Kholofelo Sethebe 3137355

**Question 2 (6 marks)**

ii) Explain each figure using your knowledge of how the ‘landscape was created’.

1. The species richness

* Species richness is the number of different species represented in an [ecological community](https://en.wikipedia.org/wiki/Community_(ecology)), landscape or region. The diagram shows that more of different species were found in the middle of the sampling environment (This can be a result of most paper shapes not being controlled by the generated wind, as the wind blew them upwards and back down allowing them to land in the middle and only allowing the different shapes to mix in the air, and the sweets being too heavy to be controlled by the generated wind and the carpet not allowing the sweets to bounce outwards).

1. The Shannon-Weaver

* The Shannon diversity index (H) is another index that is commonly used to characterize species diversity in a community.  Like Simpson's index, Shannon's index accounts for both abundance and evenness of the species present.
* Looking at the data, it shows that the species were more diverse in the middle of the sampling environment and this may result from the same reason mentioned above. The diagram again shows that the abundance of species decreased with increasing x-coordinates (This resulted from the fact that the species distributors and wind generator were standing too close to 50 and 100 x coordinates and next to 350cm of the Y- Coordinates.

1. The Simpson’s Index.

* Simpson's Diversity Index is a measure of diversity which takes into account the number of species present, as well as the relative abundance of each species. As species richness and evenness increase, so diversity increases.
* The data shows that the sweets species were more diverse in the middle and resulting from them being too heavy to be affected by the generated wind, and not being able to bounce harder due to the ground being carpet. Therefore that only allowed most of them to land in the middle. Most of the different paper shapes species were also more diverse and abundant in the middle of the sampling environment. Most of the crumpled species were able to disperse nicely and were found everywhere, and this was due to the fact that their weight was enough for the generated wind to affect them, same as the “elephants”.

**Question 3 (8 marks)**

iii) Briefly describe a few of the more ‘obvious’ positive and negative species associations. If possible, add an ‘ecological’ explanation for why this is so. (4)

Magenta = Dissimilarity close to 0 (maximum similarity)

Cyan = Dissimilarity close to 1 (minimum similarity)

* The differences are due to the log transformation. In the untransformed distance matrix, small differences in abundant species have the same importance as small differences in species with few individuals.
* a value of 0 means that the two communities we are comparing share all their species, while a value of 1 means they share none.
* As the both diagrams shows, the ordered dissimilarity matrix contain species with maximum dissimilarity in the middle and the, while in the dissimilarity matrix, species with maximum dissimilarity are scattered around the sampling environment.

**Question 4 (20)**

1. Undertake a PCA on the species data. Comprehensively explain the various (all) components of the summary () of the PCA object. (20)

* Total Inertia Proportion = 5.421 => “Inertia” is either the sum of the variances of the variables (PCA on a covariance matrix) or, as in this case (PCA on a correlation matrix), the sum of the diagonal values of the correlation matrix, i.e. the sum of all correlations of the variables with themselves, which corresponds to the number of variables (11 in this example).
* Unconstrained Inertia Proportion = 5.421 => In PCA, the analysis is unconstrained, and so are the results.
* Eigenvalues: PC1 = 1.3536 PC2 = 0.8904 PC3 = 0.8032 .These are measures of the importance (variance) of the axes. They can be expressed as Proportions Explained, or proportions of variation accounted for, by dividing them by the total inertia.
* General scaling constant of score = 3.478191, “Scaling” refers to the way ordination results are projected in the reduced space for graphical display. There is no single way to optimally display objects and variables together in a PCA biplot, i.e. a plot showing two types of results, here the sites and the variables. Two main types of scaling are generally used. Each of them has properties that must be kept in mind for proper interpretation of the biplots.
* Species scores: coordinates of the arrow heads of the variables. For historical reasons, response variables are always called “species” in vegan, no matter what they represent.
* Site scores: coordinates of the sites in the ordination diagram. Objects are always called “Sites” in vegan output files.

1. Provide plots of the PCA. How many axes must be retained, and why? What patterns become visible from the biplot? Explain your findings in detail. (10)

* Only 6 axes must be retained (PC1 – PC6) because their Eigenvalues fall under average and their information can be used alone without the others.
* Scaling 1 - distances among objects (sites) in the biplot are approximations of their Euclidean distances in multidimensional space; the angles among descriptor (species) vectors are meaningless. I chose this scaling with the **main interest being to interpret relationships among objects.** The diagram shows that the sites shoe not pattern as they scattered around. There are no sites pulling in different direction which. All sites are pulling in the positive direction( I am not if that is clear ☹)
* Scaling 2 - distances among objects in the biplot are not approximations of their Euclidean distances; the angles between descriptor (species) vectors reflect their correlations. I chose this scaling with the **main interest focuses on the relationships among descriptors (species).** The diagram shows that most species are clustered together,besides the once pulling outwards. The center of the diagram show many species associated together. There is no species pulling in different direction in this diagram, which means that the species do not have an enemy (no one will increase with a decrease in another) which is understandable due to the data we had.

**Question 5 (10)**

1. What patterns become visible from the ordination plot? Explain your findings in detail. (10)

Function goodness.meta MDS finds a goodness of fit statistic for observations (points). This is defined so that sum of squared values is equal to squared stress. Large values indicate poor fit. The diagram shows most large values (being in the centre). For instance, value 17 and 16, 11 and 5 ect.

Function stressplot is a wrapper to [Shepard](http://cc.oulu.fi/~jarioksa/softhelp/MASS/html/Shepard.html) function in MASS package. It plots ordination distances against original dissimilarities, and draws a step line of the nonlinear fit. In addition, it adds to the graph two correlation-like statistics on the goodness of fit. The nonmetric fit is based on stress *S* and defined as *sqrt(1-S^2)*. The “linear fit” is the correlation between fitted values and ordination distances. The diagram shows that the fit is perfect as there is perfect rank order coordination between the two axis and the points are falling on monotonic line. The r2 is close to 1, which indicates that the model explains all the variability of the response data around its mean.

**Question 6 (40)**

* Do a constrained analysis on the data, and comprehensively explain all the findings as per the of the summary () of the NMDS object. (20)
* The constrained fraction is the amount of variance of the **Y** matrix explained by the explanatory variables.
* Eigenvalues and their contribution to the variance: this analysis yielded 12 canonical axes (with eigenvalues labelled RDA1 to RDA12) and 16 additional, unconstrained axes for the residuals (with eigenvalues labelled PC1 to PC16).
* *Accumulated constrained eigenvalues*: these are cumulative amounts of variance expressed as proportions of the total *explained* variance, as opposed to their contribution to the *total* variance described above.
* Species scores are the coordinates of the tips of the vectors representing the response variables in the bi- or triplots. As in PCA, they depend on the scaling chosen.
* Site scores (weighted sums of species scores): coordinates of the sites as expressed in the space of the response variables Y.
* Site constraints (linear combinations of constraining variables): coordinates of the sites in the space of the explanatory variables X. These are the fitted site scores.
* *Biplot scores for constraining variables*: coordinates of the tips of the vectors representing the explanatory variables. These coordinates are obtained as follows: correlations are computed between the explanatory variables and the fitted site scores, and then these correlations are transformed to produce the biplot scores.
* Run the necessary permutation tests. Explain the outcomes. (6)

## Triplots of the rda results (wa scores)

## Site scores as weighted averages (vegan's default)

* Produce the necessary biplots. Explain the findings. (14)

## Triplots of the rda results (lc scores)

## Site scores as linear combinations of the environmental variables

* These triplots show that the Carpet and cardboard play an important role in the dispersion of the species along the sites.
* Scaling 1 shows that beans and more species were associated more with the carpet and less species were associated with the cardboard
* Scaling 2 shows b square and s­-rect were associated more with raw 18, and more species were clustered together in the centre.

## Triplots of the rda results (wa scores)

## Site scores as weighted averages (vegan's default)

**Question 7 (10)**

Write down your understanding of the ‘ecology’ of the landscape that was sampled and analysed in the various steps, above.

The floor of the sampling environment was covered entirely by carpet which made it hard for species to bounce off hard. There was a box incline in the middle of the sampling environment that acted as a bridge and there was no species that landed on top of the bridge, this may result from the incline being higher than the ground and the wind pushing all the species off. It was observed that most of the non crumpled species landed in the middle of the sampling environment, and this can be due to the fact that the paper turn to not have much of a direction when blown away, and being too light allows them to blow upwards and back down instead of blowing outward. The crumpled shapes and “elephants” were the only ones that showed a pattern of moving outwards and being close to the edges, and this can be resulting from the fact that crumpling the shapes gave them more weight than the normal, even though the weight was not too heavy to not get affected by the wind, they were not light either, therefore they were perfect to be affected by the wind. On the other hand, the generated wind did not have any effect on the sweets species due to them being heavy in weight, their dispersal came from them bouncing off the incline box in the middle and slightly bouncing off the ground. Hence most of them were found in the middle. If the floor did not have carpet on, it was going to be very easy for the sweets to bounce off and actually disperse in all directions. Most species were found between 50 and 100 with a decrease in species richness on both 150 and 200; reason being that people that were dispersing the species were standing too close to 50 and 100 and the wind generator (Prof) was standing there as well, therefore most species ended up landing right next to where they were dispersed.