

CHALLENGE

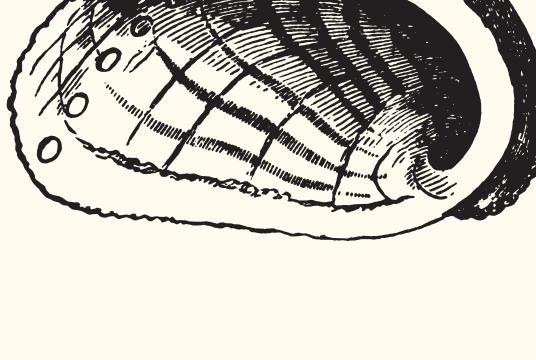
Dimension Reduction

The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope - a time-consuming task. Other measurements, which are easier to obtain, are used to predict the age.

VARIABLE TRANSFORMATION

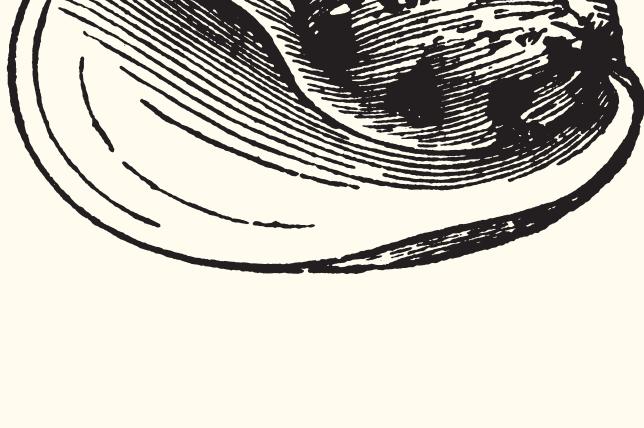
01

We need to prepare the data set.
Transform the gender variable into a suitable form and remove the rings variable.



02

CORRELATION AND STANDARDIZATION



Let's have a look at the correlations among the variables. That will enable us to see which variables carry similar information. The common problem for Dimension Reduction is multicollinearity.
Then, standardize the variables, a necessary step for PCA.

03 OPTIMAL NUMBER OF COMPONENTS

A necessary step for PCA is to indicate the optimal number of components.

Use the elbow rule to determine how many to use.



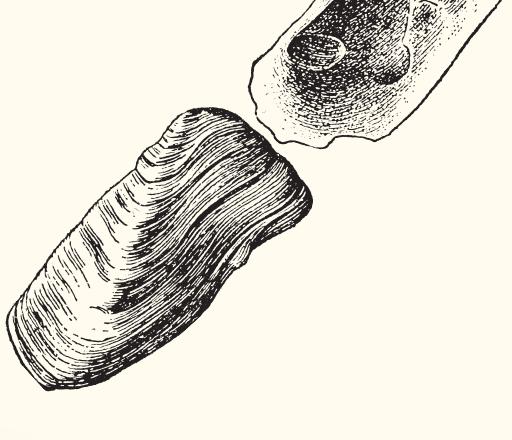
04 PCA

You should now be ready to create the PCA model. Next comes an important step: the interpretation of the components created.



05 MANIFOLD LEARNING AND

Though still in research, Manifold Learning is one the best ways to perform Dimension Reduction. Unfortunately, there is still no best way to determine the optimal number of components for Manifold Learning. Use 2 components and visualize the results.



Diogo Resende

