Debris Graphs & Stats

2023-01-09

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 1.0.0   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.5.0   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(readxl)  
library(plotrix)

macroplastics <- read\_excel("~/Desktop/MB5001 Macroplastics Counts.xlsx")  
plastics\_mass <- read\_excel("~/Desktop/MB5001 Macroplastics Mass.xlsx")

trash\_categories <- macroplastics %>%   
 group\_by(Trash\_type) %>%   
 summarise(sum = sum(Count))

# ABUNDANCE

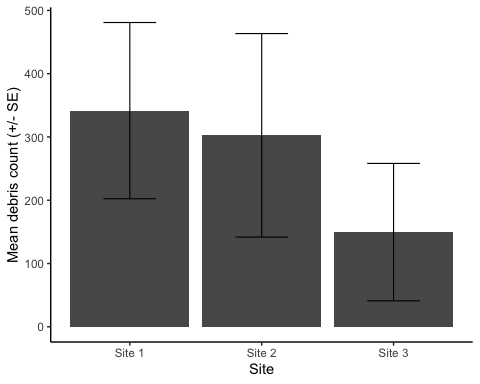
plastics\_summary <- macroplastics %>%   
 group\_by(Tidal\_line, Site) %>%   
 summarise(sum = sum(Count))

## `summarise()` has grouped output by 'Tidal\_line'. You can override using the  
## `.groups` argument.

mean\_for\_sites <- plastics\_summary %>%   
 group\_by(Site) %>%   
 summarise(mean = mean(sum), SE = std.error(sum))

ggplot(mean\_for\_sites, aes(x = Site, y = mean)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = mean - SE, ymax = mean + SE), width = 0.4, size = 0.4) +  
 labs(x = "Site", y = "Mean debris count (+/- SE)") +  
 theme\_classic()

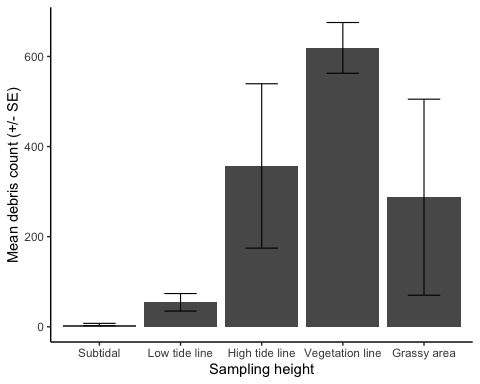
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.



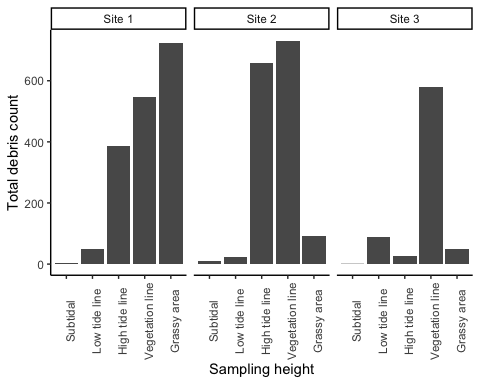
mean\_for\_heights <- plastics\_summary %>%   
 group\_by(Tidal\_line) %>%   
 summarise(mean = mean(sum), SE = std.error(sum))

mean\_for\_heights$Tidal\_line = factor(mean\_for\_heights$Tidal\_line, levels = c('Subtidal', 'Low tide line', 'High tide line', 'Vegetation line', 'Grassy area'))

ggplot(mean\_for\_heights, aes(x = Tidal\_line, y = mean)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = mean - SE, ymax = mean + SE), width = 0.4, size = 0.4) +  
 labs(x = "Sampling height", y = "Mean debris count (+/- SE)") +  
 theme\_classic()



plastics\_summary$Tidal\_line = factor(plastics\_summary$Tidal\_line, levels = c('Subtidal', 'Low tide line', 'High tide line', 'Vegetation line', 'Grassy area'))  
  
ggplot(plastics\_summary, aes(x = Tidal\_line, y = sum)) +  
 geom\_bar(stat = "identity") +  
 labs(x = "Sampling height", y = "Total debris count") +  
 facet\_wrap(~ Site) +  
 theme\_classic() +  
 theme(axis.text.x = element\_text(angle = 90))



## Difference in abundance between sites: chi-square goodness of fit test

site\_totals <- macroplastics %>%   
 group\_by(Site) %>%   
 summarise(total = sum(Count))

chisq.test(site\_totals$total, p = c(1/3, 1/3, 1/3))

##   
## Chi-squared test for given probabilities  
##   
## data: site\_totals$total  
## X-squared = 389.23, df = 2, p-value < 2.2e-16

Conclusion: there is a significant difference between the observed frequency of debris across the sites and the expected frequency (expected = there will be an equal abundance of debris across all sites)

## Difference in abundance between sampling heights: chi-square goodness of fit test

height\_totals <- macroplastics %>%   
 group\_by(Tidal\_line) %>%   
 summarise(total = sum(Count))

chisq.test(height\_totals$total, p = c(1/5, 1/5, 1/5, 1/5, 1/5))

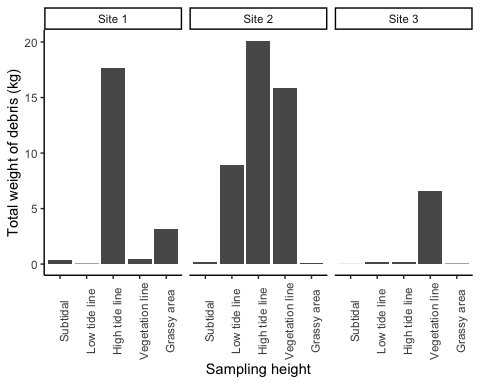
##   
## Chi-squared test for given probabilities  
##   
## data: height\_totals$total  
## X-squared = 2792.2, df = 4, p-value < 2.2e-16

Conclusion: there is a significant difference between the observed frequency of debris across the sampling heights and the expected frequency (expected = there will be an equal abundance of debris across all sampling heights)

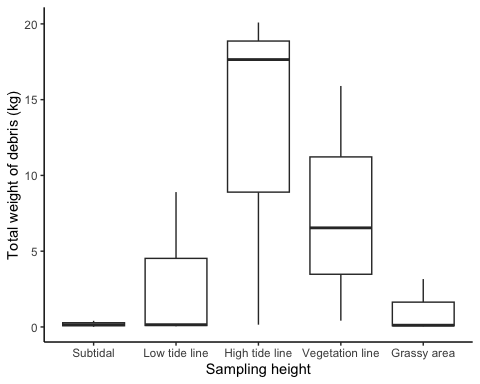
# MASS

plastics\_mass$Tidal\_line = factor(plastics\_mass$Tidal\_line, levels = c('Subtidal', 'Low tide line', 'High tide line', 'Vegetation line', 'Grassy area'))

ggplot(plastics\_mass, aes(x = Tidal\_line, y = Total\_weight\_kg)) +  
 geom\_bar(stat = "identity") +  
 facet\_wrap(~ Site) +  
 theme\_classic() +  
 theme(axis.text.x = element\_text(angle = 90)) +  
 labs(x = "Sampling height", y = "Total weight of debris (kg)")



ggplot(plastics\_mass, aes(x = Tidal\_line, y = Total\_weight\_kg)) +  
 geom\_boxplot() +  
 theme\_classic() +  
 labs(x = "Sampling height", y = "Total weight of debris (kg)")



plastics\_mass %>%   
 group\_by(Tidal\_line) %>%   
 summarise(mean = mean(Total\_weight\_kg))

## # A tibble: 5 × 2  
## Tidal\_line mean  
## <fct> <dbl>  
## 1 Subtidal 0.188  
## 2 Low tide line 3.03   
## 3 High tide line 12.6   
## 4 Vegetation line 7.62   
## 5 Grassy area 1.11

## Difference in mass of debris between sampling heights: kruskal-wallis test

kruskal.test(Total\_weight\_kg ~ Tidal\_line, data = plastics\_mass)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Total\_weight\_kg by Tidal\_line  
## Kruskal-Wallis chi-squared = 5.9111, df = 4, p-value = 0.2059

Conclusion: there is not a significant difference in mass of debris between sampling heights

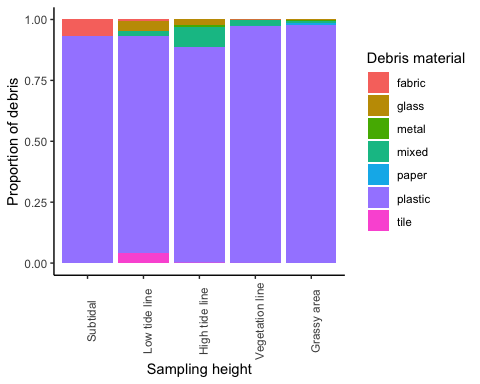
# MATERIAL

macroplastics$Tidal\_line = factor(macroplastics$Tidal\_line, levels = c('Subtidal', 'Low tide line', 'High tide line', 'Vegetation line', 'Grassy area'))

material\_height\_totals <- macroplastics %>%   
 group\_by(Tidal\_line, material) %>%   
 summarise(total = sum(Count)) %>%   
 mutate(percentage = total/sum(total)\*100)

## `summarise()` has grouped output by 'Tidal\_line'. You can override using the  
## `.groups` argument.

material\_height\_totals %>%   
 ggplot() +  
 geom\_bar(aes(x = Tidal\_line, y = total, fill = material), stat = 'identity', position = "fill") +  
 theme\_classic() +  
 labs(x = "Sampling height", y = "Proportion of debris", fill = "Debris material") +  
 theme(axis.text.x = element\_text(angle = 90))



## Whether all materials were found in equal proportions: chi-square goodness of fit test

material\_totals <- macroplastics %>%   
 group\_by(material) %>%   
 summarise(total = sum(Count)) %>%   
 mutate(percentage = total/sum(total)\*100)

chisq.test(material\_totals$total, p = c(1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7))

##   
## Chi-squared test for given probabilities  
##   
## data: material\_totals$total  
## X-squared = 20950, df = 6, p-value < 2.2e-16

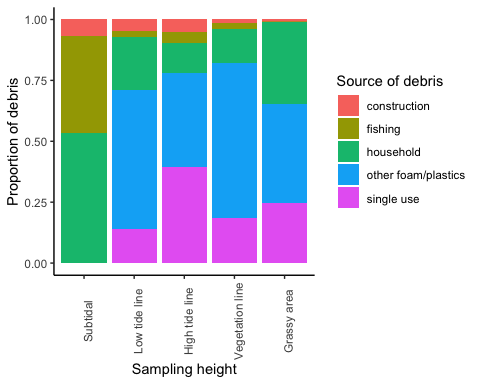
Conclusion: There is a significant difference between the observed probabilities and the expected probabilities, meaning that the 7 materials are not found in equal proportions (then refer to the graph, which clearly shows there’s more plastic than anything else, but we can’t call this conclusion (that it’s mostly plastic) significant since we didn’t run a test specifically about plastic being greater than others)

# CATEGORY/SOURCE

category\_height\_totals <- macroplastics %>%   
 group\_by(Tidal\_line, category) %>%   
 summarise(total = sum(Count)) %>%   
 mutate(percentage = total/sum(total)\*100)

## `summarise()` has grouped output by 'Tidal\_line'. You can override using the  
## `.groups` argument.

category\_height\_totals %>%   
 ggplot() +  
 geom\_bar(aes(x = Tidal\_line, y = total, fill = category), stat = 'identity', position = "fill") +  
 theme\_classic() +  
 labs(x = "Sampling height", y = "Proportion of debris", fill = "Source of debris") +  
 theme(axis.text.x = element\_text(angle = 90))



## Whether all categories/sources of debris were found in equal proportions: chi-square goodness of fit test

category\_totals <- macroplastics %>%   
 group\_by(category) %>%   
 summarise(total = sum(Count)) %>%   
 mutate(percentage = total/sum(total)\*100)

chisq.test(category\_totals$total, p = c(1/5, 1/5, 1/5, 1/5, 1/5))

##   
## Chi-squared test for given probabilities  
##   
## data: category\_totals$total  
## X-squared = 3200, df = 4, p-value < 2.2e-16

Conclusion: There is a significant difference between the observed probabilities and the expected probabilities, meaning that the 5 categories are not found in equal proportions (then refer to the graph, which shows it’s mostly household & single use than anything else, but we again can’t call this conclusion significant since we didn’t run a test specifically about these 2 categories being larger than the others)