### 1. Exploring the data

##### Load libraries

# install these if not already present on your machine by uncommenting this

# install.packages(c("PieGlyph", "DImodels", "DImodelsVis",
# "dplyr", "ggplot2"))

# Load necessary libraries
library(dplyr) # for data manipulation
library(ggplot2) # for standard visulisations
library(PieGlyph) # for creating pie-glyphs
library(DImodels) # for fitting Diversity-Interactions models
library(DImodelsVis) # for interpreting Diversity-Interactions models

##### Read data

# Change your working directory
# setwd()

# Uncomment this if you wish to choose the file manually
# raw\_data <- read.csv(file.choose())

# Read data
model\_data <- read.csv("Dataset\_4species\_simulated.csv")

# Summary of the data
summary(model\_data)

# View a snippet of data
head(model\_data)

##### Visualisations to explore data

###### Distribution of the response

# Create plot template
ggplot(# data is used to specify the data to be used for this visualisation
 data = model\_data,
 # aes is used to specify the variables to map to various plot components
 # for example what variable to show on the x-axis or y-axis
 aes(# response if mapped on x-axis
 x = yield,
 # after\_stat is internal keyword to clarify we want density on y-axis instead of counts
 y = after\_stat(density))) +
 # geom functions are used to specify type of visualisation
 # geom\_histogram means plotting a histogram
 geom\_histogram(fill = "darkolivegreen3") +
 # Overlaying the histogram with a density curve
 geom\_density(colour = "steelblue2", linewidth = 1) +
 # Adjust the theme of the plot
 theme\_classic(14) +
 # Changing axes labels, title, subtitle, etc. of the plot
 labs(x = "Yield (t/ha)", y = "Density",
 title = "Histogram of response")

###### Distribution of the response by treatment

# We will create a separate histogram for each treatment level
ggplot(data = model\_data,
 aes(x = yield,
 y = after\_stat(density))) +
 geom\_histogram(fill = "darkolivegreen3") +
 geom\_density(colour = "steelblue2", linewidth = 1) +
 # facet\_wrap is used to specify creating a separate plot for each category of a given categorical variable
 facet\_wrap(~ treatment, labeller = label\_both ) +
 theme\_classic(14) +
 labs(x = "Yield (t/ha)", y = "Density",
 title = "Histogram of response by treatment")

##### Yield vs richness

ggplot(data = model\_data,
 # yield is shown on y-axis now with richness on the x-axis
 aes(x = richness, y = yield)) +
 # show observations as points; size adjusts the size of points
 geom\_point(size = 3) +
 # separate plot for each treatment
 facet\_wrap(~ treatment, labeller = label\_both ) +
 theme\_classic(14) +
 labs(x = "Richness", y = "Yield (t/ha)", fill = "Species",
 title = "Yield vs richness")

##### Yield vs different dimensions of biodiversity

# species name and associated colour scheme to be used throughout the document
species <- c("p1", "p2", "p3", "p4")
pie\_cols <- c("#006B4E", "#00D198", "#A24700", "#FF7609")

ggplot(data = model\_data,
 # yield is shown on y-axis now with richness on the x-axis
 aes(x = richness, y = yield)) +
 # Showing points as pie-glyphs instead of points to show relative proportions of species in each community
 # slices parameter is used to specify the columns names containing species proportions
 geom\_pie\_glyph(slices = c("p1", "p2", "p3", "p4"),
 colour = "black") +
 # colours of slices within pie-glyphs
 scale\_fill\_manual(values = pie\_cols) +
 # separate plot for each treatment
 facet\_wrap(~ treatment, labeller = label\_both ) +
 theme\_classic(14) +
 labs(x = "Richness", y = "Yield (t/ha)", fill = "Species",
 title = "Yield vs richness, species composition and abundances")