Rock Crabs 2

Parallel axes

This is the **second** in a series of questions on the exploration of some data on rock crabs.

The context for the data as well as the data itself is explained in more detail in the **background** document accompanying the assignment.

Since you will be using loon's interactive graphics to explore the data, you might also want to occasionally save some plots from your interactive analysis to include in your solution(s). Information on how to do this appears in the document **SavingLoonPlots**. Be sure to read that document (and even its "Rmd" version) before attempting to save loon plots.

25 marks

a. *(2 marks)* Parallel coordinate plots. Change the axes layout on `sa1` (scaling = variable) to `parallel` from `radial` and adjust the window shape to be wider than it is tall.

   Can you distinguish groups in this interactive parallel coordinate display? Explain why you can or cannot.

b. Consider the following function. It takes a data set and multiplies every other column by -1, thus reversing the order of the magnitudes of its contents.

```
flipAxes <- function(data) {
                dataCopy <- data
                for (i in 1:ncol(data)) {
                    if (i%%2 == 0) dataCopy[,i] <- (-1) * dataCopy[,i]
                    }
                dataCopy
                }
```

   Suppose `data` is the data set you used in the construction of `sa1`(i.e. with every variate appearing beside every other variate), construct a `parallel` axes display of `flipAxes(data)`, linking this display as part of the `linkingGroup` "lepto". Note that the argument `axesLayout` can be used to choose the "parallel" axes layout.

   i. *(2 marks)* Provide your code for producing the parallel axes system for the flipped axes.

      Assign the display to the symbol `sa3`.

      Display the resulting plot.

   ii. *(3 marks)* What geometric structure(s) do the points appear to lie near in the 5 dimensional space?

      What features of the display allow you to conclude this?

   iii. *(3 marks)* Consider the line segments between the parallel axes of "rear width (mm)" and "carapace length (mm)". The configuration of the line segments suggests that the corresponding points in Cartesian coordinates lie near two different lines.

      - What feature of the line segments for these parallel axes suggest the two lines in the Cartesian axis system?
      - Are the two lines parallel in the Cartesian system? Why or why not?
      - What is the sign of the slope of the line(s) in the Cartesian system?

   iv. *(4 marks)* Make sure all lines in `sa3` are the same colour.

      Focussing on area between the two parallel axes "rear width (mm)" and "carapace length (mm)", select all line segments (as best you can) which lie on one of the Cartesian lines. Colour this group a different colour from the rest. Select the group again (by colour) and deactivate them from the plot.

The rest of the lines should appear to lie on the other Cartesian line. If any do not, select these, colour them the same as the first group you selected and deactivate them as well.

The remaining lines should now appear as a single colour group near the second Cartesian line in the plane of "rear width (mm)" and "carapace length (mm)".

The code `sa3['color']` will return the colours that are assigned to every curve in `sa3`. The code `unique(sa3['color'])` should return only two different colours. These colours can be used to identify the row numbers in each group you have identified.

- How many points have you identified in each group and what are their row numbers in `lepto` (i.e. which points have you assigned to the two groups)?

  Show the code you used to determine the groups.

  Call the groups of row numbers `group1` and `group2`.

v. *(4 marks)* Have your parallel axes plot show only the second group you created.

To get this you may need to first reactivate the cases so that your parallel axes plot again shows all lines. There should be only two colours showing the two groups you created. Select `by color` the first group you created. Deactivate this group so that only the second group remains in the plot.

Now look at the line segments appearing between the two parallel axes "carapace width (mm)" and "body depth (mm)". For this group, at least, it would appear that the points in the corresponding Cartesian plane appear near two lines.

- Are these lines parallel in the plane? Why or why not?
- Select the segments which appear in one of the lines and colour these a third colour (i.e. different from the previous two).
- The second group is now divided into two subgroups (by colour). These two subgroups together form either `group1` or `group2` from the previous question above. Suppose they are subgroups of your previous `group1` (if not, change all 1s to 2s below), then we will call these `subgroup1a` and `subgroup1b`.

- Show the code you use to create the subgroups, give the number in each subgroup, and the row number from `lepto` for each subgroup.

vi. *(4 marks)* Repeat (v) above but this time divide the other major group into two subgroups.

In the end you should have 4 colours in total each identifying a different subgroup (2 from part v and 2 more from subdividing the remaining major group).

Answer all questions raised in part (v) but now for the remaining major group you first identified.

vii. *(3 marks)* Recall from the background material that a 1974 paper suggested that the species "Leptograpsus variegatus" should in fact be divided into two different species, each one being identified largely by the crab's colour (blue or orange).

In `loon.data`, there is also a data set called `crabSpecies` that contains the proposed species (variate `Species`) as well as the `Sex` of each of the 200 crabs in `lepto`.

Using the groups you identified in your parallel axes analysis, determine whether your findings are consistent or not with this conjecture. You can do this by simply examining `summary(crabSpecies[subgroup, ])` for each subgroup you created.

Describe your findings. That is, describe how well you did in defining the subgroups.