Preliminary Analysis

#The biological hypothesis that are being tested are as follows: #1. The lower the depth the higher the abundance of plastics will be found. #2. The higher the salinity the higher the abundance of plastics will be found. #3. The higher the temperature the higher the abundance of plastics will be found.

#The dependent variable of this test is the plastic abundance #The independent variables of this test are environmental conditions (sea temperature, depth and salinity)

#The statistical analysis being used in this test is the Kendall Rank Correlation Test which is a non-parametric test used to measure the correlation of two variables #The statistical null hypothesis is there is no statistically significant relationship between the median abundance of plastics and the median environmental variables (sea temperature, depth and salinity). #The assumptions of the statistical analysis is: #1. The data is continuous or ordinal #2. The data is monotonic

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v purrr   0.3.4     v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.1.2

## Warning: package 'dplyr' was built under R version 4.1.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

#Import datasets (nettows_info.csv and plastics_info.csv)

```
nettows_info <- read_csv("data/nettows_info.csv")
```

```
## Rows: 171 Columns: 32
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (4): # VesselTripName, CollectedBy, DateUTC, NetTypeMeshSizeDimensions...
## dbl  (26): NetStation, Replicate, Duration, VesselSpeed, VesselDirection, Wi...
## time  (2): StartTimeUTC, EndTimeUTC
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
plastics_info <- read_csv("data/plastics_info.csv")
```

```
## Rows: 839 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): Type, Colour, PolymerType
## dbl (4): # Plastic, NetStation, Replicate, Length
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

#We want to create a new dataset containing only the data we need and take the average from the three replicates of each site. #We want to take the total plastic abunance at each site and the average sea temperature, depth and salinity at each site.

```
plastics_abund <- count(plastics_info, NetStation)

avtemp <- nettows_info %>%
  group_by(NetStation) %>%
  summarize(averaged.ST = mean(SeaTemperature))

avdepth <- nettows_info %>%
  group_by(NetStation) %>%
  summarize(averaged.depth = mean(Depth))

avsalinity <- nettows_info %>%
  group_by(NetStation) %>%
  summarize(averaged.sal = mean(Salinity))
```

#Compile all of the data into a single dataset.

```
list <- list(plastics_abund, avtemp, avdepth, avsalinity)
data <- Reduce(function(x, y) merge(x, y, all=TRUE), list)
```

#Now we can view the datasets #Table of Site, plastic abundance, average sea temperature, depth and sailinity

```
head(data, 5)
```

```
##   NetStation  n averaged.ST averaged.depth averaged.sal
## 1          1  3    15.00000      2619.4333     35.24667
```

```
## 2          2  7    15.30000      4633.6667     35.36667
## 3          3  4    15.26667      4383.6667     35.45667
## 4          4  3    15.10000      4863.3333     35.47867
## 5          5 62    13.16667       709.0667     35.37333
```

#All of the data is measured on a continuous scale

#We also will test for normality of the datasets

```
shapiro.test(data$n)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$n
## W = 0.73623, p-value = 2.043e-08
```

```
shapiro.test(data$averaged.ST)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$averaged.ST
## W = 0.86879, p-value = 6.089e-05
```

```
shapiro.test(data$averaged.depth)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$averaged.depth
## W = 0.85079, p-value = 5.162e-06
```

```
shapiro.test(data$averaged.sal)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$averaged.sal
## W = 0.87806, p-value = 0.0001137
```

#From the output, the four p-values are less than the significance level 0.05 implying that the distribution of the data are significantly different from normal distribution. #In other words, we can assume the data is NOT normally distributed.

#Create scatter plots of the environmental conditions vs the plastic abundance.

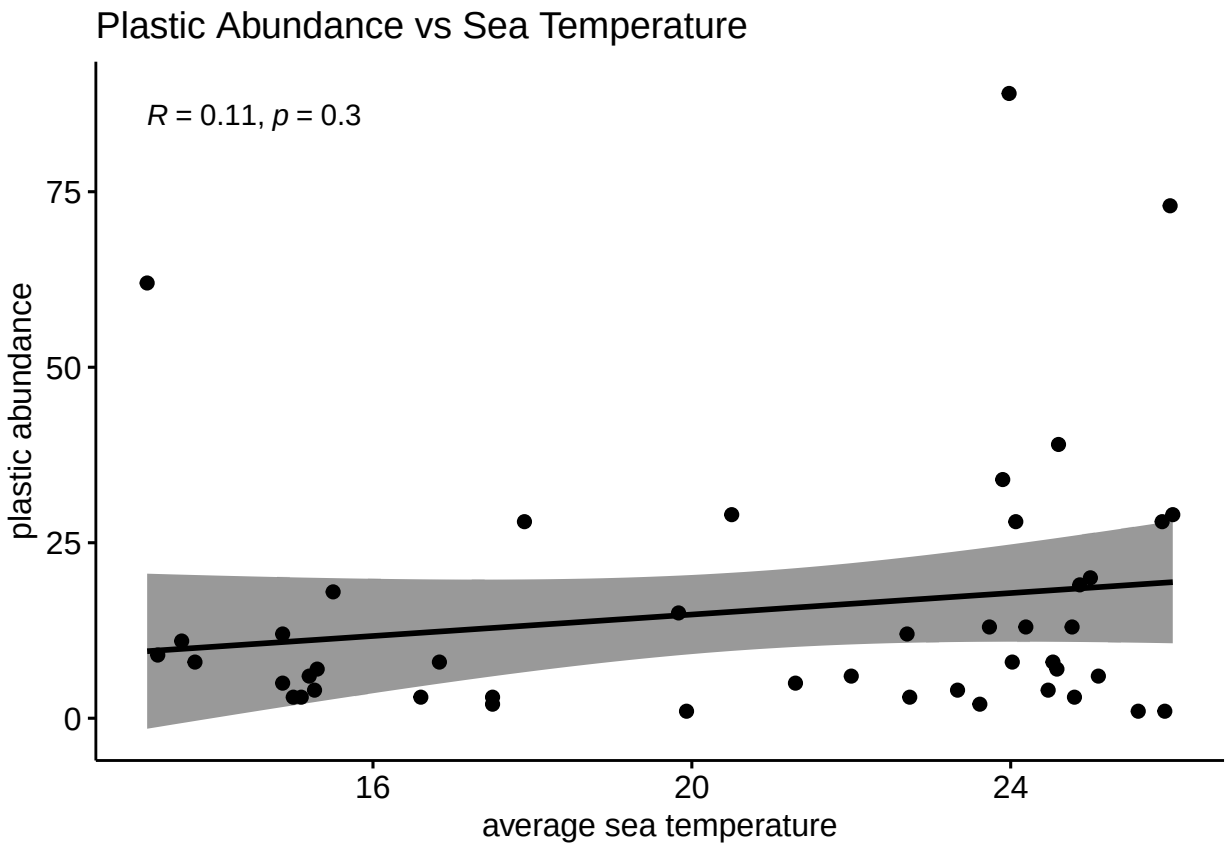
```
#Scatter plot of plastic abundance and average sea temperature
ggscatter(data, x = "averaged.ST", y = "n",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "kendall",
          title = "Plastic Abundance vs Sea Temperature",
          xlab = "average sea temperature", ylab = "plastic abundance")
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 11 rows containing non-finite values (stat_smooth).

## Warning: Removed 11 rows containing non-finite values (stat_cor).

## Warning: Removed 11 rows containing missing values (geom_point).
```



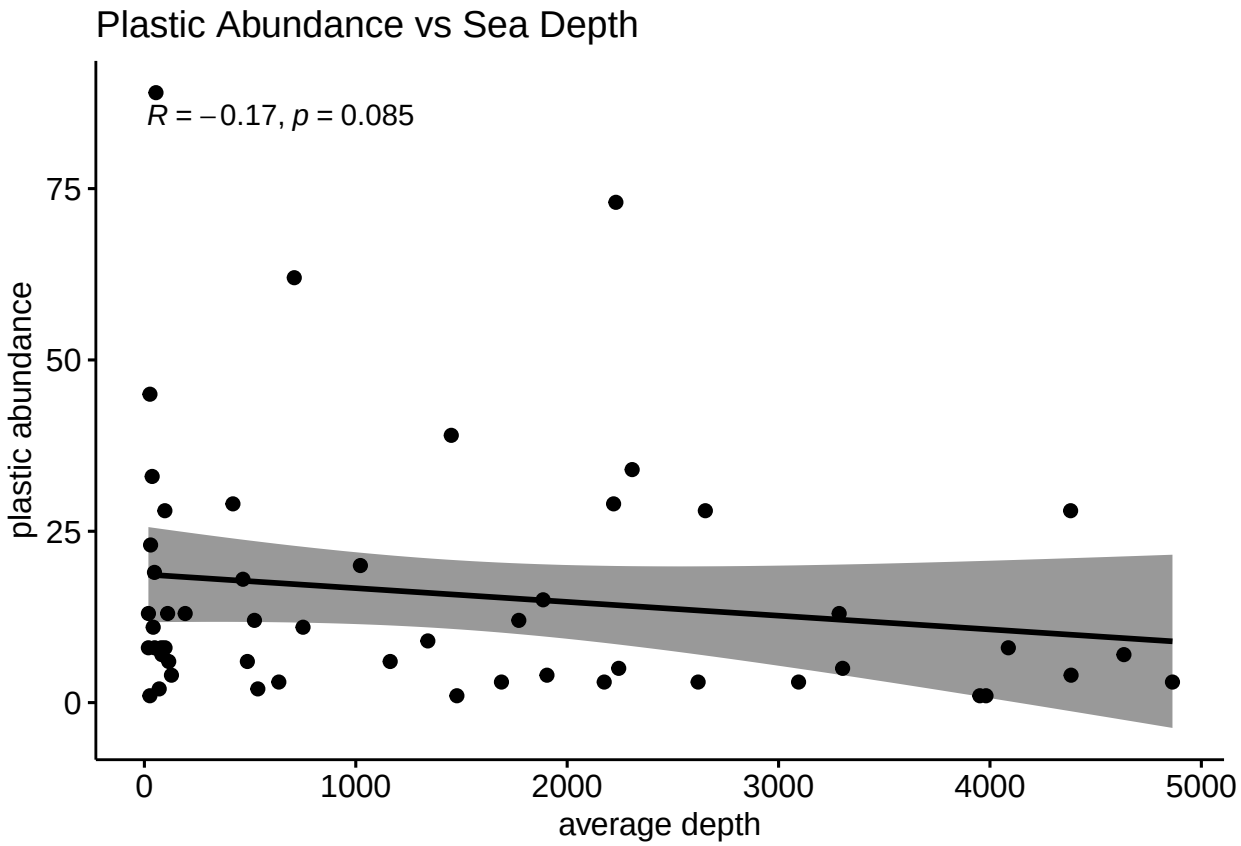

```
#Scatter plot of plastic abundance and average depth
ggscatter(data, x = "averaged.depth", y = "n",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "kendall",
          title = "Plastic Abundance vs Sea Depth",
          xlab = "average depth", ylab = "plastic abundance")
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 4 rows containing non-finite values (stat_smooth).

## Warning: Removed 4 rows containing non-finite values (stat_cor).

## Warning: Removed 4 rows containing missing values (geom_point).
```



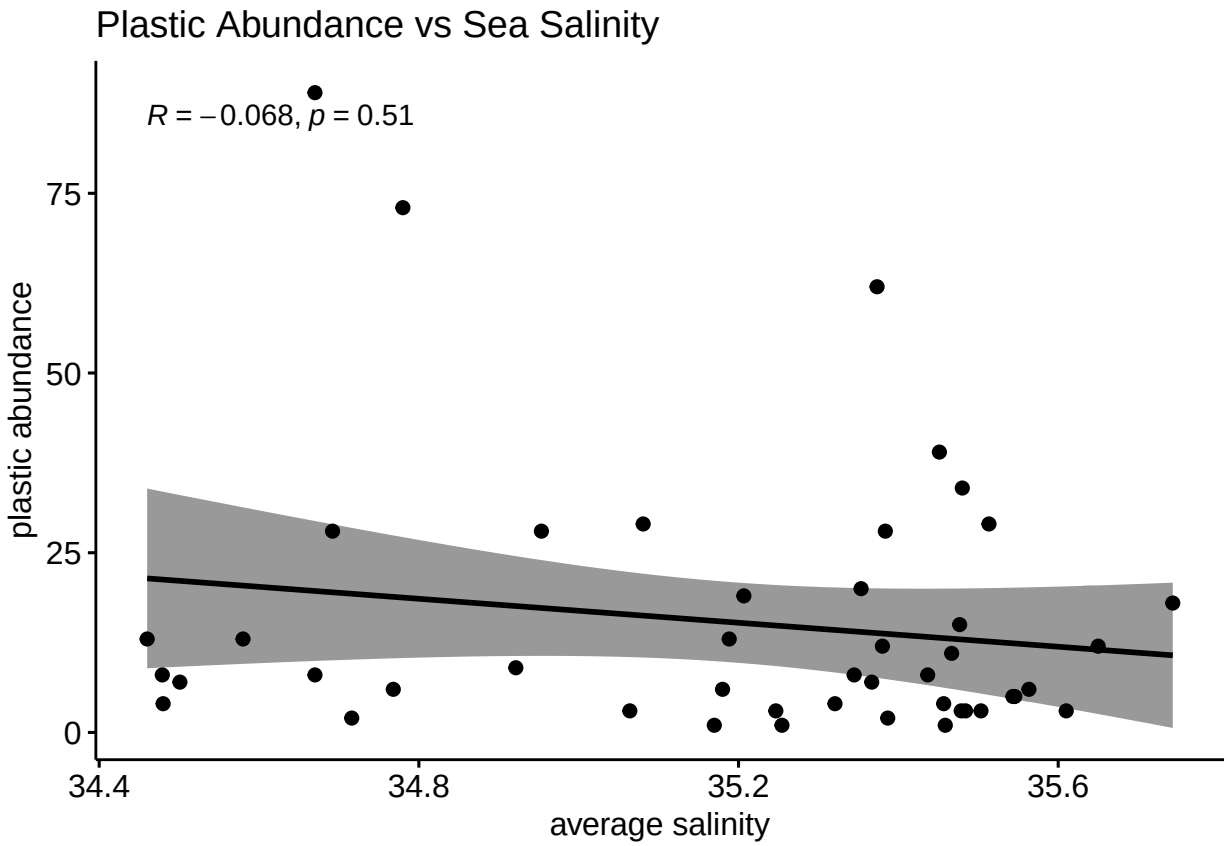

```
#Scatter plot of plastic abundance and average salinity
ggscatter(data, x = "averaged.sal", y = "n",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "kendall",
          title = "Plastic Abundance vs Sea Salinity",
          xlab = "average salinity", ylab = "plastic abundance")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 11 rows containing non-finite values (stat_cor).
```

```
## Warning: Removed 11 rows containing missing values (geom_point).
```



#From the above plots we can see that the above plots are linear and monotonic in relationship.

#Therefore, we can say that the data is not normally distributed, continuous in scale and have a monotonic relationship. #Therefore, all assumptions have been met to use the Kendall Rank Correlation Test.

```
temperature_kendall <- cor.test(data$n, data$averaged.ST,
                method = "kendall")
```

```
## Warning in cor.test.default(data$n, data$averaged.ST, method = "kendall"):
## Cannot compute exact p-value with ties
```

```
depth_kendall <- cor.test(data$n, data$averaged.depth,
                method = "kendall")
salinity_kendall <- cor.test(data$n, data$averaged.sal,
                method = "kendall")
```

```
## Warning in cor.test.default(data$n, data$averaged.sal, method = "kendall"):
## Cannot compute exact p-value with ties
```

```
temperature_kendall
```

```
##
##  Kendall's rank correlation tau
##
## data:  data$n and data$averaged.ST
```

```
## z = 1.0444, p-value = 0.2963
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.1085549
```

```
#tau = 0.1086, p = 0.2963
depth_kendall
```

```
##
##  Kendall's rank correlation tau
##
## data:  data$n and data$averaged.depth
## z = -1.7223, p-value = 0.08502
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##        tau
## -0.1657111
```

```
#tau = -0.1657, p = 0.08502
salinity_kendall
```

```
##
##  Kendall's rank correlation tau
##
## data:  data$n and data$averaged.sal
## z = -0.6551, p-value = 0.5124
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##         tau
## -0.06806058
```

```
#tau = -0.06806, p = 0.5124
```

#Plastic abundance and sea temperature do not have a sgignificant relationship (tau = 0.1086, p = 0.2963) #Plastic abundance and depth do not have a sgignificant relationship (tau = -0.1657, p = 0.08502) #Plastic abundance and salinity do not have a sgignificant relationship (tau = -0.06806, p = 0.5124) #All p-values obtained are greater than the significance level 0.05. #Therefore, the null hypothesis is not rejected. #Therefore, we can determine that there is no statistically significant relationship between plastics abundance and environmental conditions (sea temperature, depth and salinity)