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.

   There is no obvious pattern, as there is no xvalues where all lines pass through, or lines that are parelel,

   there is weak pattern at the bottom of graph where the lines seem to have positive coefficients.

   we also notice some points have significantly different rear vs length ratio

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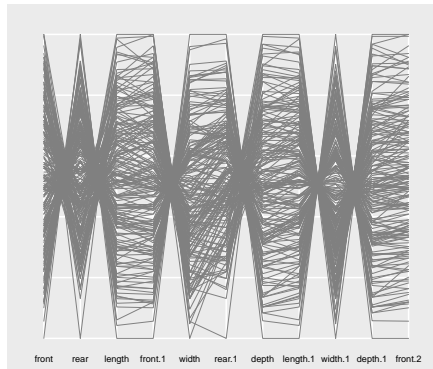
   Suppose `data` is the data set you used in the construction of `sa1`(i.e. with every variate appearing beside every other variate), construct a `parallel` axes display of `flipAxes(data)`, linking this display as part of the `linkingGroup` "lepto". Note that the argument `axesLayout` can be used to choose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

```
library(loon)
library(loon.data)
library(gridExtra)
library(loon.data); data("lepto")
library(PairViz)
ar = eulerian (5)
sa3 <- l_serialaxes(flipAxes(lepto)[,ar] ,color = "grey", linkingGroup = "lepto", axesLayout = "parall

plot(sa3)
```

front rear length front.1 width rear.1 depth length.1 width.1 depth.1 front.2

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?

they look like parrelell lines. As there are places between features where all lines meet, indicating a line with negative coefficient. There are also roughly parallel lines between some coefficients, indicating a positive coeficient line.

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?

almost all lines in the parrellel axis meet at one of two dots between rear and length

- Are the two lines parallel in the Cartesian system? Why or why not?

no, there are intersection of these lines (lines that go through both dots mentioned above)

- What is the sign of the slope of the line(s) in the Cartesian system?

negative

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.

```
g1lab = unique(sa3['color'])[1]
g2lab = unique(sa3['color'])[2]

g1.ind = sa3['color'] == g1lab
sum(g1.ind)
group1 = (1:200)[g1.ind]
```

2

```
g2.ind = sa3['color'] == g2lab
sum(g2.ind)
group2 = (1:200)[g2.ind]
group1
group2
```

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

group1: [1] 1 2 4 5 6 7 9 11 12 13 15 17 18 20 23 27 31 37 40 41 42 43 45 47 49 50 52 53 54 56 [31] 57 60 62 63 64 66 67 75 77 79 82 85 86 87 88 93 97 98 100 101 105 108 109 110 113 114 115 116 117 120 [61] 124 126 128 129 132 138 139 144 145 149 153 156 157 158 161 162 165 167 169 172 173 174 175 177 180 181 183 185 186 189 [91] 190 192 195 197 198 199 200

length =97

group2: [1] 3 8 10 14 16 19 25 26 28 29 32 34 35 39 44 46 51 59 61 65 68 69 70 73 74 78 83 84 89 90 [31] 95 96 103 104 106 107 112 123 125 127 131 133 134 135 136 140 142 143 146 148 150 151 155 159 160 164 166 170 171 176 [61] 178 179 184 187 188 191 196 21 22 24 30 33 36 38 48 55 58 71 72 76 80 81 91 92 94 99 102 111 118 119 [91] 121 122 130 137 141 147 152 154 163 168 182 193 194

length = 103

    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?

yes, there is almost no in common points in these two groups

- 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.

```
g1a.lab = unique(sa3['color'])[1]
g1b.lab = unique(sa3['color'])[3]

group1a.ind = sa3['color'] == g1a.lab
sum(group1a.ind)
group1a = (1:200)[group1a.ind]

group1b.ind = sa3['color'] == g1b.lab
sum(group1b.ind)
group1b = (1:200)[group1b.ind]

print("1a")
group1a
print("1b")
group1b
print("2")
```

```
group2
print("1")
group1
```

[1] "1a" [1] 1 2 15 18 20 23 27 31 37 41 42 43 47 53 56 60 62 63 66 67 79 85 86 87 88 100 101 105 110 113 [31] 114 115 116 117 124 126 128 129 132 144 145 149 169 174 177 180 183 185 186 190 192 195 197 198 199 200

length is 56

[1] "1b" [1] 4 5 6 7 9 11 12 13 17 40 45 49 50 52 54 57 64 75 77 82 93 97 98 108 109 120 138 139 153 156 [31] 157 158 161 162 165 167 172 173 175 181 189

length is 41

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

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

- Are these lines parallel in the plane? Why or why not?

yes, there is little intersection of points in these two lines. But these two parellel lines are close, it's not very different from one single line

- 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`.

```
g2a.lab = unique(sa3['color'])[2]
g2b.lab = unique(sa3['color'])[4]

group2a.ind = sa3['color'] == g2a.lab
sum(group2a.ind)
group2a = (1:200)[group2a.ind]

group2b.ind = sa3['color'] == g2b.lab
sum(group1b.ind)
group2b = (1:200)[group2b.ind]

print("group2a")
group2a
print("length")
sum(group2a.ind)

print("group2b")
group2b
print("length")
sum(group2b.ind)
```

```
[1] "group2a"
 [1]    3    8   10   14   16   19   25   26   28   29   32   34   35   39   44   46   51   59   61   65   68   69   70   73
[31]   95   96  103  104  106  107  112  123  125  127  131  133  134  135  136  140  142  143  146  148  150  151  155  159
[61]  178  179  184  187  188  191  196
[1] "sum"
[1] 67
[1] "group2b"
 [1]   21   22   24   30   33   36   38   48   55   58   71   72   76   80   81   91   92   94   99  102  111  118  119  121
[31]  154  163  168  182  193  194
[1] "sum"
```

```
[1] 36
```

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.

```r
library(loon.data); data("crabSpecies")
crabSpecies

blue.ind = (1:200)[  crabSpecies$Species=="blue" ]
orange.ind = (1:200)[  crabSpecies$Species=="orange" ]
male.ind = (1:200)[  crabSpecies$Sex=="male" ]
female.ind = (1:200)[  crabSpecies$Sex=="female" ]

#blue male
bm = intersect(blue.ind, male.ind)
bf = intersect(blue.ind, female.ind)
om = intersect(orange.ind, male.ind)
of = intersect(orange.ind, female.ind)

m=matrix(0, nrow = 4, ncol = 4)
rownames(m) = c("bm","om","bf","of")
colnames(m) = c("1a","1b","2a","2b")
m[1,1] = length(intersect(bm, group1a))
m[1,2] = length(intersect(bm, group1b))
m[1,3] = length(intersect(bm, group2a))
m[1,4] = length(intersect(bm, group2b))


m[3,1] = length(intersect(bf, group1a))
m[3,2] = length(intersect(bf, group1b))
m[3,3] = length(intersect(bf, group2a))
m[3,4] = length(intersect(bf, group2b))

m[2,1] = length(intersect(om, group1a))
m[2,2] = length(intersect(om, group1b))
m[2,3] = length(intersect(om, group2a))
m[2,4] = length(intersect(om, group2b))

m[4,1] = length(intersect(of, group1a))
m[4,2] = length(intersect(of, group1b))
m[4,3] = length(intersect(of, group2a))
m[4,4] = length(intersect(of, group2b))
m


length( intersect(male.ind, c(group1)))
length( intersect(female.ind, c(group1)))

length( intersect(blue.ind, c(group1)))
```

```
length( intersect(orange.ind, c(group1)))


length( intersect(male.ind, c(group2)))
length( intersect(female.ind, c(group2)))


length( intersect(blue.ind, c(group2)))
length( intersect(orange.ind, c(group2)))
```

our matrix:

$$\begin{pmatrix} & 1a & 1b & 2a & 2b \\ bluemale & 40 & 0 & 10 & 0 \\ orangemale & 1 & 39 & 0 & 10 \\ bluefemale & 14 & 0 & 34 & 2 \\ orangefemale & 1 & 2 & 23 & 24 \end{pmatrix}$$

when grouping group1 = male, group2 = female, we grouped about 80% data correct

for the subspecies we are most successful in grouping orange male, we only grouped 2 data out of 100 incorrectly

we did worse on blue female, we only group half of the data correctly

we grouped about 2/3 of data correctly, for other sub categories