# STATS 2DA3 Winter 2023

# ASSIGNMENT 2

**Submit through Avenue to Learn.**

**Due before 11pm on Friday, February 10th.**

**Your assignment must conform to the Assignment Standards listed below.**

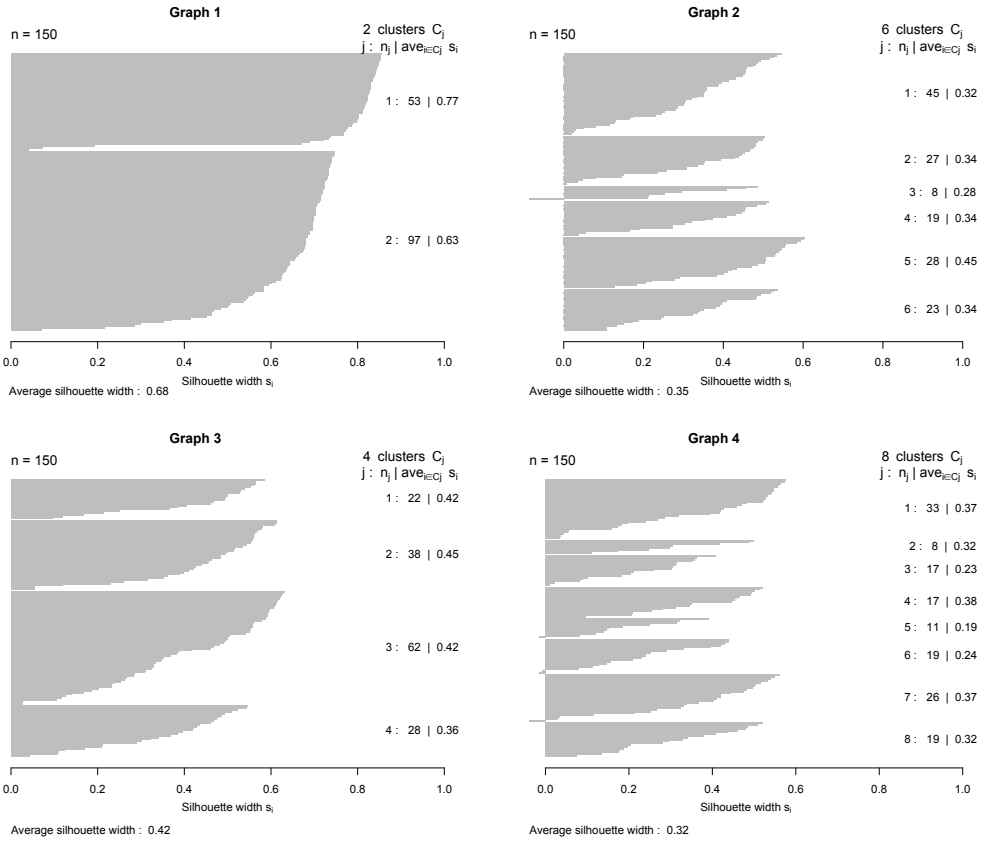
**Assignments submitted up to 24 hours late will incur a 30% penalty.**
**Assignments submitted more than 24 hours late will receive a zero grade.**

**Answer all questions. Not all questions carry equal marks.**
**All graphs must be labelled (including axes).**

1. **(4 MARKS)** Consider the four silhouette plots shown below:

   (a) What is the best number of clusters to choose and why?

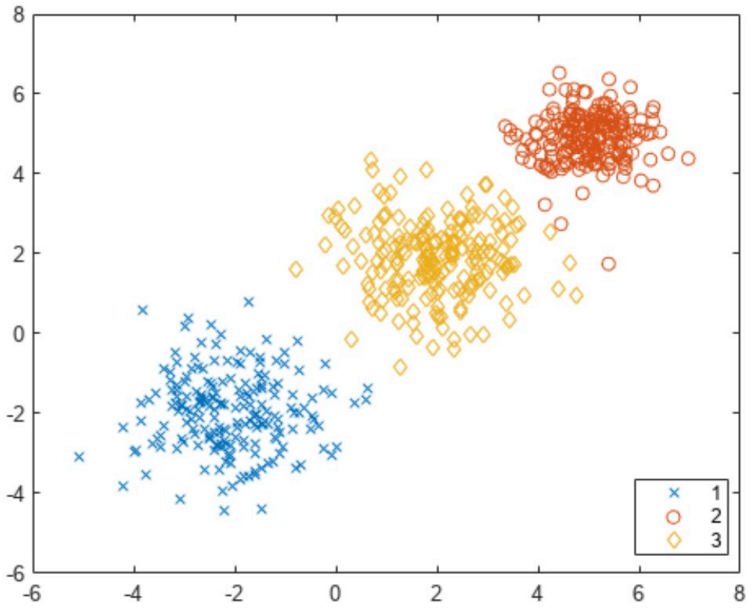
   (b) What is the second best number of clusters to choose and why?



2. **(6 MARKS)**

   (a) Describe $k$-means clustering and $k$-medoids clustering.

   (b) Would you expect $k$-means clustering to work reasonably well on the 3 clusters depicted below? Explain your answer.



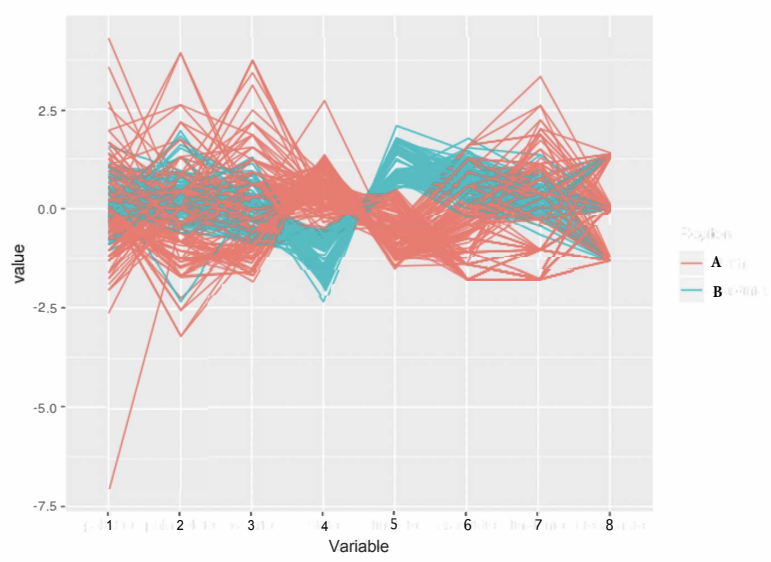

3. **(9 MARKS)** Using the finch dataset from the dynRB package, complete the following tasks: (Note: you may need to load additional libraries to answer the questions.)

   (a) Create box and whisker plots for body length (BodyL) against Species, and for wing length (WingL) against Species. Display the 2 graphs in one image, placing the plot for body length on top, and the plot for wing length underneath, (i.e. image with 1 column, 2 rows).

   (b) Create a parallel co-ordinates plot using all 9 predictor variables. The plot should be colour coded by Species.

4. **(1 MARK)** Consider the Parallel Co-ordinates plot below; it displays 2 different types of response, Response A and Response B, and 8 predictor variables.

   (a) Name a good predictor variable for separating out the responses.




---

**Assignment Standards**

- LaTeX is strongly recommended but not strictly required. The use of Markdown in R studio is also recommended.
- Submit your assignment as one **.pdf document**. **All R code should be included and organized either at the end of the assignment or inline (if using R Markdown).**
- Eleven-point font (times or similar) must be used with 1.5 line spacing and margins of at least 1 inch all around.
- Do not include a title page. The title and your name should be printed at the top of the first page.
- Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.

---