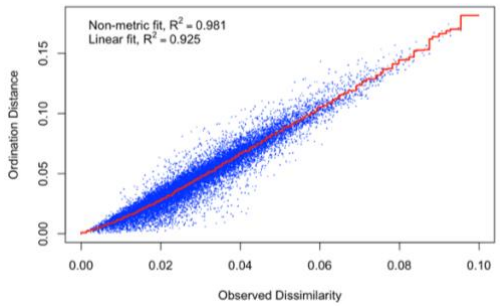
Coding sample of coral analysis – NMDS

```r
#load libraries and set seed
library(vegan)

library(readxl)
library(ggplot2)
set.seed(123)

#read master data and provide column names
master_data <- read.csv("Master datatable - Combined_current.csv")
colnames(master_data) <- c("sample_id", "family", "genus", "species", "region", "temp", "pCO2", "pCO2_grou
ps", "pH_sw", "H2CO3", "HCO3_minus", "CO3_2_minus", "TA", "DIC_sw", "net_calc", "Li_Ca", "Li_Ca_sd", "B_Ca
", "B_Ca_sd", "Mg_Ca", "Mg_Ca_sd", "Sr_Ca", "Sr_Ca_sd", "U_Ca", "U_Ca_sd", "Li_Mg", "Sr_U", "d11B", "d11B_
sd", "pH_cf", "DIC_cf")
trace_elements <- c("Li_Ca", "B_Ca", "Mg_Ca", "Sr_Ca", "U_Ca", "Li_Mg", "Sr_U")

#create function to extract significant p-values
extract_p_value <- function(data, element, temp) {
  temp_data <- subset(data, temp == temp)
  if (nrow(temp_data) > 1) {
    regression <- lm(reformulate("CO3_2_minus", response = element), data = temp_data)
    return(summary(regression)$coefficients[2, 4]) #return p-value if data is sufficient
  } else {
    return(NA) #return NA if data is insufficient
  }
}

#create an empty data frame to store p-values
regression_data_frame <- data.frame(
    element = character(),
    genus = character(),
    family = character(),
    temp = integer(),
    p_value = numeric(),
    stringsAsFactors = FALSE)

#for loop to obtain p-values for all genus at 28 and 31 degrees celsius
for (element in trace_elements) {
  element_data <- master_data[, c("family", "genus", "region", "temp", "CO3_2_minus", element)]
  data <- element_data[!is.na(element_data[[element]]) & !is.na(element_data[["CO3_2_minus"]]), ]
  for (genus in unique(data$genus)) {
    genus_data <- data[data[["genus"]] == genus,]
    p_value_result <- extract_p_value(genus_data, element, 28)
    regression_data_frame <- rbind(regression_data_frame, data.frame(
        element = element,
        genus = genus,
        family = unique(genus_data[["family"]]),
        temp = 28,
        p_value = p_value_result))
    p_value_result <- extract_p_value(genus_data, element, 31)
    regression_data_frame <- rbind(regression_data_frame, data.frame(
        element = element,
        genus = genus,
        family = unique(genus_data[["family"]]),
        temp = 31,
        p_value = p_value_result))
  }
}

#extract rows with significant p-values
significant_p_values <- regression_data_frame[regression_data_frame[["p_value"]] <= 0.05, ]
head(significant_p_values)

##    element       genus         family temp      p_value
## 21    B_Ca Siderastrea Siderastreidae   28 0.0008917985
## 22    B_Ca Siderastrea Siderastreidae   31 0.0008917985
## 23    B_Ca  Stylophora Pocilloporidae   28 0.0030373766
```

```
## 24    B_Ca  Stylophora Pocilloporidae   31 0.0030373766
## 45   Sr_Ca Siderastrea Siderastreidae   28 0.0019483064
## 46   Sr_Ca Siderastrea Siderastreidae   31 0.0019483064
```

```
#data subset for NMDS
nmds_data <- master_data[c("species", "pCO2_groups", "Li_Ca", "B_Ca", "Mg_Ca", "Sr_Ca", "U_Ca", "Li_Mg", "
Sr_U")]
nmds_data <- na.omit(nmds_data)

#create community matrix using bray-curtis dissimilarity method
community_matrix_all_data <- as.matrix(nmds_data[c("Li_Ca", "B_Ca", "Mg_Ca", "Sr_Ca", "U_Ca", "Li_Mg", "Sr
_U")])
nmds_all_data <- metaMDS(community_matrix_all_data, distance = "bray", k = 2)

stressplot(nmds_all_data)
```




```
#store data scores into a data frame
data_scores <- as.data.frame(scores(nmds_all_data, "sites"))
data_scores$species = nmds_data$species
data_scores$pCO2_groups = nmds_data$pCO2_groups
head(data_scores)
```

```
##         NMDS1        NMDS2    species pCO2_groups
## 1 -0.06195967 -0.002531977 tenuifolia         250
## 2 -0.02536985 -0.021462959 tenuifolia         250
## 3 -0.06683929 -0.003041270 tenuifolia         250
## 5 -0.07951625 -0.003375560 tenuifolia         400
## 6 -0.06894125  0.000727178 tenuifolia         400
## 7 -0.05930964  0.002682089 tenuifolia         400
```

```
#plot NMDS graph using ggplot2
ggplot(data_scores, aes(x = NMDS1, y = NMDS2)) + geom_point(size = 4, aes(shape = species, colour = as.fac
tor(pCO2_groups))) + theme(axis.text.y = element_text(colour = "black", size = 12, face = "bold"),
    axis.text.x = element_text(colour = "black", face = "bold", size = 12),
    legend.text = element_text(size = 12, face ="bold", colour ="black"),
    legend.position = "right", axis.title.y = element_text(face = "bold", size = 14),
    axis.title.x = element_text(face = "bold", size = 14, colour = "black"),
    legend.title = element_text(size = 14, colour = "black", face = "bold"),
    panel.background = element_blank(), panel.border = element_rect(colour = "black", fill = NA, size = 1.
2),
  legend.key=element_blank()) + labs(colour = "pCO2_groups")
```