at Fremantle, West Australia. The data is complete, containing no missing values.

Name	Full.name	Data.type	Unit	Description
SP	Species	chr	N/A	Blue and Orange define the two species of purple rock crab.
sex	Sex	chr	N/A	Male and Female determine the sex of the purple rock crab.
index	Unique Row Identifier	int	num	Unique row identifier for the purple rock crab.
FL	Frontal lobe size	dbl	mm	The frontal lobe size of the purple rock crab.
RW	Rear width	dbl	mm	The Rear width of the purple rock crab.
CL	Carapace length	dbl	mm	The carapace length of the purple rock crab.
CW	Carapace width	dbl	mm	The carapace width of the purple rock crab.
BD	Body depth	dbl	mm	he vertical distance between the dorsal and ventral margins of the purple rock crab.

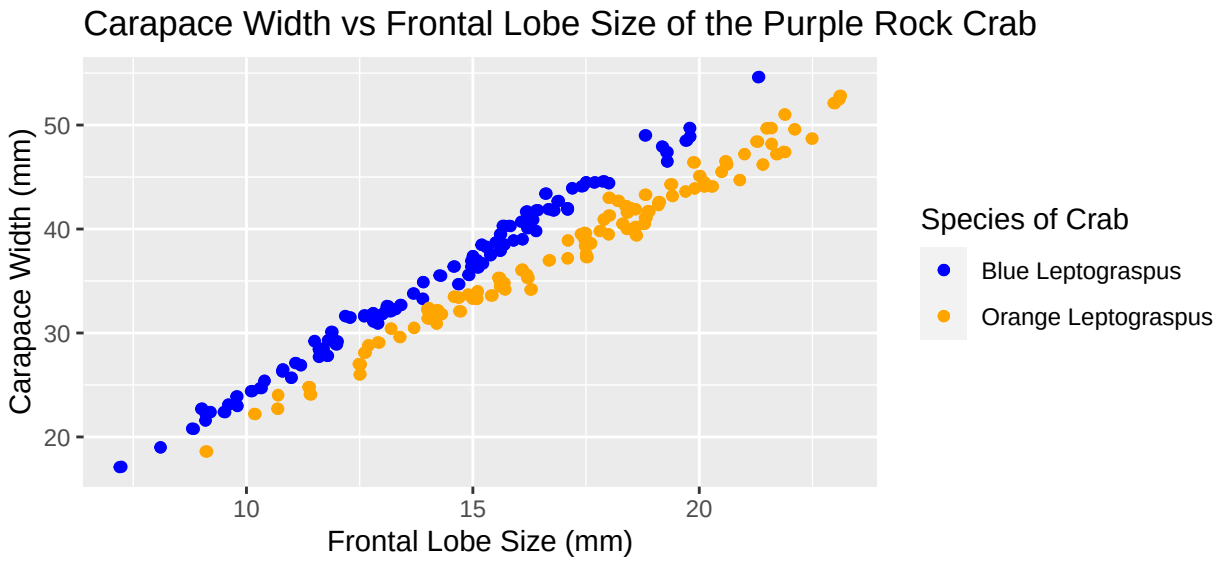
The crabs have been measured on their frontal lobe size (FL), rear width (RW), carapace length (CL), carapace width (CW) and body depth (BD) in mm. Furthermore their gender and colour have been recorded.

Visualisation

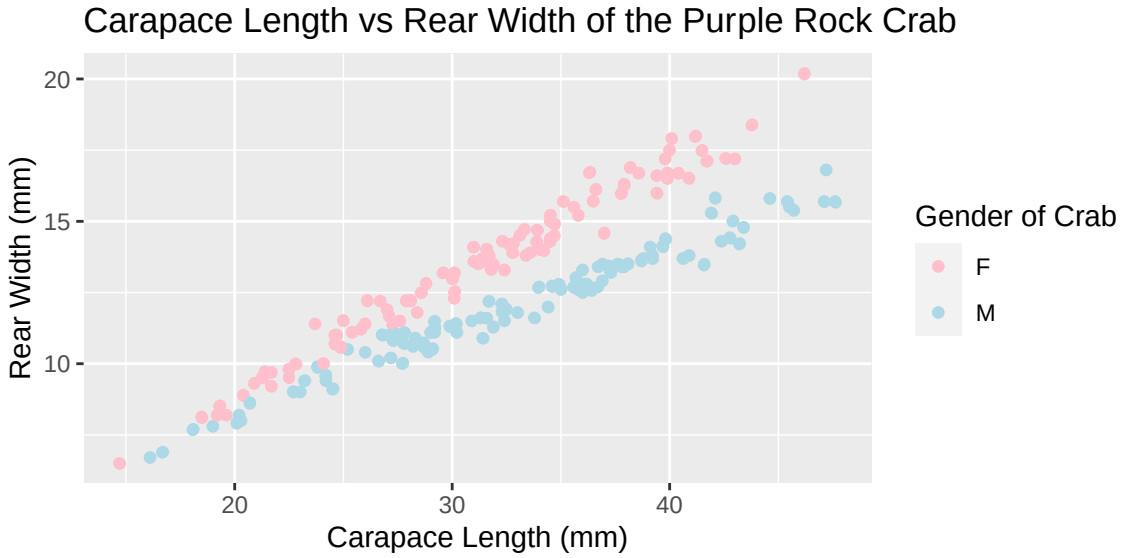
Distribution of the Crab body metrics



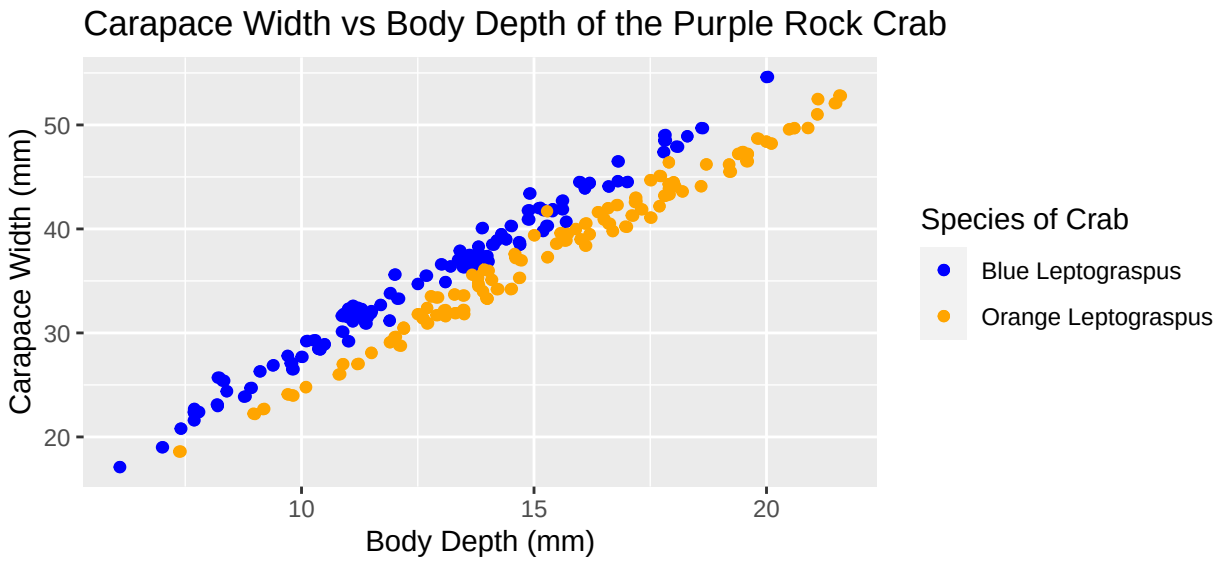
There is an even distribution of records from both species of Leptograspus, with 50% Blue and 50% orange.



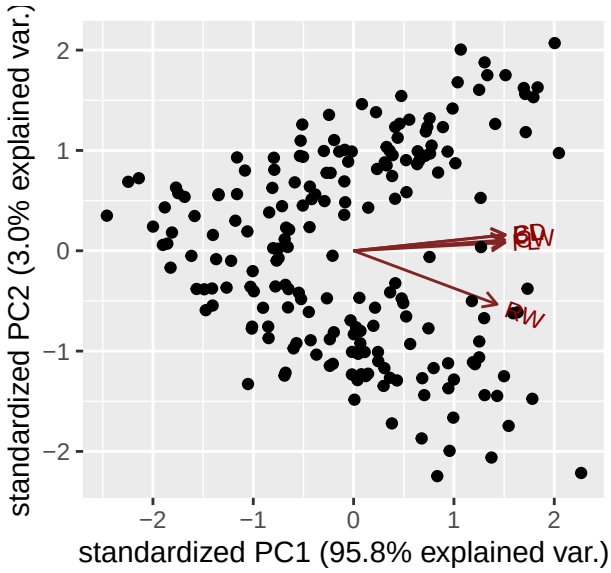
The most prominent separation between the blue and orange species can be visualized when discriminating upon the carapace width (CW) and the frontal lobe size (FL). These are the optimal variables with which one could classify the species of Leptograpsus.



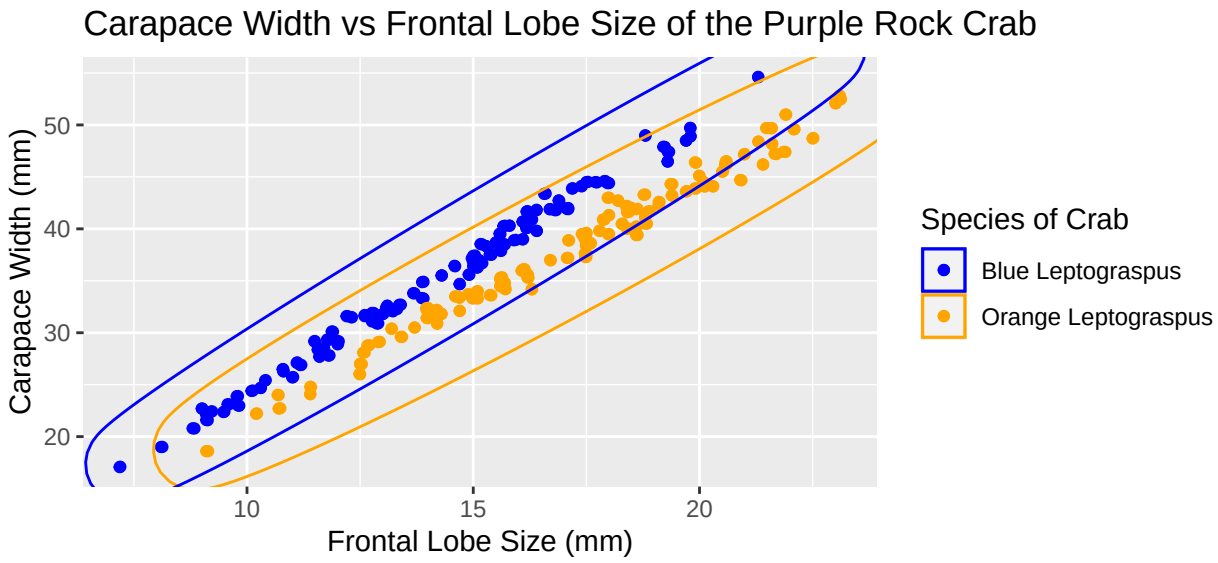
Discriminating upon carapace length (CL) and rear width (RW) shows the gender difference and can be used to identify individuals of either crab species' gender.



Although discriminating upon carapace width (CW) and body depth (BD) too can be used to determine the species of the Purple rock crab, it is less reliant and separate than discriminating on carapace width (CW) and the frontal lobe size (FL).



As seen in the PCA, all variables have about the same impact upon the dotplot, with the exception of Rear Width, which makes the RW value the one with the most information gain.



With a stricter formula, one could easily identify any new Crab measurements and identify them by species through clustering.