



INTRODUCTION TO PYTHON & R

DAY 1 (29 Sept.): Introduction to R



paul j. van staden (PhD)
Department of Statistics
University of Pretoria





DATA MANIPULATION & VISUALISATION

- In this section we will cover the following:
 - Visualisation for categorical variables:
 - Bar plots
 - Visualisation for numerical variables:
 - Box plots
 - Histograms
 - Visualisation for paired variables:
 - Scatterplots





Data sets

Data on South African animals.

DATA compiled from:

South African National Parks & Pick n Pay (2017). South African Super Animals Trump Cards.

Data on the level of anxiety experienced by arachnophobes when interacting with spiders.

DATA adapted from:

Field, A. (2009). *Discovering Statistics using SPSS (and sex and drugs and rock 'n' roll)*, SAGE Publications Ltd, London, UK.



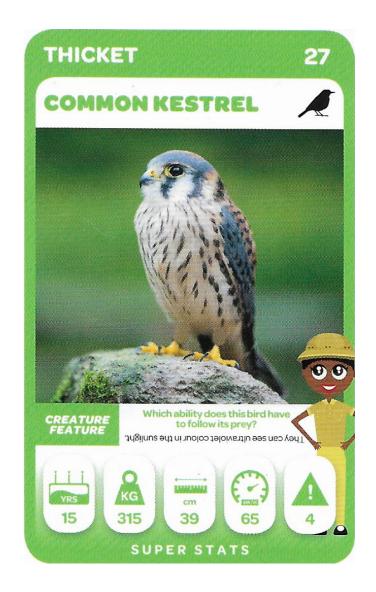


EXAMPLE: South African animals

- Consider the following variables for 108 animals:
 - Card number
 - Type of animal
 - Habitat
 - Species
 - Age in years
 - Weight in kilograms
 - Size in centimeters
 - Speed in kilometer per hour
 - Vulnerability status scaled from 1 to 4 (highest to lowest risk)
- The focus will be on the 36 birds in the data set.















EXERCISE: Arachnophobia

- In a study on arachnophobia (the fear of spiders), 24 arachnophobes (persons fearing spiders) had to interact with spiders of different sizes.
- During each interaction, the level of anxiety of the arachnophobe was measured through galvanic skin response (GSR).
- Consider the following two variables for i = 1, 2, ..., 24:
 - y_i : The GSR measurement for the level of anxiety of the i^{th} arachnophobe.
 - x_i : The size of the spider in centimeters (cm) for the i^{th} arachnophobe.
- Draw box plots for y and for x and a scatterplot for y against x.