MSc by Thesis-Rscript

Clare Collins

2023-08-10

R script structure and packages

Single # is a print of the console

Double ## are my notes

Assumes this markdown file is saved in the root folder, data is saved in ./data and items are written out to ./outputs

To shorten the knitted document, only published plots and visual comparisons will show up, everything else can be run through R, but the markdown is set not to show the results within the markdown document ", include=FALSE".

Install required packages

```
install.packages("tidyverse") (includes ggplot2 and dplyr)
install.packages("ggpubr")  ## for ggarrange - arranging plots and checking for normality - recommen
install.packages("rempsyc") ##for publication standard tables see <https://rempsyc.remi-theriault.co
install.packages("magittr")
install.packages("knitr") ##to create markdown document
install.packages("car")
install.packages("ggsignif")
install.packages("grid")
install.packages("devtools")
devtools::install_github("thomasp85/patchwork") # patches plots together, like ggarrange, but quickl
install.packages("ggtext")
install.packages("patchwork")
install.packages("egg") ## to layer graphs...but stops ggarrange, so remove where issues arise
```

Load packages

References for R and packages used

Citation and references - Bibtex for importing to Zotero

Literature Data

## Recent Interest in MPs

Import data "WoK_RelPubFishMPs.csv" and sort

Figure 1: Microplastics publications have increased more than the general publication rate

Create barchart to identify whether Microplastics publications have increased themselves, or inline with the increase in publication rate in general

Topic search "(*plastic OR *plastics) AND fish* AND (ingest* OR consum*)" within Web of Knowledge for Years 1933 (start)-2022 (end) Searched 2023-06-22 (n= 2367 in 135 categories) then refined by the top two WoK categories (Environmental Sciences (n= 1317 publications) and Marine Freshwater Biology (n= 705 publications)) compared to all publications in those categories 1,724,817 and 387,288 respectively per year to assess whether there is a relative increase in microplastics publications when considering the actual increase in all publications. No data before 1983 so this is the first year. No microplastics fish ingestion papers before 1990 so maybe worth limiting to this.

```
ggplot(Publication_Relative_Numbers, aes(x = Year, y = Relative_Publications))+
  geom_col(aes(fill=Category), colour="black", position = "dodge") +
  scale_fill_viridis_d()+
  scale_y_continuous(expand = c(0,0), labels = scales::percent)+
  ylab("Proportion of publications\non fish ingesting plastics\n")+
  xlab("\nPublication Year")+
  xlim(2000,2023)+
  theme(
      axis.ticks = element_line(colour = NA),
      axis.title.x = element_text(size = rel(1.1)),
      axis.title.y = element_text(size = rel(1.1)),
      legend.position = "top",
      legend.title = element_blank())
```
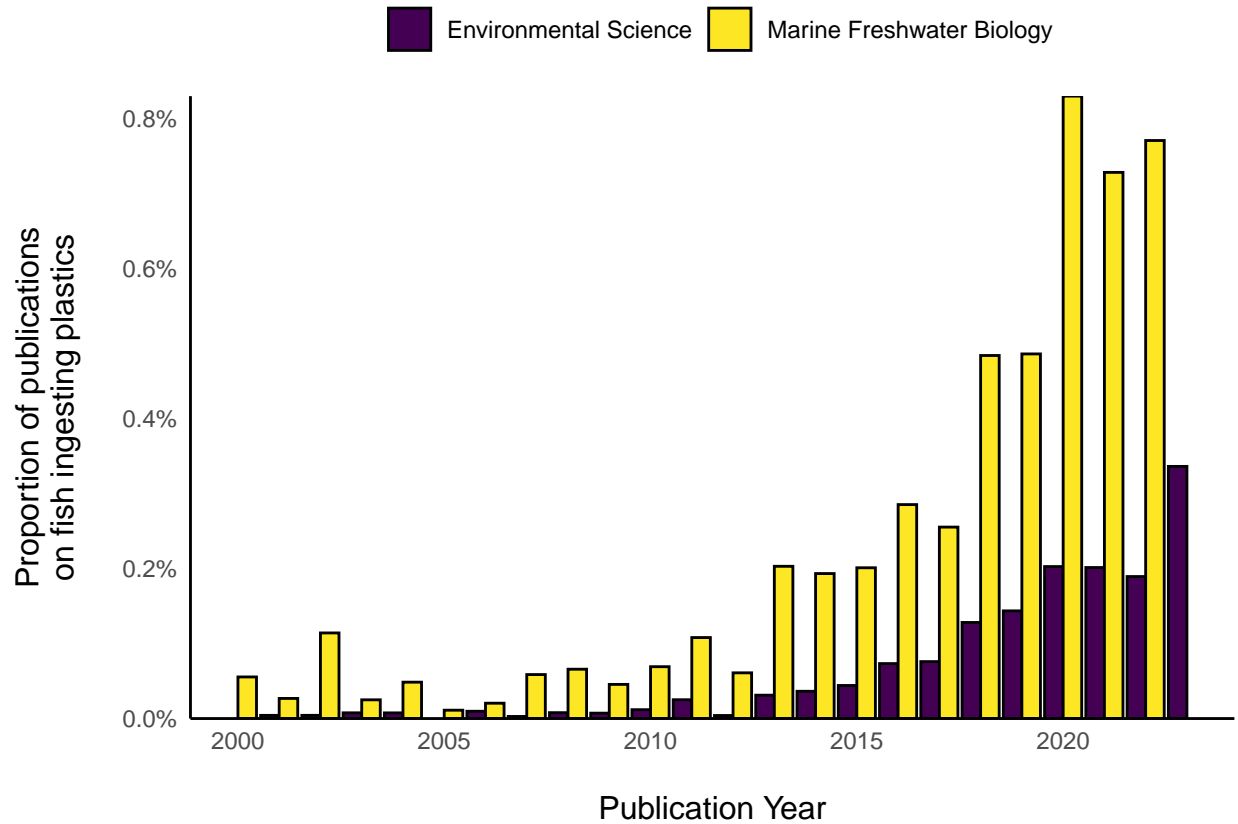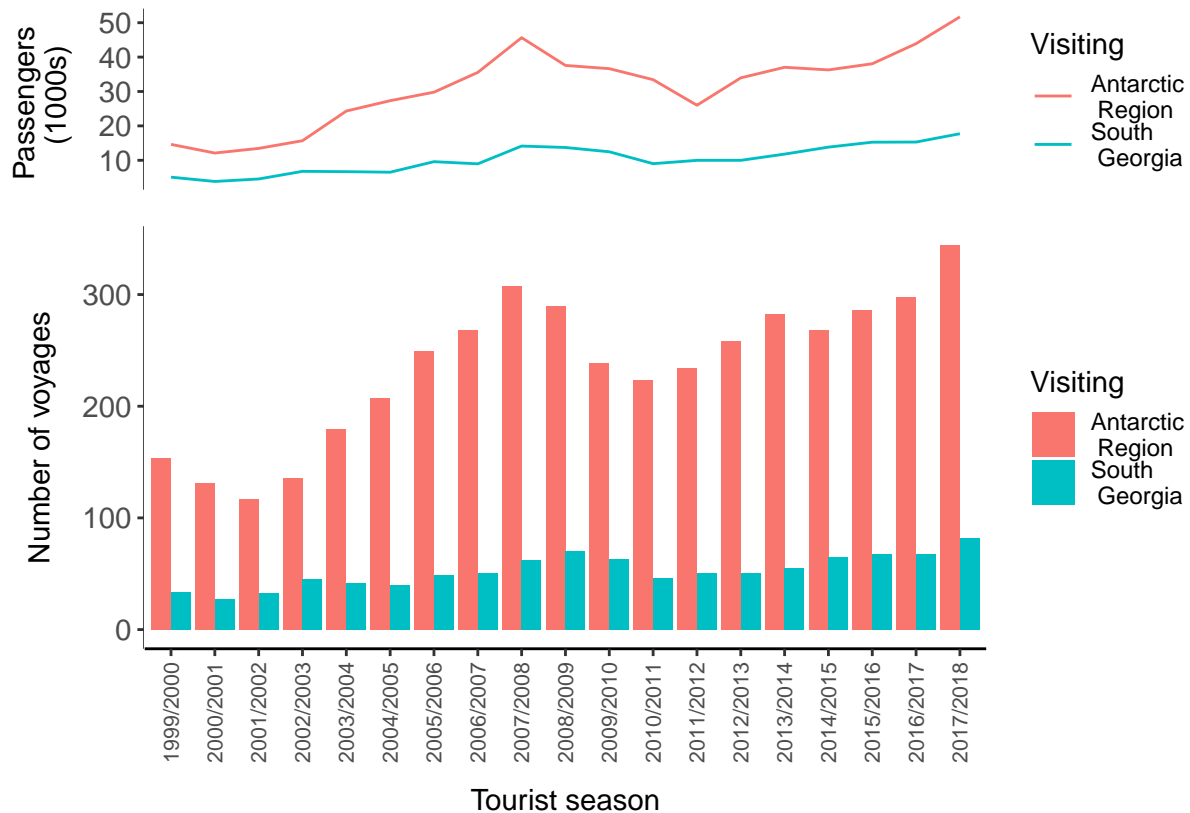
Many failed attempts to get each year to be displayed on x axis using scale_x_continuous and breaks; Google and ChatGPT suggest these should work and are not providing other solutions, but they're not, unsure why; also unsure why if I limit the years to 2022, the MFB data isn't showing.

Export png size = 750 x 300

Figure 3: Antarctica Tourism increase

```
tourism
```

Export png size = 600 x 400

Figures 7-10: Literature Data: Microplastics ingestion by fish

Load data and libraries

Figure 7: Extraction method barplot

```
extract_bar <-  ggplot(literature, aes(Extraction)) +
    labs(x="Extraction method", ##label x axis
         y="Papers using method (n=56)")+ ##label y axis
    geom_bar()+
    geom_text(aes(label= after_stat(count)), stat="count", nudge_y = -1, colour = "white")+ ##label the
    theme(
      axis.title.y = element_text(margin = margin(r = 10)), ## increase space between axis labels and a
      axis.title.x = element_text(margin = margin(t = 10)))  # increase space between axis labels and t
extract_bar ##export plot at 600x400
```
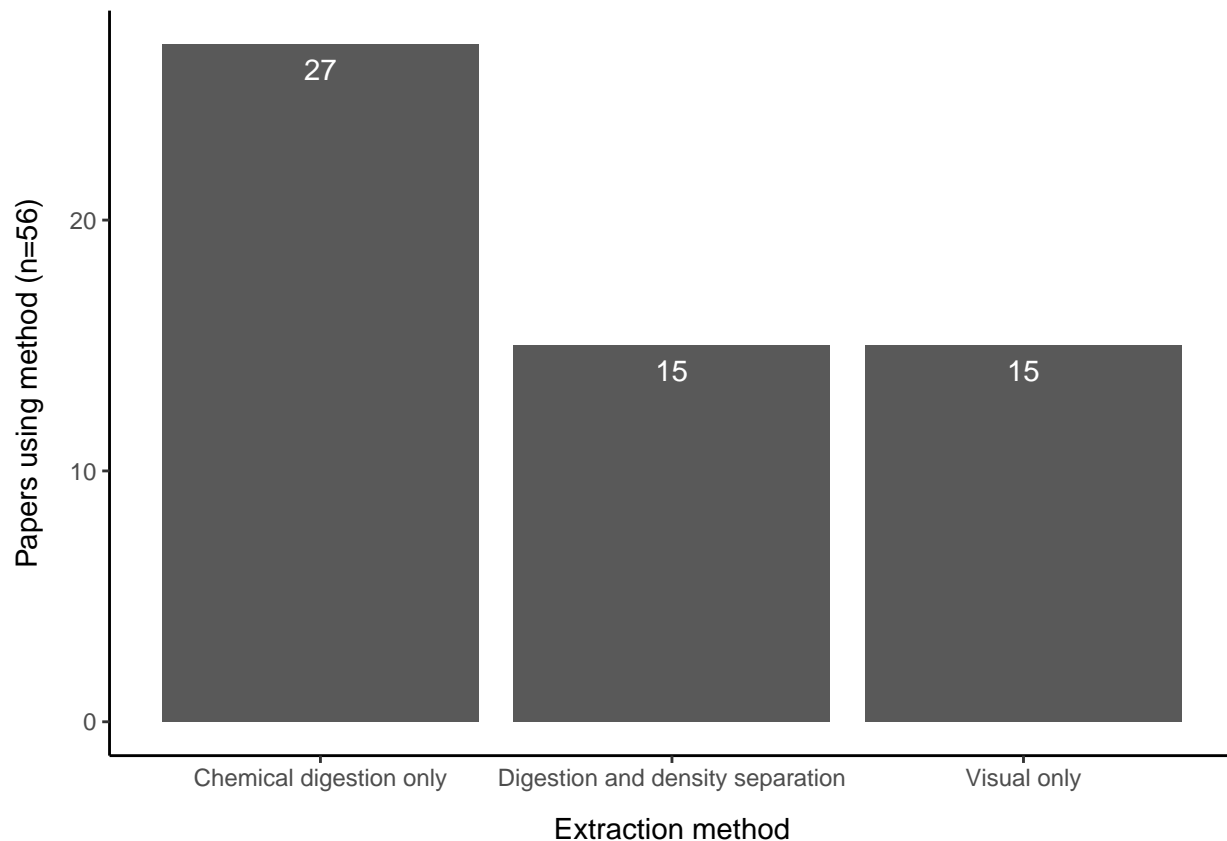
Export png size = 600 x 400

Figure 8 A: Chemicals used for digestion

```
chemdig_bar <-
  ggplot(data=subset(literature, !(ChemicalDigestant =="")), aes(x=ChemicalDigestant))+ ##plot Chemical
    labs(x="Chemical digestant", ##label x axis
         y="Papers using chemical\ndigestion (n=41)")+ ##label y axis
    geom_bar()+
    scale_x_discrete(limits = c("Potassium hydroxide", "Hydrogen peroxide", "Sodium hydroxide", "Protei
    geom_text(aes(label= after_stat(count)), stat="count", nudge_y = 1, colour = "black")+
    theme(
      axis.text.x = element_text(hjust = 0, angle = -45),
      axis.title.y = element_text(margin = margin(r = 10)), ## increase space between axis labels and a
      axis.title.x = element_text(margin = margin(t = 10)))  # increase space between axis labels and t
```

Figure 8 B: Highest temperature used during digestion

Figure 8 C: Digestion duration

Compile Figure 8 A-C

```
Figure8 <- ggarrange (chemdig_bar, temp_bar, digdur_bar, ncol=3, nrow=1, labels = c("A","B","C"), align=
Figure8 ## plot size 900 x 400
```
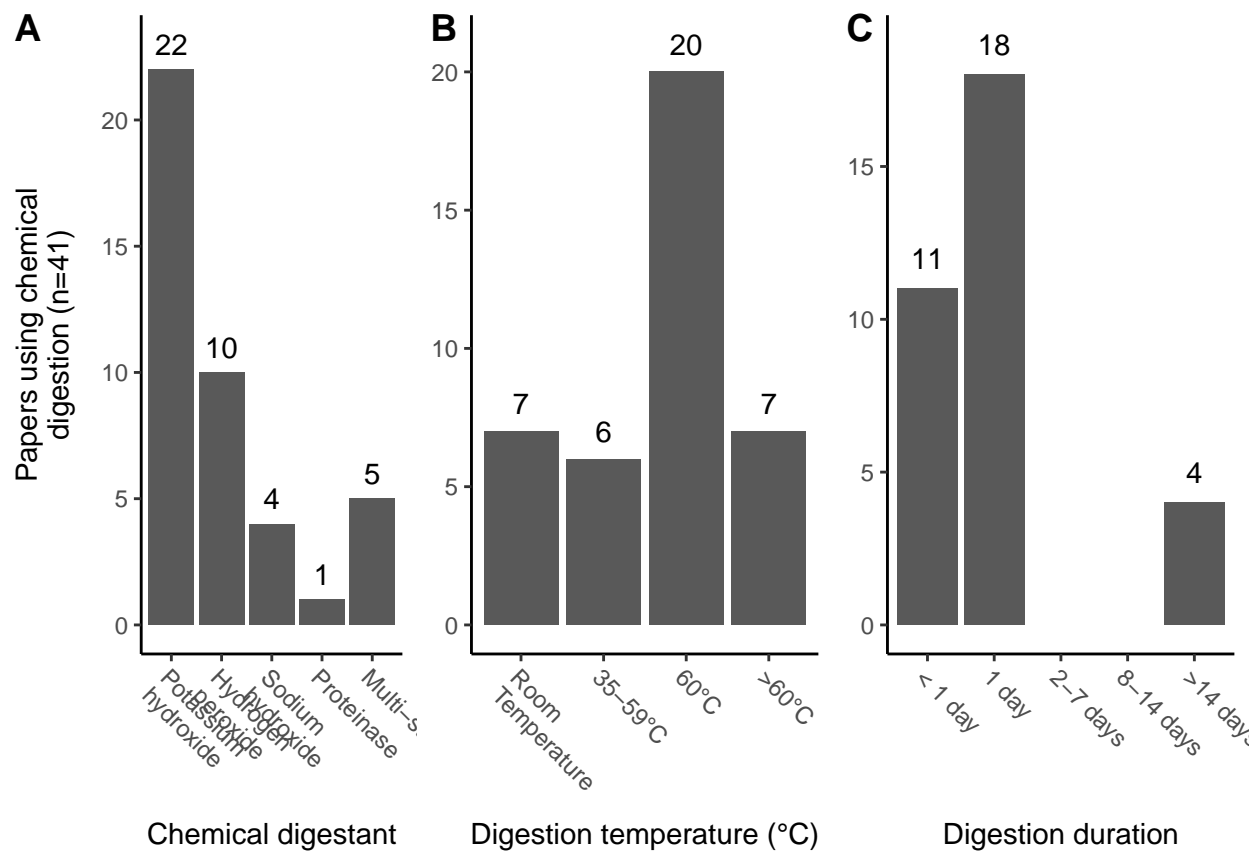

Export png size = 900 x 400

Figure 9A: Control methods employed across studies

Figure 9B: Number of controls employed in each study

Compile Figure 9 A & B

```
Figure9 <- ggarrange (controlmeth_bar, controlnum_bar, ncol=2, nrow=1, labels = c("A","B"), align="h")
Figure9 ## plot size 900 x 400
```
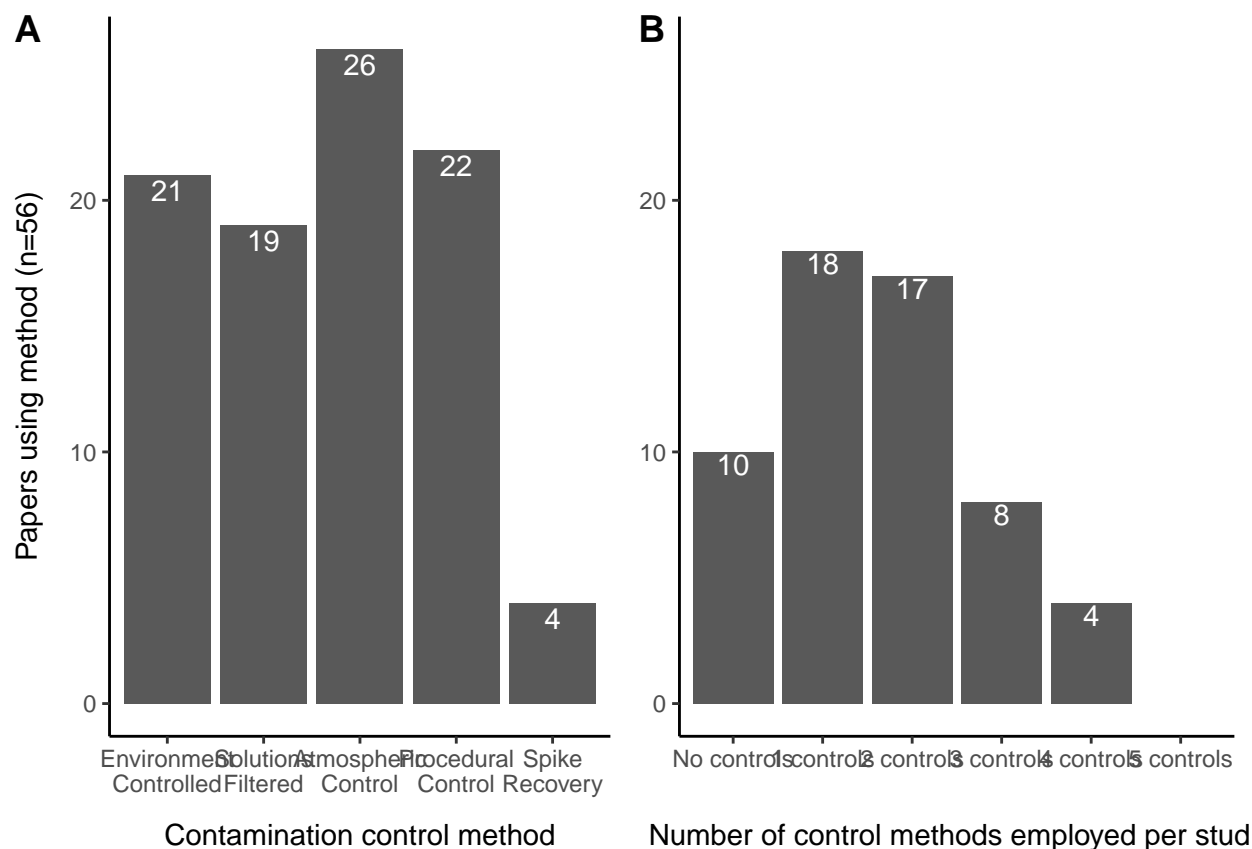
Figure 10: Plastic Polymer Confirmation Methods

```r
##Create df with method, for partial or all particles and frequency for each (possibly a quicker method

##Get counts by summing all but blank cells
VisAll <- sum(literature$VisualIDOnly == "All") #11
VisPart <- sum(literature$VisualIDOnly == "Partial") #19
HNAll <- sum(literature$HotNeedle == "All") #4
HNPart <- sum(literature$HotNeedle == "Partial") #3
FTIRAll <- sum(literature$FTIR == "All") #16
FTIRPart <- sum(literature$FTIR == "Partial") #20
RamanAll <- sum(literature$Raman == "All") #6
RamanPart <- sum(literature$Raman == "Partial") #5

PlasticPolymerMethod <- literature[,c(17:20)] ##create df with just the columns we need

PlasticPolymerMethod <- as.data.frame(t(PlasticPolymerMethod)) ##transpose ready for counts but as a df

PlasticPolymerMethod$method <- row.names(PlasticPolymerMethod) ##insert column with row names

PlasticPolymerMethod2 <- PlasticPolymerMethod ##Create copy of database ready to combine later with rbi

PlasticPolymerMethod$Particles <- c("All", "All", "All", "All") ##insert column with All
```

```r
PlasticPolymerMethod$count <- c(11, 4, 16, 6) ##insert counts as a column

PlasticPolymerMethod <- PlasticPolymerMethod[,c(58:60)] ## create final df with just the method name and

PlasticPolymerMethod2$Particles <- c("Partial", "Partial", "Partial", "Partial") ##insert column with P

PlasticPolymerMethod2$count <- c(19, 3, 20, 5) ##insert counts as a column

PlasticPolymerMethod2 <- PlasticPolymerMethod2[,c(58:60)] ## create final df with just the method name

rownames(PlasticPolymerMethod)<-NULL ##remove row names
rownames(PlasticPolymerMethod2)<-NULL ##remove row names

PlasticPolymerMethod <- rbind(PlasticPolymerMethod, PlasticPolymerMethod2)

## Create plot

polymerconf_bar <-  ggplot(PlasticPolymerMethod, aes(method,count, fill = Particles)) +
    labs(x="Plastic polymer identification method", ##label x axis
         y="Papers using method (n=56)")+ ##label y axis
    geom_col()+
    scale_x_discrete(limits = c("VisualIDOnly", "HotNeedle", "FTIR", "Raman"), labels = c("Visual Only"
    geom_label(aes(label= paste(count)), position = position_stack(vjust = 0.5), colour = "black", fill=
  scale_fill_viridis_d(option="D")+
    theme(
      axis.title.y = element_text(margin = margin(r = 10)), ## increase space between axis labels and a
      axis.title.x = element_text(margin = margin(t = 10)))  # increase space between axis labels and t
polymerconf_bar  ##export plot at 600x400
```
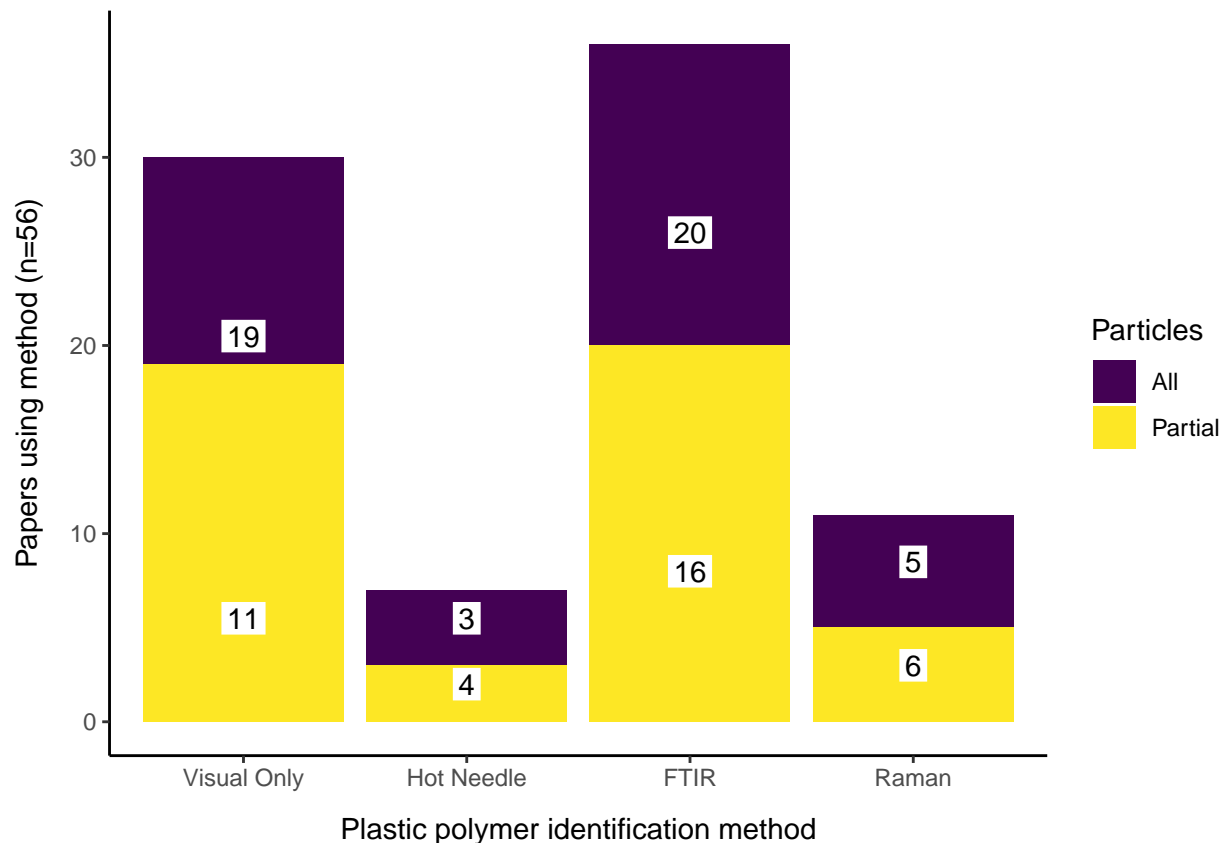
Thesis Data: Import and subset data

Fish and plastic particles data.

Create factors for Kruskal-Walis later (weight, mouth area and species specific condition).

Separate data for the different contamination control quantification, species and locations.

Import data "sample_summary.csv" and sort

Import data "particle_data.csv" and sort

Import data "control_summaries.csv" and sort

Import data "sample_summary_combinedplastic.csv" and sort

This data is where plastics and composites are counted together rather than separated, so total count is the combined total

Fish Measurement Data Analysis

1. Explore fish measurements, summarising, checking assumptions and plotting.

2. Look at the plastics in the fish.

3. Look at relationships between plastics ingested and fish species, locations and measurements

Fish measurements

First plotting lengths to show similarity across locations and species

Figure 16 - Boxplot lengths across locations and species

```
boxloclength
```
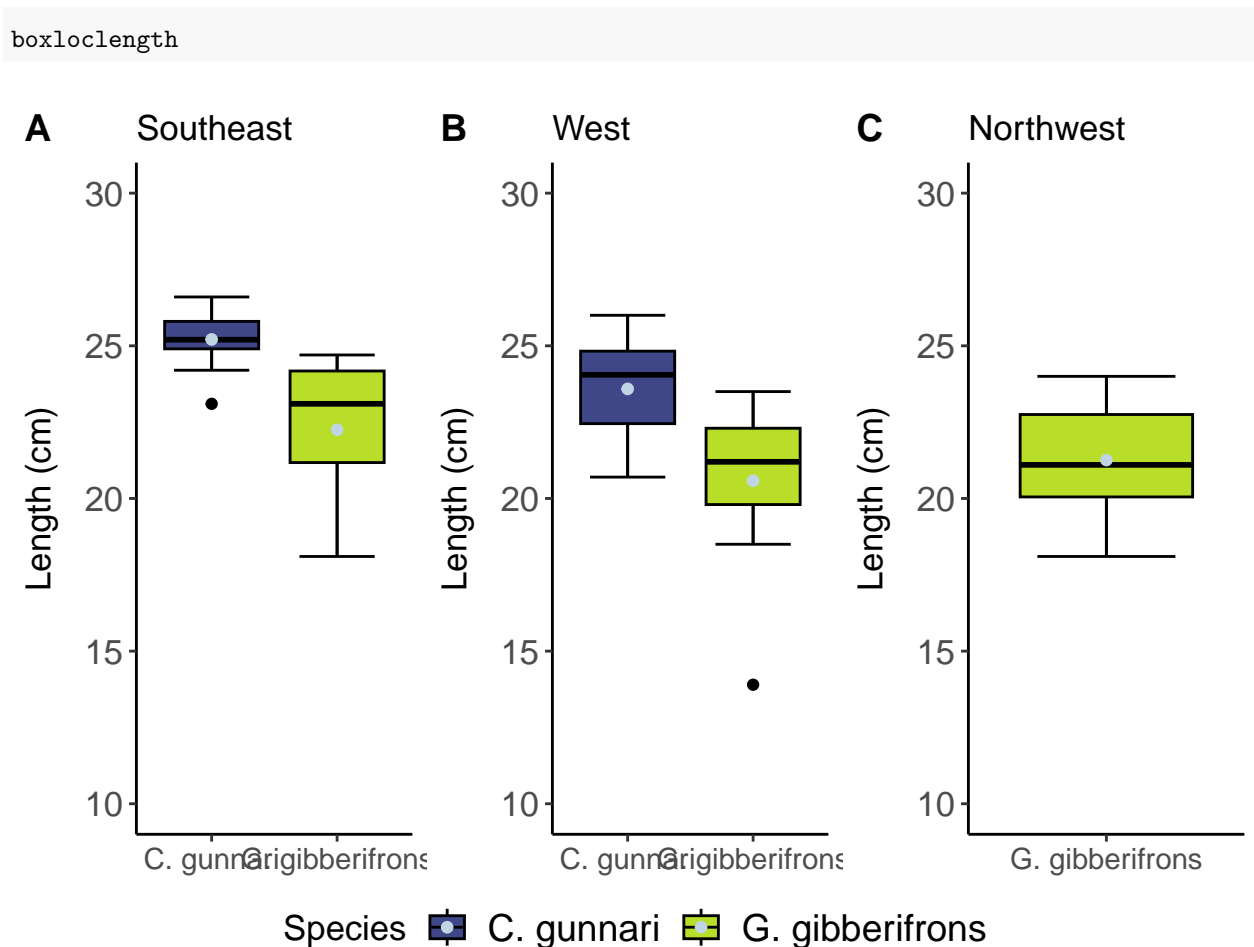

Table 7 T-tests

The morphometric data was checked for normality using visual (histogram, density plots and QQplots) and statistical (Shapiro-Wilks) methods and for homoscedasticity using Bartlett's test. Fish measurements were then checked between species for differences, visually using boxplots and 1460 statistically using the Student t-test (for normally distributed and homoscedastic data), Welch's t-test (for normally distributed heteroscedastic data) or a Mann-Whitney U/ Wilcoxon Rank Sum test (for not normally distributed, but similarly distributed data, examined visually using histograms).

Check visually for normal distribution of fish measurements per species Histograms plots species measurements:

```
## Set up multiple plots side by side with histogram to check for bell-shape
hislen <- ggplot(DFfishmeasure, aes(x = length_cm, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Length (cm)")
hiswei <- ggplot(DFfishmeasure, aes(x = weight_g, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Weight (wet)")
hisma <- ggplot(DFfishmeasure, aes(x = moutharea, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Mouth Area")
hisgut <- ggplot(DFfishmeasure, aes(x = gut_g, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Gut weight (wet)")
hisg2s <- ggplot(DFfishmeasure, aes(x = gapetosnout_mm, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Gape to snout")
hisvmo <- ggplot(DFfishmeasure, aes(x = vmo_mm, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Vertical Mouth Opening")
hishmo <- ggplot(DFfishmeasure, aes(x = hmo_mm, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Horizontal Mouth Opening")
hissfi <- ggplot(DFfishmeasure, aes(x = stomach_fullness_index, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Stomach Fullness Index")
hishea <- ggplot(DFfishmeasure, aes(x = heart_g, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Heart weight")
hiscon <- ggplot(DFfishmeasure, aes(x = condition, fill = species, colour = species)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Body condition (factor K)")

ggarrange (hislen, hiswei, hisma, hisgut, hisg2s, hisvmo, hishmo, hissfi, hishea, hiscon, ncol=5, nrow=
```
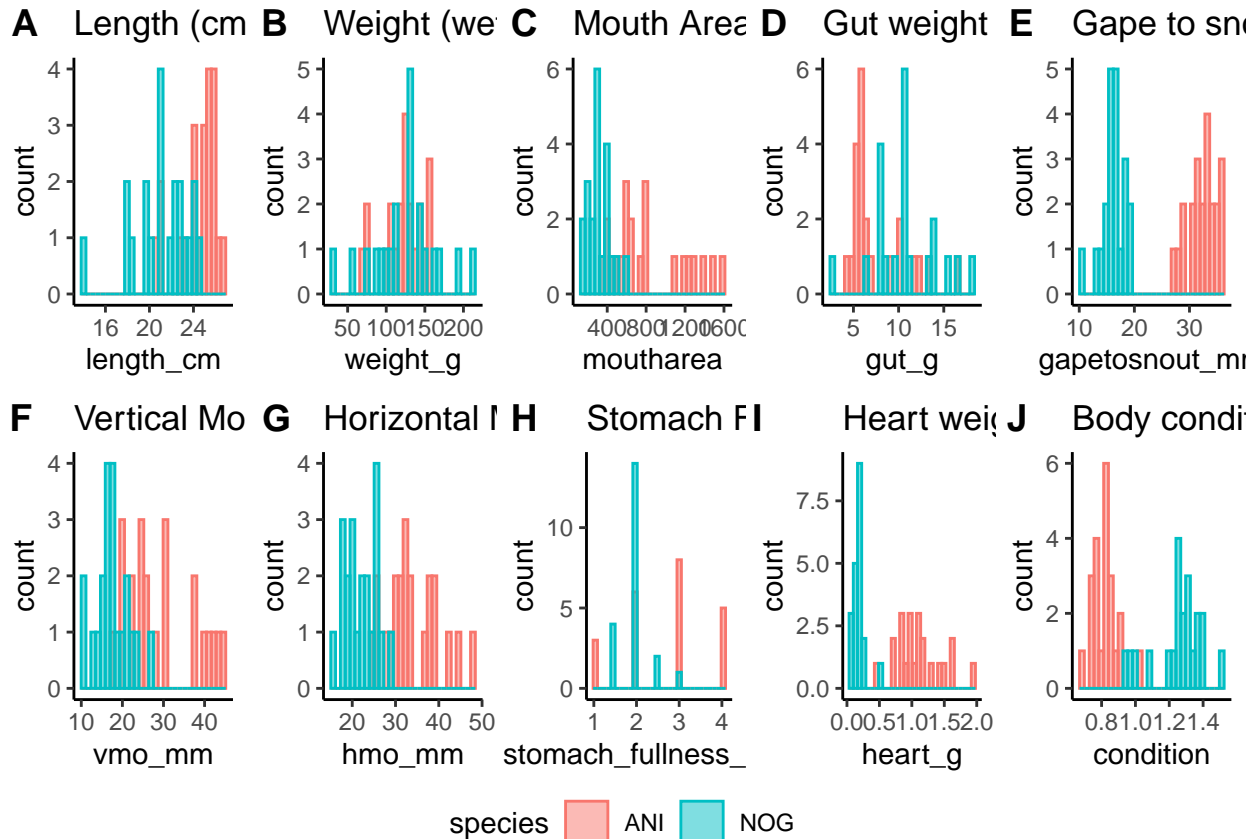
Density plots species measurements

```
####### Density Plots Fish Measurements-------
denlen <- ggplot(DFfishmeasure, aes(x=length_cm, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Length")
denwei <- ggplot(DFfishmeasure, aes(x=weight_g, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Wet weight")
denma <- ggplot(DFfishmeasure, aes(x=moutharea, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Mouth Area")
dengut <- ggplot(DFfishmeasure, aes(x=gut_g, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Gut weight(wet)")
deng2s <- ggplot(DFfishmeasure, aes(x=gapetosnout_mm, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Gape to snout length")
denvmo <- ggplot(DFfishmeasure, aes(x=vmo_mm, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Vertical Mouth Opening length")
denhmo <- ggplot(DFfishmeasure, aes(x=hmo_mm, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Horizontal Mouth Opening length")
densfi <- ggplot(DFfishmeasure, aes(x = stomach_fullness_index, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
```

```
    ggtitle("Stomach Fullness Index")
denhea <- ggplot(DFfishmeasure, aes(x = heart_g, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Heart weight")
dencon <- ggplot(DFfishmeasure, aes(x = condition, fill = species, colour = species)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Body condition (factor K)")

ggarrange (denlen, denwei, denma, dengut, deng2s, denvmo, denhmo, densfi, denhea, dencon, ncol=5, nrow=
```
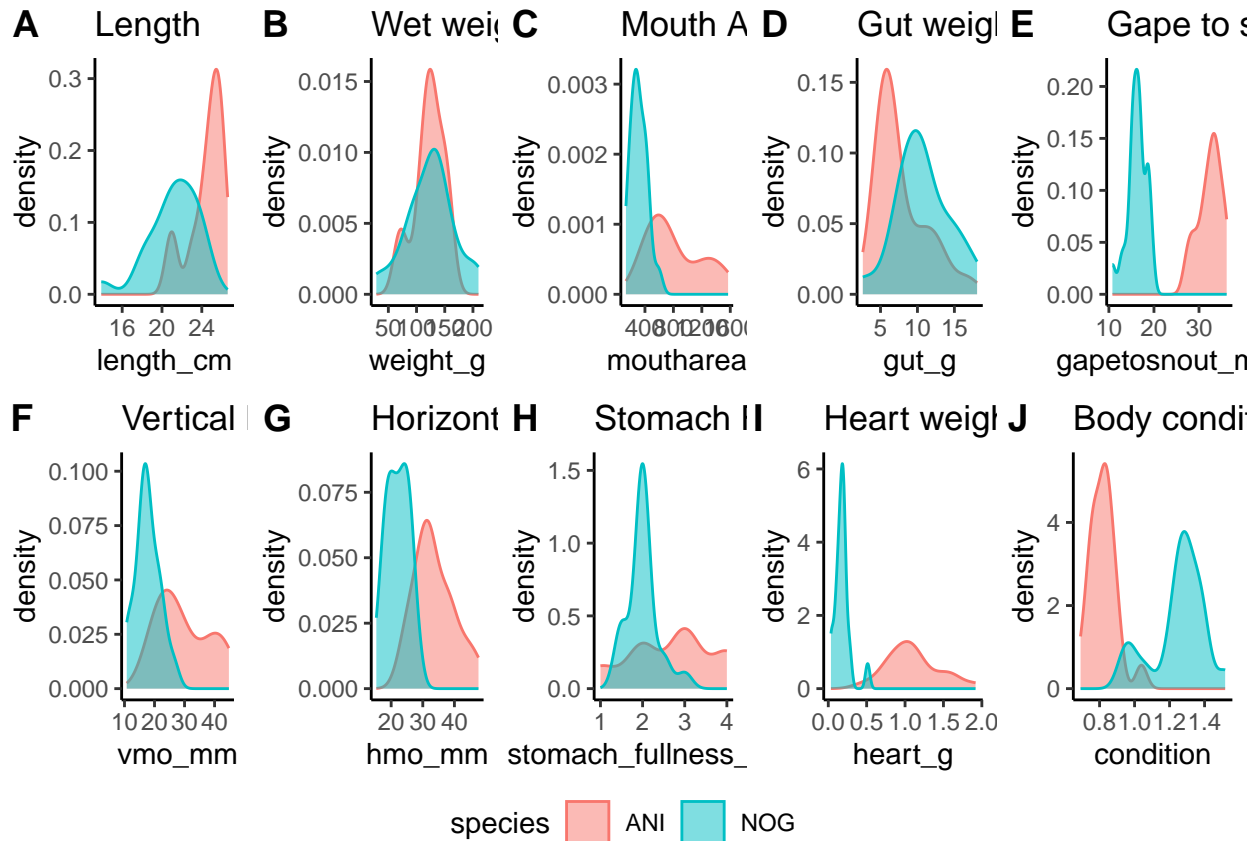

QQplots species measurements

```
####### QQ plot Fish Measurements -------
qqlen <- ggplot(DFfishmeasure, aes(sample=length_cm, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Length")
qqwei <- ggplot(DFfishmeasure, aes(sample=weight_g, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Weight(wet)")
qqma <- ggplot(DFfishmeasure, aes(sample=moutharea, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Mouth Area")
qqgut <- ggplot(DFfishmeasure, aes(sample=gut_g, colour = factor(species))) +
```

```
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gut weight (wet)")
qqg2s <- ggplot(DFfishmeasure, aes(sample=gapetosnout_mm, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gape to Snout Length")
qqvmo <- ggplot(DFfishmeasure, aes(sample=vmo_mm, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Vertical mouth opening")
qqhmo <- ggplot(DFfishmeasure, aes(sample=hmo_mm, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Horizontal mouth opening")
qqsfi <- ggplot(DFfishmeasure, aes(sample=stomach_fullness_index, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Stomach Fullness Index")
qqhea <- ggplot(DFfishmeasure, aes(sample=heart_g, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Heart weight")
qqcon <- ggplot(DFfishmeasure, aes(sample=condition, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Body condition (factor K)")

ggarrange (qqlen, qqwei, qqma, qqgut, qqg2s, qqvmo, qqhmo, qqsfi, qqhea, qqcon, ncol=5, nrow=2, labels =
```
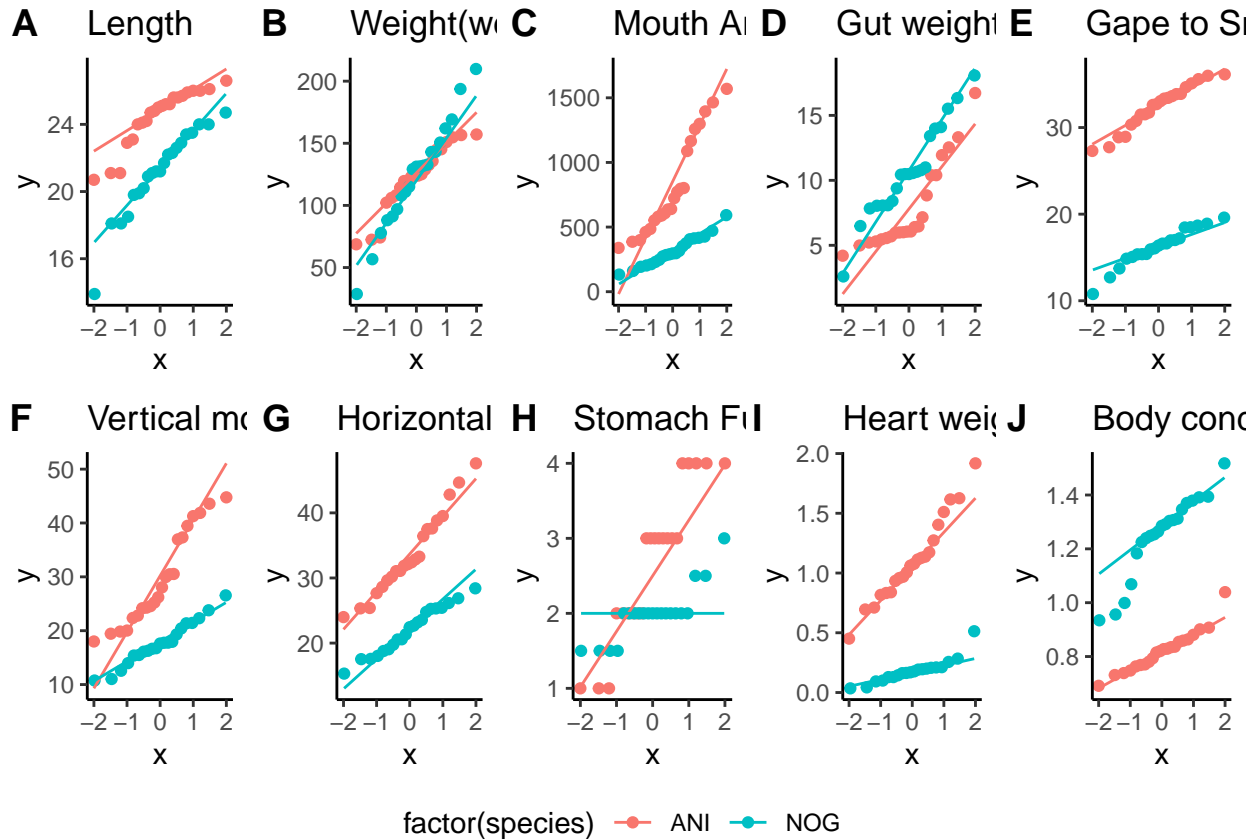
factor(species) — ANI — NOG

Checking for normality statistically (Shapiro-Wilks) per species

Test for homoscedasticity between species (equal variance across groups) using Bartlett's test. If p-value >= 0.05, group measurements are homoscedastic, use var.equal=TRUE in the T Test.

Check for significant differences in measurements between species using Student t-test for homoscedastic variables and Welch's t-test for heteroscedastic https://rcompanion.org/rcompanion/d_02.html.

using MWU for not normally distributed, but similarly distributed data

length, gut length, mouth area, stomach fullness and heart are not normal in at least one species

Checking the shape of the distributions are similar:

Fine - run MWU across all non-normal measures:

All measures, except body weight, are significantly different between species p<0.05

Table 7: Morphometric data showing mean measurements and significant differences (t-test) between species Despite non-normal distributions, the t-test appears a suitable option following advice from McDonald, John, H. (2014) Handbook of biological statistics. 3rd ed. Maryland: Sparky House Publishing.

"The t-test assumes that the observations within each group are normally distributed. Fortunately, it is not at all sensitive to deviations from this assumption, if the distributions of the two groups are the same (if both distributions are skewed to the right, for example). I've done simulations with a variety of non-normal

distributions, including flat, bimodal, and highly skewed, and the two-sample t-test always gives about 5% false positives, even with very small sample sizes." AND "The Mann-Whitney U-test is a non-parametric alternative to the two-sample t-test that some people recommend for non-normal data. However, if the two samples have the same distribution, the two-sample t-test is not sensitive to deviations from normality, so you can use the more powerful and more familiar t-test instead of the Mann-Whitney U-test. If the two samples have different distributions, the Mann-Whitney U-test is no better than the t-test. So there's really no reason to use the Mann-Whitney U-test unless you have a true ranked variable instead of a measurement variable."

So MWU for SFI but others can be checked using t-tests.

Means and SD   Provide the means and SD for the t-test results

Figure 17 - Differences in morphometric data across species - visual   Boxplot

```
plot(figure16)
```
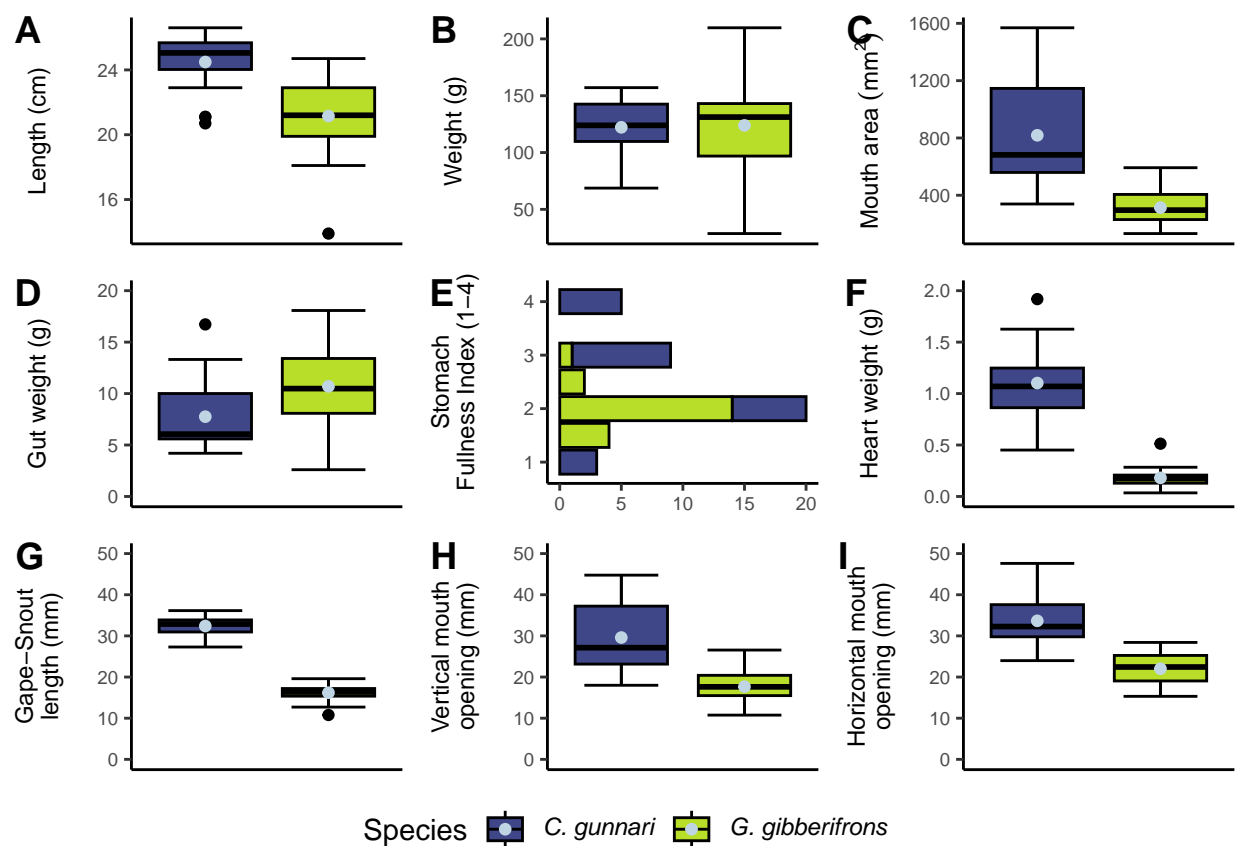


```
## plot size 753 x 553
```

Median measurements per species

Check for significant differences in fish measurements between locations   The measurements were also checked between locations for differences again visually using boxplots and then statistically using a

one-way ANOVA for the normally distributed weight data and a Kruskal-Wallis for the other non-normally distributed morphometric data.

Shapiro-Wilks (normal distribution) locations:

Homogeneity of variance across locations

Check density and qqplots per species across locations to see if that data is normally distributed

Density *C. gunnari*

```r
## Checking density plots per event per species to see if that data is normally distributed

####### Density Plots ANI Measurements -------
denlenani <- ggplot(DFfishmeasureani, aes(x=length_cm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Length")
denweiani <- ggplot(DFfishmeasureani, aes(x=weight_g, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Wet weight")
dengutani <- ggplot(DFfishmeasureani, aes(x=gut_g, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Gut weight(wet)")
deng2sani <- ggplot(DFfishmeasureani, aes(x=gapetosnout_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Gape to snout length")
denvmoani <- ggplot(DFfishmeasureani, aes(x=vmo_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Vertical Mouth Opening length")
denhmoani <- ggplot(DFfishmeasureani, aes(x=hmo_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Horizontal Mouth Opening length")
densfiani <- ggplot(DFfishmeasureani, aes(x = stomach_fullness_index, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Stomach Fullness Index")
denheaani <- ggplot(DFfishmeasureani, aes(x = heart_g, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Heart weight")
denconani <- ggplot(DFfishmeasureani, aes(x = condition, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Body condition (factor K)")

ggarrange (denlenani, denweiani, dengutani, deng2sani, denvmoani, denhmoani, densfiani, denheaani, denc
```
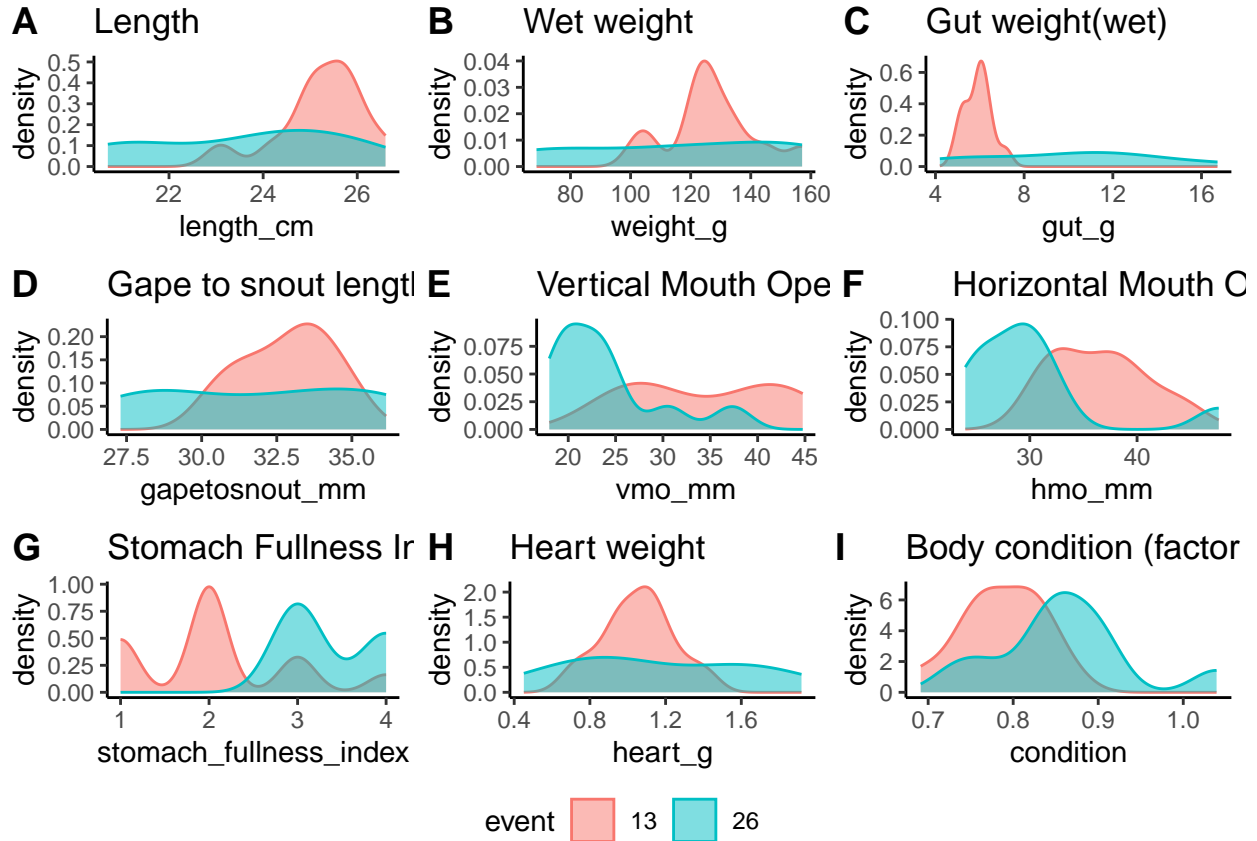
Density *G. gibberifrons*

```
## Checking density per event for G. gibb to see if that data is normally distributed
####### Density Plots NOG Measurements-------
denlennog <- ggplot(DFfishmeasurenog, aes(x=length_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Length")
denweinog <- ggplot(DFfishmeasurenog, aes(x=weight_g, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Wet weight")
dengutnog <- ggplot(DFfishmeasurenog, aes(x=gut_g, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Gut weight(wet)")
deng2snog <- ggplot(DFfishmeasurenog, aes(x=gapetosnout_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Gape to snout length")
denvmonog <- ggplot(DFfishmeasurenog, aes(x=vmo_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Vertical Mouth Opening length")
denhmonog <- ggplot(DFfishmeasurenog, aes(x=hmo_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Horizontal Mouth Opening length")
densfinog <- ggplot(DFfishmeasurenog, aes(x = stomach_fullness_index, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Stomach Fullness Index")
denheanog <- ggplot(DFfishmeasurenog, aes(x = heart_g, fill = event, colour = event)) +
```

```
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Heart weight")
denconnog <- ggplot(DFfishmeasurenog, aes(x = condition, fill = event, colour = event)) +
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Body condition (factor K)")

ggarrange (denlennog, denweinog, dengutnog, deng2snog, denvmonog, denhmonog, densfinog, denheanog, dence
```
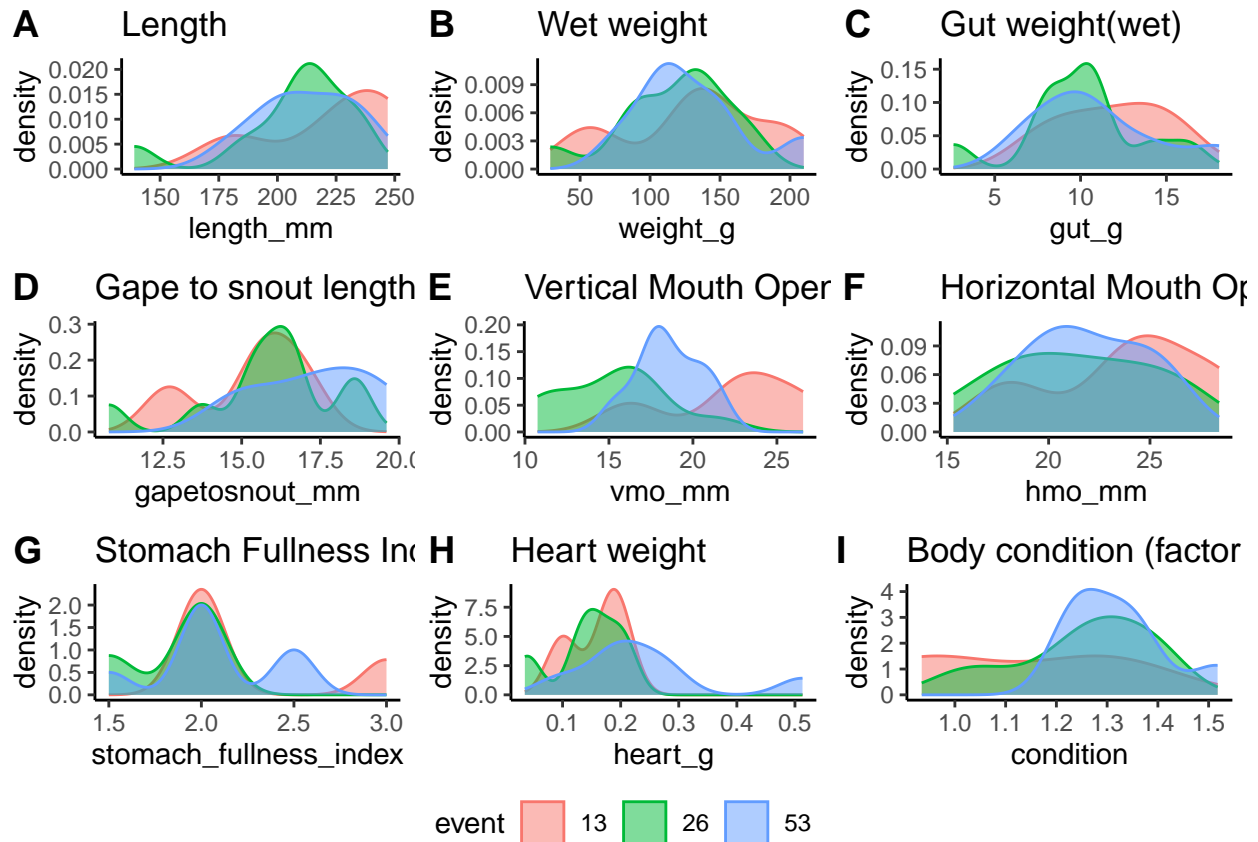

Density all fish (both species) per location

```
###### Density Plots Fish Measurements-------
denevenlen <- ggplot(DFfishmeasure, aes(x=length_mm, fill = event, colour = event)) +
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Length")
denevenwei <- ggplot(DFfishmeasure, aes(x=weight_g, fill = event, colour = event)) +
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Wet weight")
denevengut <- ggplot(DFfishmeasure, aes(x=gut_g, fill = event, colour = event)) +
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Gut weight(wet)")
deneveng2s <- ggplot(DFfishmeasure, aes(x=gapetosnout_mm, fill = event, colour = event)) +
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Gape to snout length")
denevenvmo <- ggplot(DFfishmeasure, aes(x=vmo_mm, fill = event, colour = event)) +
    geom_density(alpha = 0.5, position = "identity") +
    ggtitle("Vertical Mouth Opening length")
```

```
denevenhmo <- ggplot(DFfishmeasure, aes(x=hmo_mm, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Horizontal Mouth Opening length")
denevensfi <- ggplot(DFfishmeasure, aes(x = stomach_fullness_index, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Stomach Fullness Index")
denevenhea <- ggplot(DFfishmeasure, aes(x = heart_g, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Heart weight")
denevencon <- ggplot(DFfishmeasure, aes(x = condition, fill = event, colour = event)) +
  geom_density(alpha = 0.5, position = "identity") +
  ggtitle("Body condition (factor K)")

ggarrange (denevenlen, denevenwei, denevengut, deneveng2s, denevenvmo, denevenhmo, denevensfi, denevenh
```
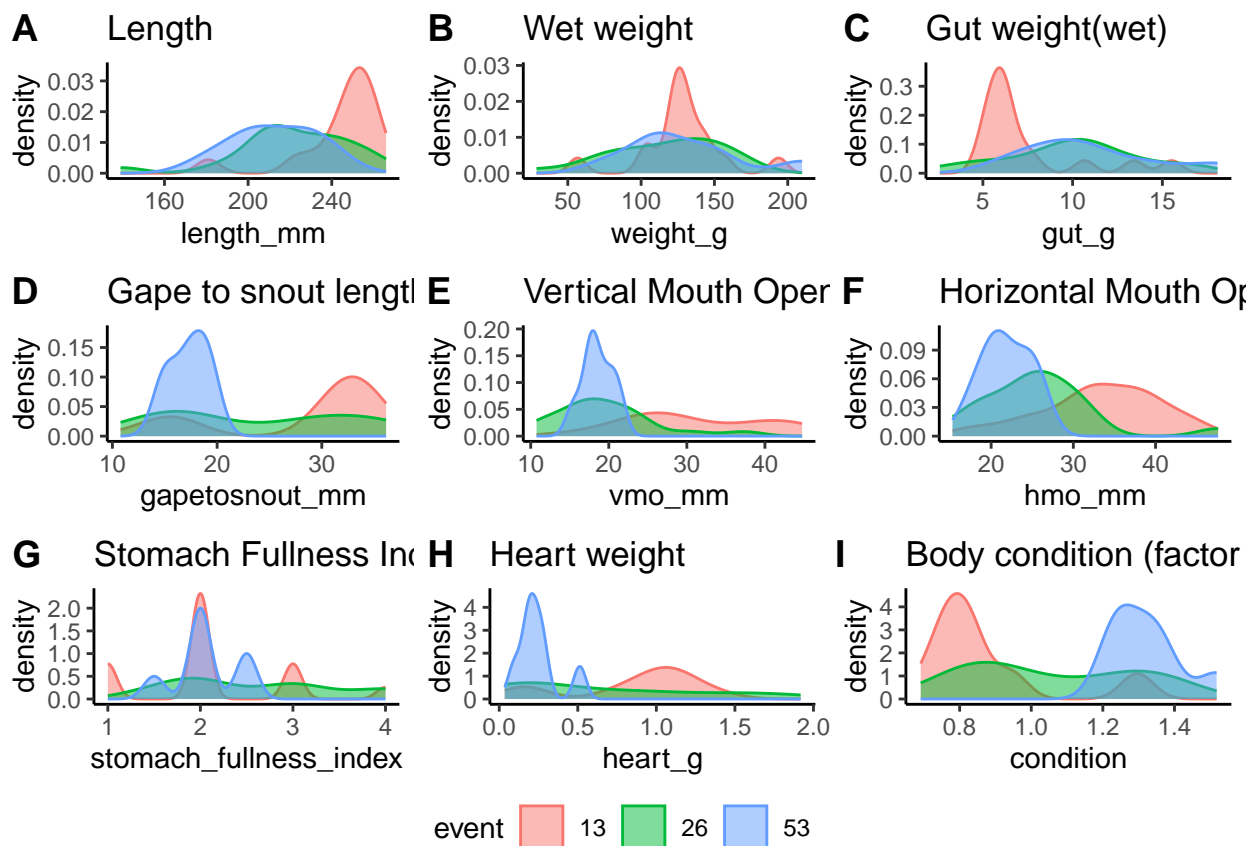

QQplots *C. gunnari*

```
####### QQ plot ANI Measurements -------
qqlenani <- ggplot(DFfishmeasureani, aes(sample=length_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Length")
qqweiani <- ggplot(DFfishmeasureani, aes(sample=weight_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Weight(wet)")
```

```r
qqgutani <- ggplot(DFfishmeasureani, aes(sample=gut_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gut weight (wet)")
qqg2sani <- ggplot(DFfishmeasureani, aes(sample=gapetosnout_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gape to Snout Length")
qqvmoani <- ggplot(DFfishmeasureani, aes(sample=vmo_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Vertical mouth opening")
qqhmoani <- ggplot(DFfishmeasureani, aes(sample=hmo_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Horizontal mouth opening")
qqsfiani <- ggplot(DFfishmeasureani, aes(sample=stomach_fullness_index, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Stomach Fullness Index")
qqheaani <- ggplot(DFfishmeasureani, aes(sample=heart_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Heart weight")
qqconani <- ggplot(DFfishmeasureani, aes(sample=condition, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Body condition (factor K)")

ggarrange (qqlenani, qqweiani, qqgutani, qqg2sani, qqvmoani, qqhmoani, qqsfiani, qqheaani, qqconani, nc
```

QQplots *G. gibberifrons*

```r
qqlennog <- ggplot(DFfishmeasurenog, aes(sample=length_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Length")
qqweinog <- ggplot(DFfishmeasurenog, aes(sample=weight_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Weight(wet)")
qqgutnog <- ggplot(DFfishmeasurenog, aes(sample=gut_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gut weight (wet)")
qqg2snog <- ggplot(DFfishmeasurenog, aes(sample=gapetosnout_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gape to Snout Length")
qqvmonog <- ggplot(DFfishmeasurenog, aes(sample=vmo_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Vertical mouth opening")
qqhmonog <- ggplot(DFfishmeasurenog, aes(sample=hmo_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Horizontal mouth opening")
```
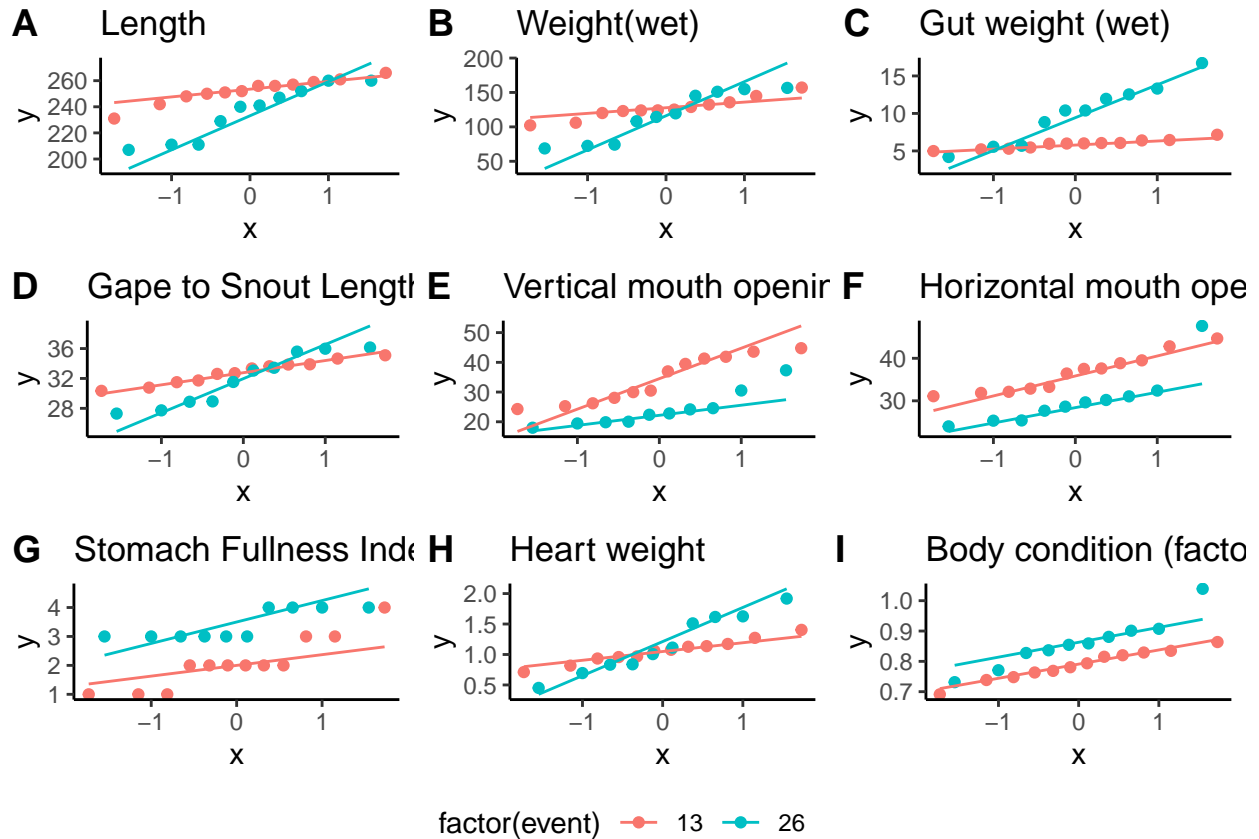
```
qqsfinog <- ggplot(DFfishmeasurenog, aes(sample=stomach_fullness_index, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Stomach Fullness Index")
qqheanog <- ggplot(DFfishmeasurenog, aes(sample=heart_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Heart weight")
qqconnog <- ggplot(DFfishmeasurenog, aes(sample=condition, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Body condition (factor K)")

ggarrange (qqlennog, qqweinog, qqgutnog, qqg2snog, qqvmonog, qqhmonog, qqsfinog, qqheanog, qqconnog, nc
```


QQplots - both species (all fish) at each location

```
qqevenlen <- ggplot(DFfishmeasure, aes(sample=length_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Length")
qqevenwei <- ggplot(DFfishmeasure, aes(sample=weight_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Weight(wet)")
qqevengut <- ggplot(DFfishmeasure, aes(sample=gut_g, colour = factor(event))) +
```
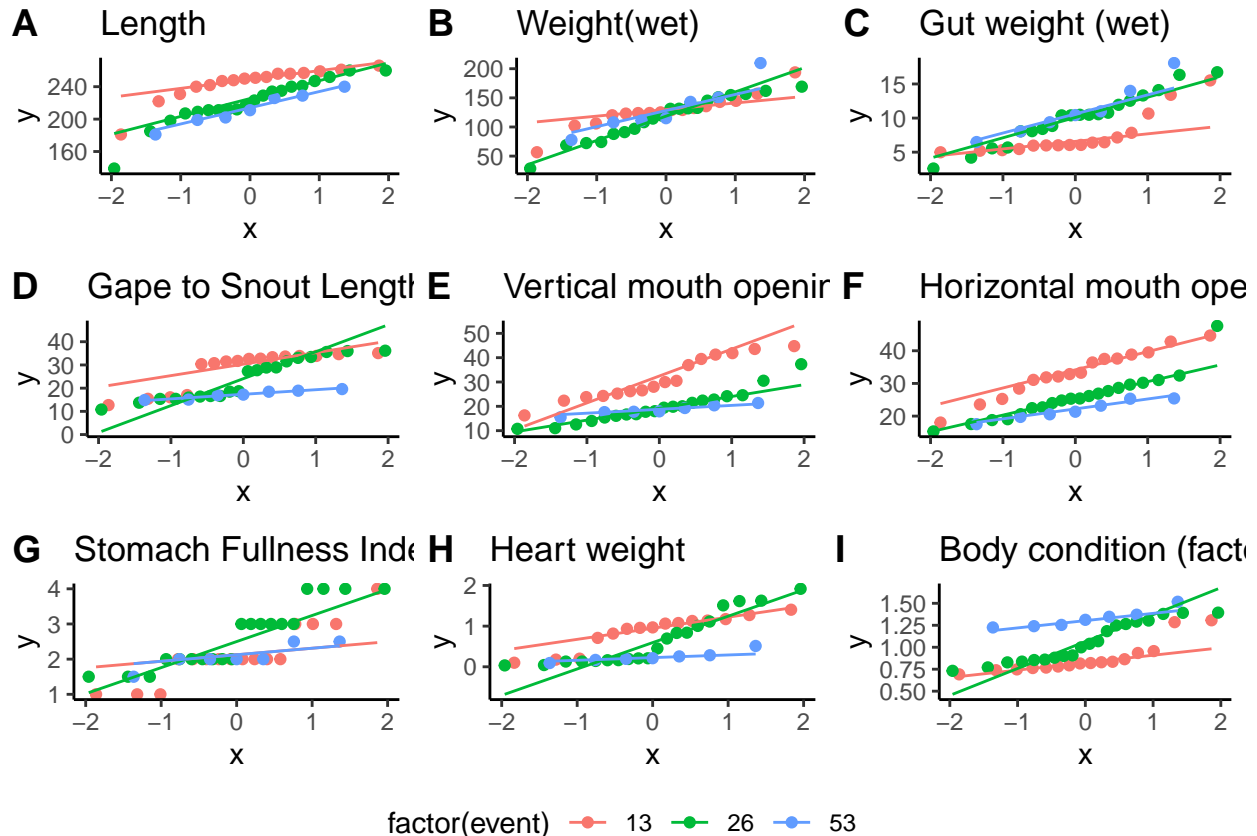
```r
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gut weight (wet)")
qqeveng2s <- ggplot(DFfishmeasure, aes(sample=gapetosnout_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Gape to Snout Length")
qqevenvmo <- ggplot(DFfishmeasure, aes(sample=vmo_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Vertical mouth opening")
qqevenhmo <- ggplot(DFfishmeasure, aes(sample=hmo_mm, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Horizontal mouth opening")
qqevensfi <- ggplot(DFfishmeasure, aes(sample=stomach_fullness_index, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Stomach Fullness Index")
qqevenhea <- ggplot(DFfishmeasure, aes(sample=heart_g, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Heart weight")
qqevencon <- ggplot(DFfishmeasure, aes(sample=condition, colour = factor(event))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Body condition (factor K)")

ggarrange (qqevenlen, qqevenwei, qqevengut, qqeveng2s, qqevenvmo, qqevenhmo, qqevensfi, qqevenhea, qqeve
```

A Length

B Weight(wet)

C Gut weight (wet)

D Gape to Snout Length

E Vertical mouth opening

F Horizontal mouth opening

G Stomach Fullness Index

H Heart weight

I Body condition (factor)

factor(event) — 13 — 26 — 53

Histogram all fish per location

```
## Set up multiple plots side by side with histogram to check for bell-shape
hisevenlen <- ggplot(DFfishmeasure, aes(x = length_mm, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Length")
hisevenwei <- ggplot(DFfishmeasure, aes(x = weight_g, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Weight (wet)")
hisevengut <- ggplot(DFfishmeasure, aes(x = gut_g, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Gut weight (wet)")
hiseveng2s <- ggplot(DFfishmeasure, aes(x = gapetosnout_mm, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Gape to snout")
hisevenvmo <- ggplot(DFfishmeasure, aes(x = vmo_mm, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Vertical Mouth Opening")
hisevenhmo <- ggplot(DFfishmeasure, aes(x = hmo_mm, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Horizontal Mouth Opening")
hisevensfi <- ggplot(DFfishmeasure, aes(x = stomach_fullness_index, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Stomach Fullness Index")
hisevenhea <- ggplot(DFfishmeasure, aes(x = heart_g, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
```

```
  ggtitle("Heart weight")
hisevencon <- ggplot(DFfishmeasure, aes(x = condition, fill = event, colour = event)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  ggtitle("Body condition (factor K)")

ggarrange (hisevenlen, hisevenwei, hisevengut, hiseveng2s, hisevenvmo, hisevenhmo, hisevensfi, hisevenh
```


Testing for differences amongst three groups, so One-way ANOVA for continuous, normal and homoscedastic = weight only; Kruskal-Wallis for non-normal data (no assumption of homogeneity) = length, gut weight, heart weight, gape to snout, vmo and hmo

```
summary(aov(DFfishmeasure$weight_g ~ DFfishmeasure$event))
```

```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## DFfishmeasure$event   2   1666   833.1   0.675  0.515
## Residuals            40  49391  1234.8
```

```
#                      Df Sum Sq Mean Sq F value Pr(>F)
# DFfishmeasure$event   2   1666   833.1   0.675  0.515
# Residuals            40  49391  1234.8
## p > 0.05 supporting the null hypothesis that there is not a significant difference between these gro
kruskal.test(DFfishmeasure$length_mm ~ DFfishmeasure$event)
```

```
##
```

```
##  Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$length_mm by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 12.697, df = 2, p-value = 0.00175
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$length_mm by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 12.697, df = 2, p-value = 0.00175 ##DIFFERENT
kruskal.test(DFfishmeasure$gut_g ~ DFfishmeasure$event)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$gut_g by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 7.6357, df = 2, p-value = 0.02197
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$gut_g by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 7.6357, df = 2, p-value = 0.02197 ##DIFFERENT

kruskal.test(DFfishmeasure$heart_g ~ DFfishmeasure$event)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$heart_g by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 5.5161, df = 2, p-value = 0.06342
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$heart_g by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 5.5161, df = 2, p-value = 0.06342

kruskal.test(DFfishmeasure$gapetosnout_mm ~ DFfishmeasure$event)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$gapetosnout_mm by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 5.7273, df = 2, p-value = 0.05706
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$gapetosnout_mm by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 5.7273, df = 2, p-value = 0.05706

kruskal.test(DFfishmeasure$vmo_mm ~ DFfishmeasure$event)
```

```
##
```

```
##   Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$vmo_mm by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 18.342, df = 2, p-value = 0.000104
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$vmo_mm by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 18.342, df = 2, p-value = 0.000104 ##DIFFERENT

kruskal.test(DFfishmeasure$hmo_mm ~ DFfishmeasure$event)
```

```
##
##   Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$hmo_mm by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 14.669, df = 2, p-value = 0.0006526
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$hmo_mm by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 14.669, df = 2, p-value = 0.0006526 ##DIFFERENT

##I think Kruskal-Wallis will work on SFI also as it is discrete and ordinal
kruskal.test(DFfishmeasure$stomach_fullness_index ~ DFfishmeasure$event)
```

```
##
##   Kruskal-Wallis rank sum test
##
## data:  DFfishmeasure$stomach_fullness_index by DFfishmeasure$event
## Kruskal-Wallis chi-squared = 2.9994, df = 2, p-value = 0.2232
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFfishmeasure$stomach_fullness_index by DFfishmeasure$event
# Kruskal-Wallis chi-squared = 2.9994, df = 2, p-value = 0.2232
```

Boxplots:

```
## create labels

species.labels <- c("C. gunnari", "G. gibberifrons")
names(species.labels) <- c("ANI", "NOG")

event.labels <- c("Southeast", "West", "Northwest")
names(event.labels) <- c("13", "26", "53")


######Are the fish measurements different across events - visual######
##Boxplots fish measure comparing events

boxeventlength <-
```

```r
  ggplot(DFfishmeasure, aes(group = event, fill=event, ##group variables as factors
                            y = length_mm,
                            factor(event,
                                   labels = event.labels))) + ##label event factors
  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Length (mm)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax

  )

boxeventweight <-
ggplot(DFfishmeasure, aes(group = event, fill = event, ##group variables as factors
                          y = weight_g,
                          factor(event,
                                 labels = event.labels))) + ##label event factors
  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Weight (g)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax

  )

boxeventgut <-
  ggplot(DFfishmeasure, aes(group = event, fill = event, ##group variables as factors
                            y = gut_g,
                            factor(event,
                                   labels = event.labels))) + ##label event factors
  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Gut weight (g)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax

  )

bareventsfi <-
  ggplot(DFfishmeasure, aes(y = stomach_fullness_index))+
  geom_bar(aes(fill=event), colour="black") +
  ylab("Stomach\nFullness Index\n(1 = empty - 4 = full)")+
  xlab(NULL)+ ##count is obvious and looks better without
  theme(
```

```
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax.
  )

boxeventhea <-
  ggplot(DFfishmeasure, aes(group = event, fill = event, ##group variables as factors
                            y = heart_g,
                            factor(event,
                                   labels = event.labels))) + ##label event factors

  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Heart weight (g)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax.

  )

boxeventg2s <-
  ggplot(DFfishmeasure, aes(group = event, fill = event, ##group variables as factors
                            y = gapetosnout_mm,
                            factor(event,
                                   labels = event.labels))) + ##label event factors
  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Gape to snout\nlength (mm)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax.

  )

boxeventvmo <-
  ggplot(DFfishmeasure, aes(group = event, fill = event, ##group variables as factors
                            y = vmo_mm,
                            factor(event,
                                   labels = event.labels))) + ##label event factors
  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Vertical mouth\nopening (mm)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax.

  )
```

```
boxeventhmo <-
  ggplot(DFfishmeasure, aes(group = event, fill = event, ##group variables as factors
                            y = hmo_mm,
                            factor(event,
                                   labels = event.labels))) + ##label event factors
  geom_boxplot(color="black") +
  stat_summary(fun=mean, geom="point", shape=20, size=2.5, color="#BFD5E3") +
  xlab(NULL)+ ##no x axis label as event already obvious
  ylab('Horizontal mouth\nopening (mm)')+ ##y axis label
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
    axis.title.y = element_text(margin = margin(r = 10), size = rel(0.8)), ## increase space between ax

  )
ggarrange (boxeventlength, boxeventweight, NULL, boxeventgut, bareventsfi, boxeventhea, boxeventg2s, bo
```
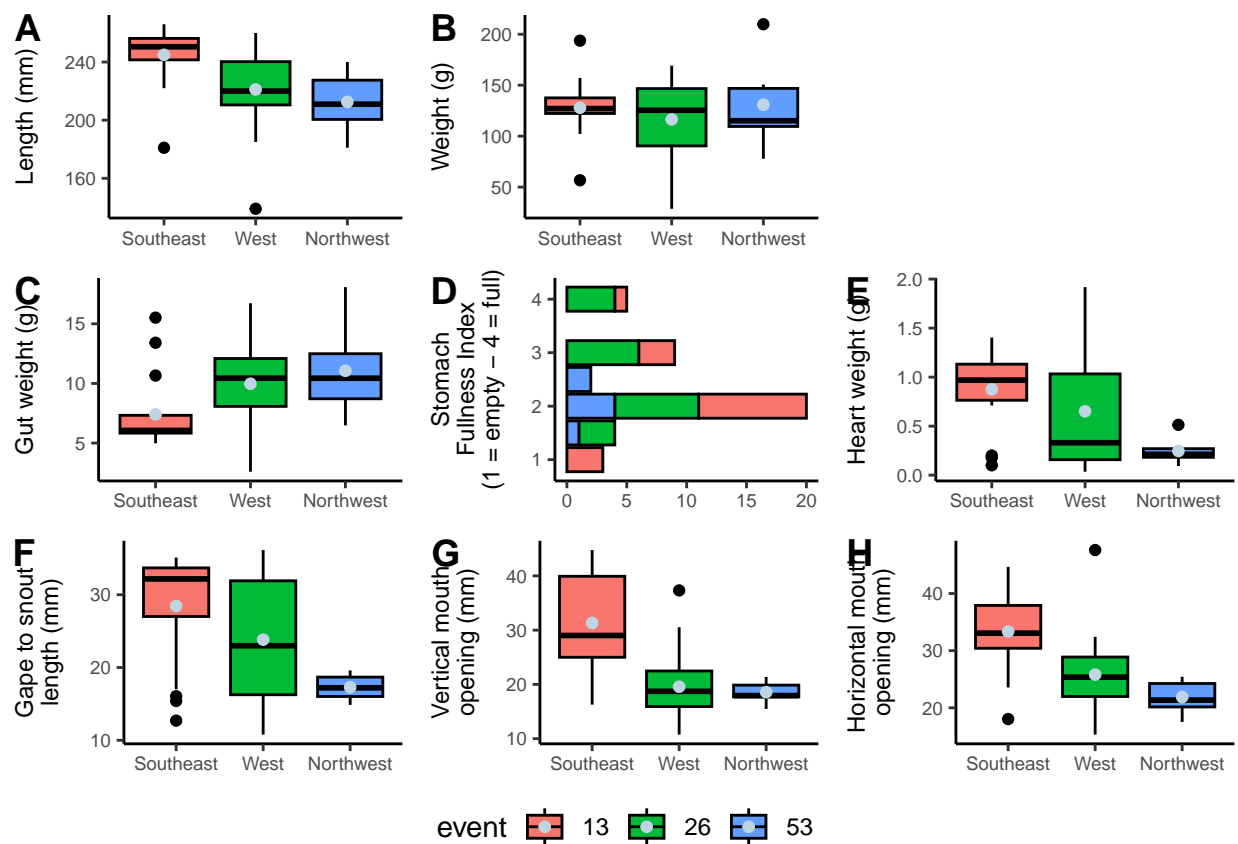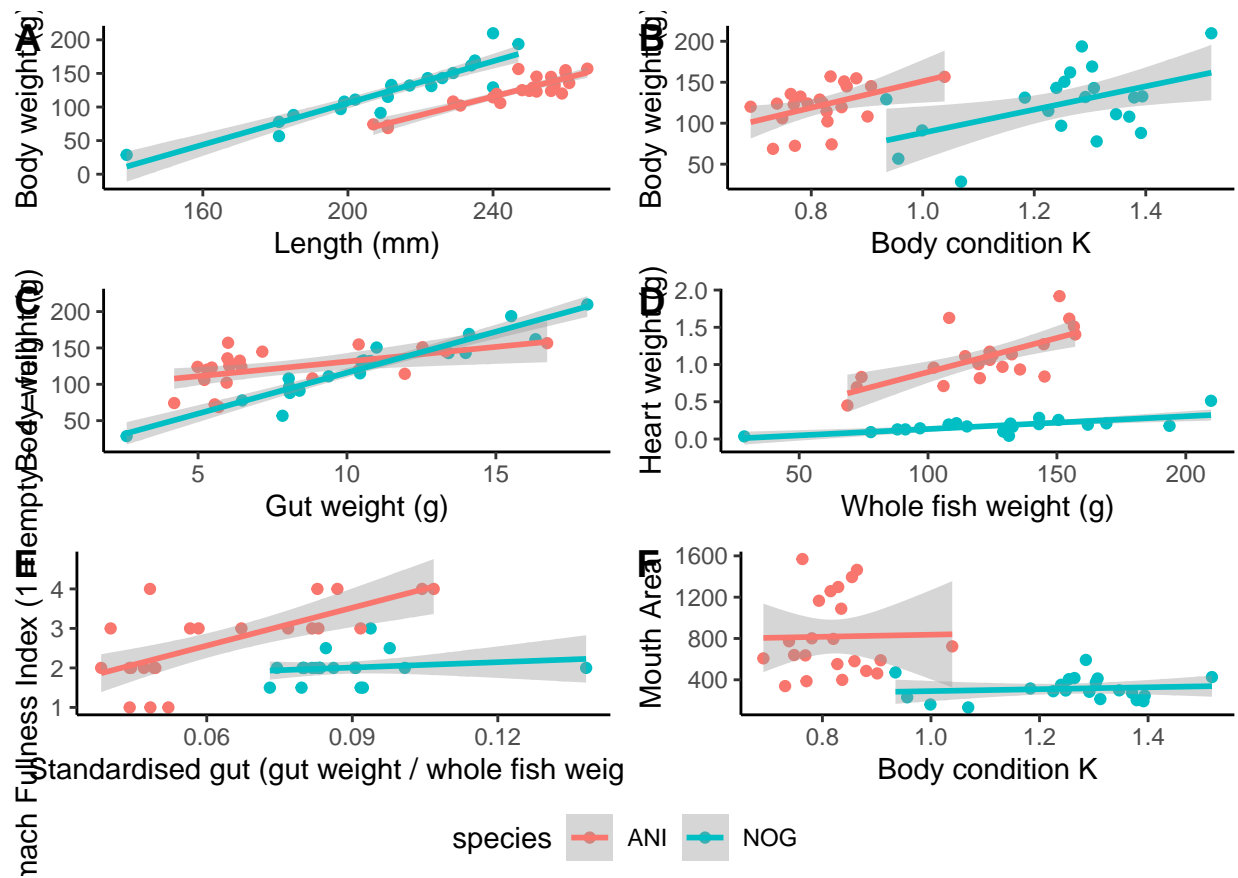

Median IQR all fish at each location

Mean ±SD per species per location

Correlations between the various measures were also run. The whole fish measurement that correlates best with other measures and is most comparable across the two species and location was used as the measurement reflective of size and age.

```
corrmeasuresplot
```


Plastics in fish

Plastic/composite particle size

Frequency of occurrence (FO)

The frequency of occurrence (FO) of plastic or composite ingestion by fish per species and location was also calculated as the percentage of fish with at least one piece of plastic inside.

Across all fish

Across *C. gunnari*

Across *G. gibberifrons*

Across southeast location (13)

Across west location (26)

Across northwest location (53)

Plastic loads (PL)

The average (mean ± SD) amount of plastic or composite particles (of any size from macro- to micro-) per fish (plastic load – PL) for each species and each location was calculated, including those fish where no plastic was present.

dataframe of plastic load (means and sd) per species and event per variable ignoring NA values

Plastic load per event

Plastic load per event per species

Plastic load all fish

Effect of location, species or morphometrics on plastic particle ingestion

Effect of trawl location, species, weight, body condition and mouth area on plastic particle ingestion was looked at using Kruskal-Wallis and where results were significant, a Mann-Whitney U was carried out to identify the pairwise results.


Plastic load (individual) correlated with morphometrics

Testing for normality (unlikely)

```
##test for normality (unlikely)
###### QQ plot ------
qqtotalP <- ggplot(DFsummarycombinedorig, aes(sample=total_plastic_particles, colour = factor(species)))
  stat_qq() +
  stat_qq_line() +
  ggtitle("Total Plastic Particles")
qqblackP <- ggplot(DFsummarycombinedorig, aes(sample=black, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Black plastic")
qqblueP <- ggplot(DFsummarycombinedorig, aes(sample=blue, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Blue plastic")
qqredP <- ggplot(DFsummarycombinedorig, aes(sample=red, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Red plastic")
qqyellowP <- ggplot(DFsummarycombinedorig, aes(sample=yellow, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Yellow plastic")
qqcolourlessP <- ggplot(DFsummarycombinedorig, aes(sample=colourless, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Colourless plastic")
qqfragP <- ggplot(DFsummarycombinedorig, aes(sample=fragment_film_bead, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Fragment/Film/Bead")
```

```
qqfibreP <- ggplot(DFsummarycombinedorig, aes(sample=fibre_rod, colour = factor(species))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Fibre/Rod")

ggarrange (qqtotalP, qqblackP, qqblueP, qqredP, qqyellowP, qqcolourlessP, qqfragP, qqfibreP, "", ncol=3
```
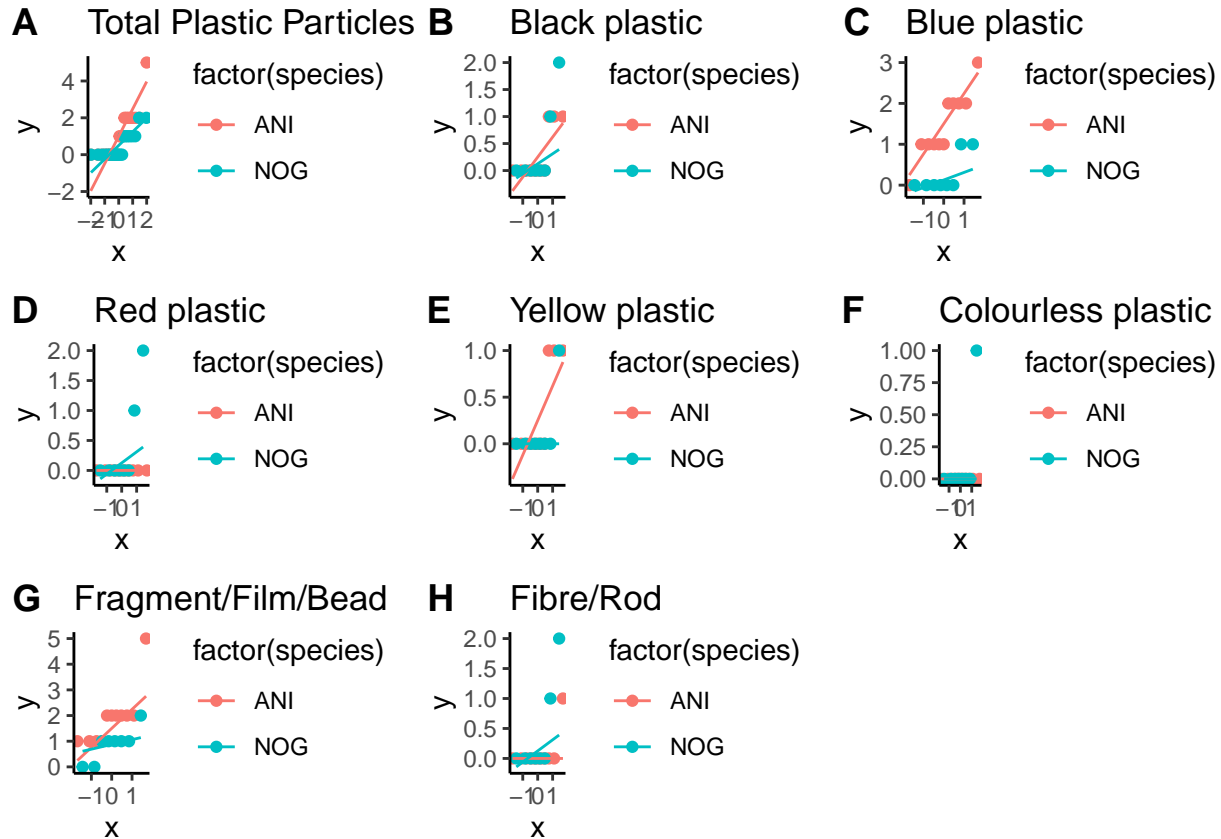

Not normal

Non-normal correlations with number of plastics against gut weight, whole weight, condition and mouth area

```
##with whole fish weight
cor.test(DFsummarycombinedorig$total_plastic_particles, DFsummarycombinedorig$weight_g, method="spearman
```

Checking for correlation in mouth area and weight against particle size

```
## Warning in cor.test.default(DFsummarycombinedorig$total_plastic_particles, :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
```

```
## 
## data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$weight_g
## S = 15287, p-value = 0.3232
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.154286
```

```
# Spearman's rank correlation rho
#
# data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$weight_g
# S = 15287, p-value = 0.3232
# alternative hypothesis: true rho is not equal to 0
# sample estimates:
#        rho
# -0.154286
```

```
## gut weight
cor.test(DFsummarycombinedorig$total_plastic_particles, DFsummarycombinedorig$gut_g, method="spearman")
```

```
## Warning in cor.test.default(DFsummarycombinedorig$total_plastic_particles, :
## Cannot compute exact p-value with ties
```

```
## 
##  Spearman's rank correlation rho
## 
## data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$gut_g
## S = 16713, p-value = 0.08971
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.2619586
```

```
# Spearman's rank correlation rho
#
# data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$gut_g
# S = 16713, p-value = 0.08971
# alternative hypothesis: true rho is not equal to 0
# sample estimates:
#        rho
# -0.2619586
```

```
## condition
cor.test(DFsummarycombinedorig$total_plastic_particles, DFsummarycombinedorig$condition, method="spearma
```

```
## Warning in cor.test.default(DFsummarycombinedorig$total_plastic_particles, :
## Cannot compute exact p-value with ties
```

```
## 
##  Spearman's rank correlation rho
## 
## data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$condition
```

```
## S = 15252, p-value = 0.3318
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.1515984
```

```
# Spearman's rank correlation rho
#
# data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$condition
# S = 15252, p-value = 0.3318
# alternative hypothesis: true rho is not equal to 0
# sample estimates:
#        rho
# -0.1515984
```

```
##mouth size
cor.test(DFsummarycombinedorig$total_plastic_particles, DFsummarycombinedorig$moutharea, method="spearma
```

```
## Warning in cor.test.default(DFsummarycombinedorig$total_plastic_particles, :
## Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$moutharea
## S = 11769, p-value = 0.4771
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.1113681
```

```
# Spearman's rank correlation rho
#
# data:  DFsummarycombinedorig$total_plastic_particles and DFsummarycombinedorig$moutharea
# S = 11769, p-value = 0.4771
# alternative hypothesis: true rho is not equal to 0
# sample estimates:
#       rho
# 0.1113681
```

not correlated

Figure 21: MP size correlations with fish morphometrics

```
spearmanpatchwork
```
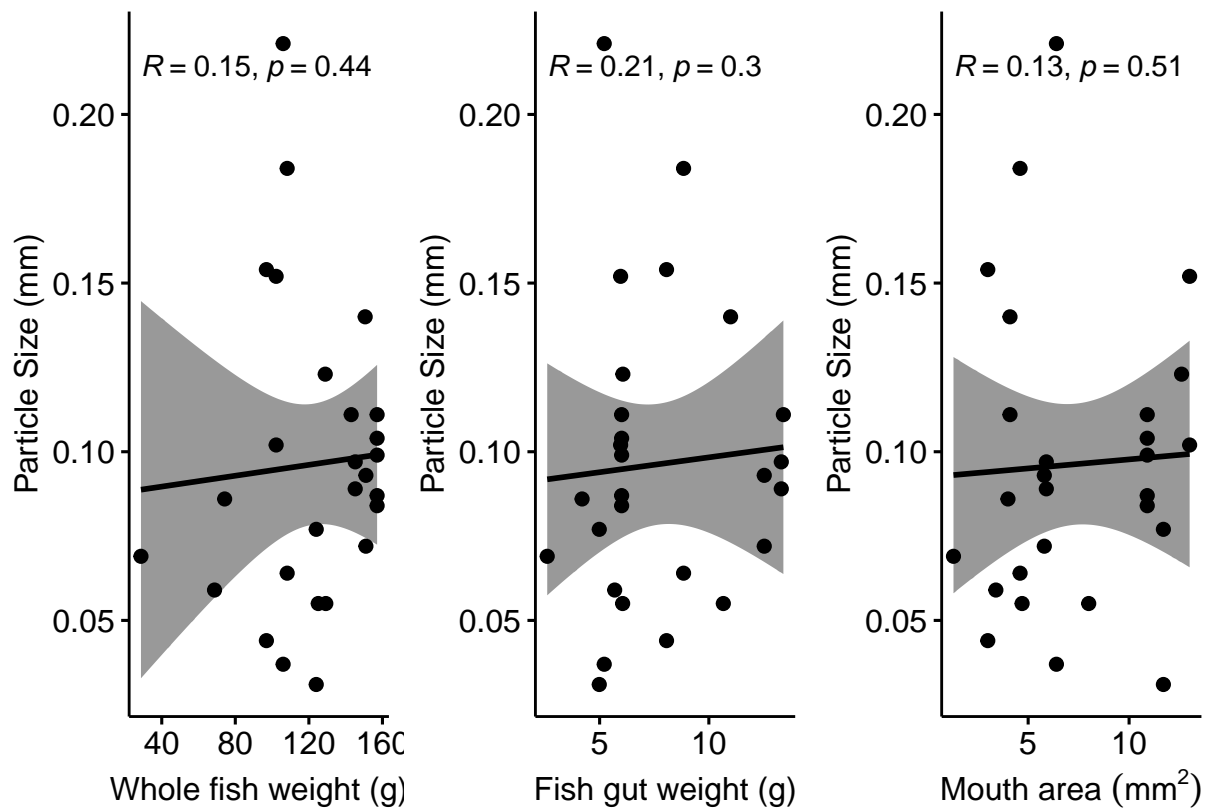
Export plot at 1000 * 300

Plastics type per species and location

Figure 22: microplastic counts in fish per polymer type, colour category, shape and size

Figure22

Testing idea of sig difference in particle size between locations   In the MDS, locations and particle size showed no significant difference (R = 0.026 (P = 33.1%) but 26 and 13 were significantly different in pairwise ANOSIMS. The initial comparison may not be significant because of the one particle at 53 and it would be good to exclude possible species bias, even though these were not significant in the ANOSIMS. So, looking at ANI (n=10 and n=12 for West (26) and southeast (13) respectively) as these are most similar in fish numbers.

Table 9: Total particle count test for differences   Looking at differences in median abundance across location, species weight, condition, mouth size and lab using the uncorrected data with Kruskal-wallis.

Kruskal-Wallis was used to compare the median microplastic ingestion across fish morphometrics (fish whole weight, body condition and mouth area categories - see Table 6).

```
##Kruskal-Wallis to test differences in location, species, weight, condition, mouth size and lab on tot
kruskal.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$event)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$event
## Kruskal-Wallis chi-squared = 3.868, df = 2, p-value = 0.1446
```

```
# Kruskal-Wallis rank sum test
#
```

```
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$event
# Kruskal-Wallis chi-squared = 3.868, df = 2, p-value = 0.1446
kruskal.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$species)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$species
## Kruskal-Wallis chi-squared = 1.7897, df = 1, p-value = 0.181
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$species
# Kruskal-Wallis chi-squared = 1.7897, df = 1, p-value = 0.181

##two groups for species so test also with Mann Whitney
wilcox.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$species)
```

```
## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot
## compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$species
## W = 280.5, p-value = 0.1854
## alternative hypothesis: true location shift is not equal to 0
```

```
# Wilcoxon rank sum test with continuity correction
#
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$species
# W = 280.5, p-value = 0.1854
# alternative hypothesis: true location shift is not equal to 0

kruskal.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$weight_category)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$weight_category
## Kruskal-Wallis chi-squared = 2.406, df = 4, p-value = 0.6615
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$weight_category
# Kruskal-Wallis chi-squared = 2.406, df = 4, p-value = 0.6615
kruskal.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$condition_category)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$condition_category
## Kruskal-Wallis chi-squared = 1.6759, df = 4, p-value = 0.7951
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$condition_category
# Kruskal-Wallis chi-squared = 1.6759, df = 4, p-value = 0.7951
kruskal.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$moutharea_category)
```

```
##
##   Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$moutharea_category
## Kruskal-Wallis chi-squared = 9.9515, df = 5, p-value = 0.07662
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$moutharea_category
# Kruskal-Wallis chi-squared = 9.9515, df = 5, p-value = 0.07662

## if you haven't before, combine species and event and test for differences
# DFsummarycombinedorig$species.event <- paste(DFsummarycombinedorig$event,DFsummarycombinedorig$specie

kruskal.test(DFsummarycombinedorig$total_plastic_particles~DFsummarycombinedorig$species.event)
```

```
##
##   Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$species.event
## Kruskal-Wallis chi-squared = 4.1232, df = 4, p-value = 0.3896
```

```
# Kruskal-Wallis rank sum test
#
# data:  DFsummarycombinedorig$total_plastic_particles by DFsummarycombinedorig$species.event
# Kruskal-Wallis chi-squared = 4.1232, df = 4,
# p-value = 0.3896

##Also test related to microplastic load
##create individual microplastic load (plastics/fish(1))
DFsummarycombinedorig$pl <- DFsummarycombinedorig$total_plastic_particles / 1

kruskal.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$event)
```

```
##
##   Kruskal-Wallis rank sum test
##
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$event
## Kruskal-Wallis chi-squared = 3.868, df = 2, p-value = 0.1446
```

```
# Kruskal-Wallis chi-squared = 3.868, df = 2, p-value = 0.1446
kruskal.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$species)
```

```
##
##   Kruskal-Wallis rank sum test
```

```
## 
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$species
## Kruskal-Wallis chi-squared = 1.7897, df = 1, p-value = 0.181
```

```
# Kruskal-Wallis chi-squared = 1.7897, df = 1, p-value = 0.18
wilcox.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$species)
```

```
## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot
## compute exact p-value with ties
```

```
## 
##  Wilcoxon rank sum test with continuity correction
## 
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$species
## W = 280.5, p-value = 0.1854
## alternative hypothesis: true location shift is not equal to 0
```

```
# W = 280.5, p-value = 0.1854
kruskal.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$species.event)
```

```
## 
##  Kruskal-Wallis rank sum test
## 
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$species.event
## Kruskal-Wallis chi-squared = 4.1232, df = 4, p-value = 0.3896
```

```
# Kruskal-Wallis chi-squared = 4.1232, df = 4, p-value = 0.3896
kruskal.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$weight_category)
```

```
## 
##  Kruskal-Wallis rank sum test
## 
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$weight_category
## Kruskal-Wallis chi-squared = 2.406, df = 4, p-value = 0.6615
```

```
# Kruskal-Wallis chi-squared = 2.406, df = 4, p-value = 0.6615
kruskal.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$condition_category)
```

```
## 
##  Kruskal-Wallis rank sum test
## 
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$condition_category
## Kruskal-Wallis chi-squared = 1.6759, df = 4, p-value = 0.7951
```

```
# Kruskal-Wallis chi-squared = 1.6759, df = 4, p-value = 0.7951
kruskal.test(DFsummarycombinedorig$pl~DFsummarycombinedorig$moutharea_category)
```

```
## 
##  Kruskal-Wallis rank sum test
## 
## data:  DFsummarycombinedorig$pl by DFsummarycombinedorig$moutharea_category
## Kruskal-Wallis chi-squared = 9.9515, df = 5, p-value = 0.07662
```

```
# Kruskal-Wallis chi-squared = 9.9515, df = 5, p-value = 0.07662
```

Control Data

Summary of control plastics: polymer, colour and shape

Read in data in the form of polymer, control type and plastic or composite and total columns

```
controls_polymer <- read.csv(file = "./data/controls_polymer.csv")
##Read in colour data across controls
controls_colour <- read.csv(file = "./data/controls_colour.csv")
##Read in shape data across controls
controls_shape <- read.csv(file = "./data/controls_shape.csv")


controlparticle_totals <- controls_polymer %>% group_by(type) %>% summarise(total=sum(total))
controlcolour_totals <- controls_colour %>% group_by(type) %>% summarise(total=sum(total))
controlshape_totals <- controls_shape %>% group_by(type) %>% summarise(total=sum(total))
```

Figure 18: Profile of plastics/composites found in controls

Stacked bar chart (from https://biostats.w.uib.no/stacking-data-series-in-bars/)

```
##polymer
controlpolystack <- ggplot(controls_polymer, aes(x = type, y = total, fill = polymer))+
  geom_col(width = .5) +
  geom_text(data=controlparticle_totals, aes(x=type, label = total, fill = NULL), nudge_y = 1, size = 3)
  scale_fill_viridis_d()+
  scale_x_discrete(limit = c("Atmospheric Control - plastic", "Atmospheric Control - composite", "Proce
                  labels = c("Atmospheric\nControl\nplastic","Atmospheric\nControl\ncomposite", "Proce
  labs(title = NULL, x = NULL, y = "Total particles")+
  theme(legend.position = "top", legend.title = element_blank())

##colour
particlecolours <- c("#000000", "#2B59C3", "#820933", "#98838F", "#CCD7C5", "#EC9F05")
controlcolourstack <- ggplot(controls_colour, aes(x = type, y = total, fill = colour))+
  geom_col(width = .5) +
  geom_text(data=controlcolour_totals, aes(x=type, label = total, fill = NULL), nudge_y = 1, size = 3)+
  scale_fill_manual(values = particlecolours)+
    scale_x_discrete(limit = c("Atmospheric Control - plastic", "Atmospheric Control - composite", "Pro
                    labels = c("Atmospheric\nControl\nplastic","Atmospheric\nControl\ncomposite", "Pro
    labs(title = NULL, x = NULL, y = "Total particles")+
    theme(legend.position = "top", legend.title = element_blank())

##shape
controlshapestack <- ggplot(controls_shape, aes(x = type, y = total, fill = shape))+
  geom_col(width = .5) +
  geom_text(data=controlshape_totals, aes(x=type, label = total, fill = NULL), nudge_y = 1, size = 3)+
    scale_fill_viridis_d()+
  scale_x_discrete(limit = c("Atmospheric Control - plastic", "Atmospheric Control - composite", "Proce
                  labels = c("Atmospheric\nControl\nplastic","Atmospheric\nControl\ncomposite", "Proce
  labs(title = NULL, x = NULL, y = "Total particles")+
```
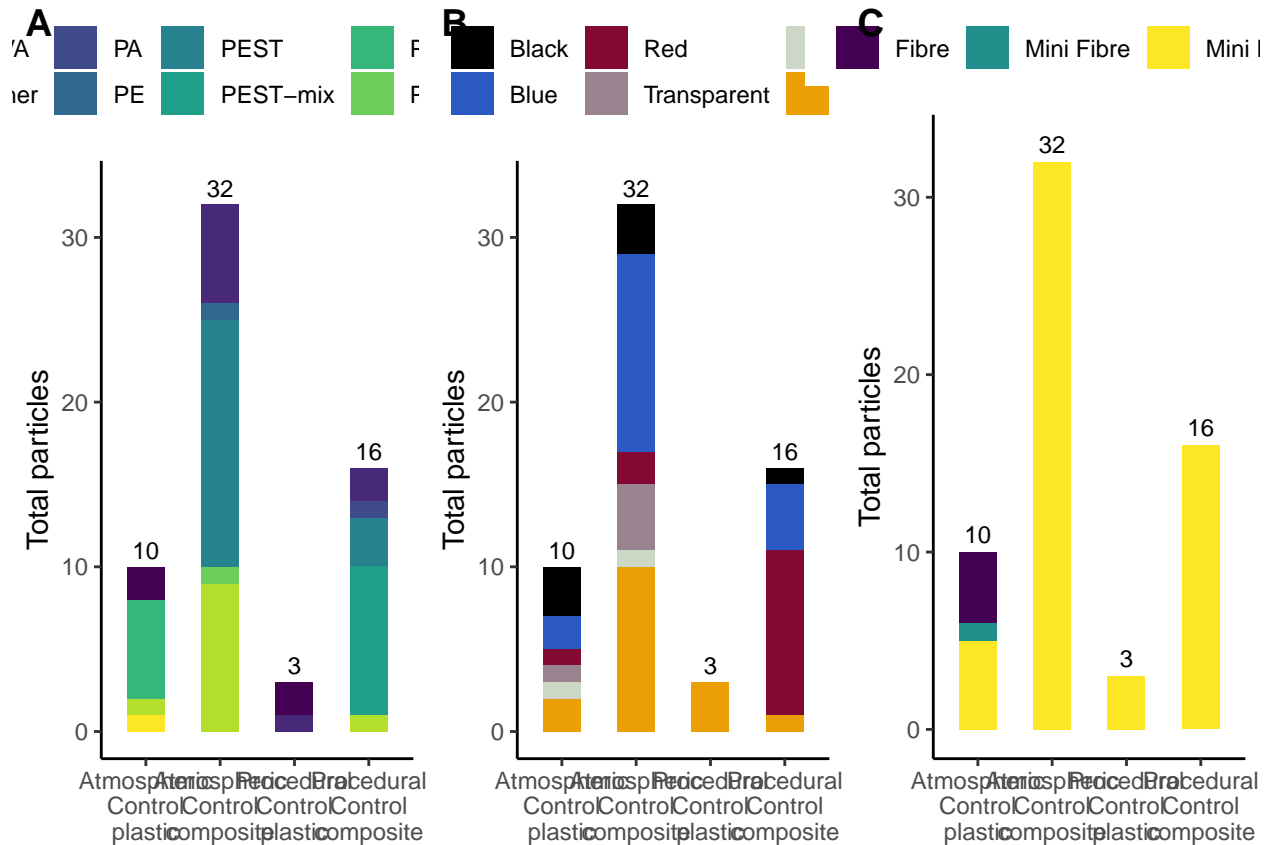
```
    theme(legend.position = "top", legend.title = element_blank())

##arrange plots 3 columns
ggarrange (controlpolystack, controlcolourstack, controlshapestack, nrow=1, ncol=3, labels = c("A","B","
```



```
## plot size = 1500 x 513
```

Check for differences in total plastic particles across labs in control filter papers

Other

Plastics and composites related to library match percentage and polymer type

Polymer per spectra match index stacked bar chart (from https://biostats.w.uib.no/stacking-data-series-in-bars/)

```
##load data file (possibly combine with particle data later to clean up)
DFspectra_match <- read.csv(file = "./data/spectra_match.csv")

legendtitlemps <- "Polymer"
matchpolystack <- ggplot(DFspectra_match, aes(x = match, y = orig, fill = polymer))+
  geom_col(width = .5) +
  geom_text(aes(label = paste(type)), colour = "white", position = position_stack(vjust = 0.5))+
```

```
  scale_fill_viridis_d(legendtitlemps)+
  scale_y_continuous(breaks = c(1:11), limits = c(0,11))+
  labs(title = NULL, x = "Spectra Match %", y = "Total particles")+
  theme(legend.position = "top")

legendtitlepms <- "Spectra Match %"
polymatchstack <- ggplot(DFspectra_match, aes(x = polymer, y = orig, fill = match))+
  geom_col(width = .5) +
  geom_text(aes(label = paste(type)), colour = "white", position = position_stack(vjust = 0.5))+
  scale_fill_viridis_d(legendtitlepms)+
  scale_y_continuous(breaks = c(1:11), limits = c(0,11))+
  labs(title = NULL, x = "Polymer", y = "Total particles")+
  theme(legend.position = "top")

ggarrange (matchpolystack, polymatchstack, nrow=1, ncol=2, labels = c("A","B"), common.legend = FALSE)
```
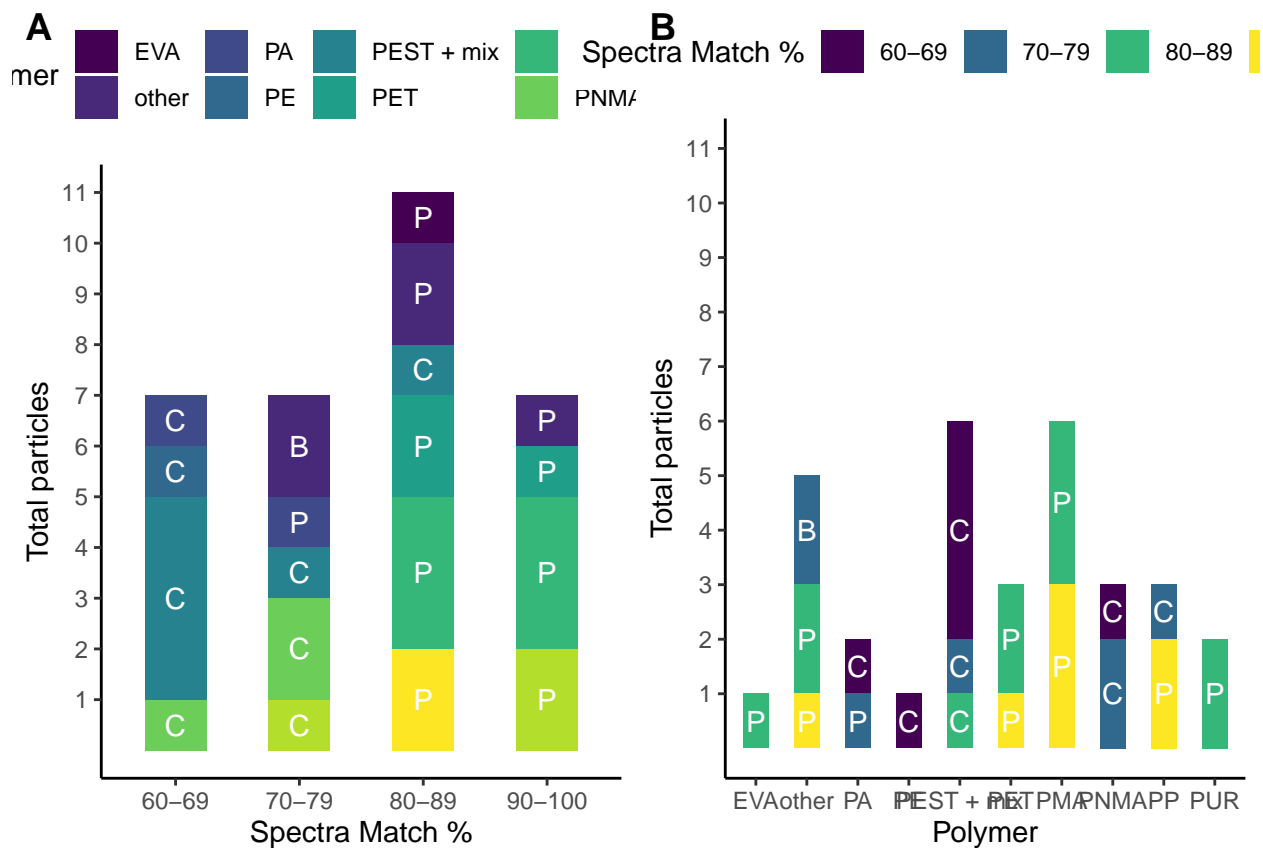

Looks like composite particles are much smaller than plastic?

```
ggplot(DFparticles, aes(x = size_mm))+
  geom_histogram(aes(fill=plastic_composite)) +
  ylab(NULL)+ ##count is obvious and looks better without
  xlab("Particle size (mm)")+
  theme(
    axis.text.x = element_text(size = rel(0.8)),
    axis.text.y = element_text(size = rel(0.8)),
```

```
    axis.title.x = element_text(margin = margin(r = 10), size = rel(0.8)),## increase space between axi
    legend.position = "top", legend.title=element_blank())+
  scale_fill_viridis_d()
```

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
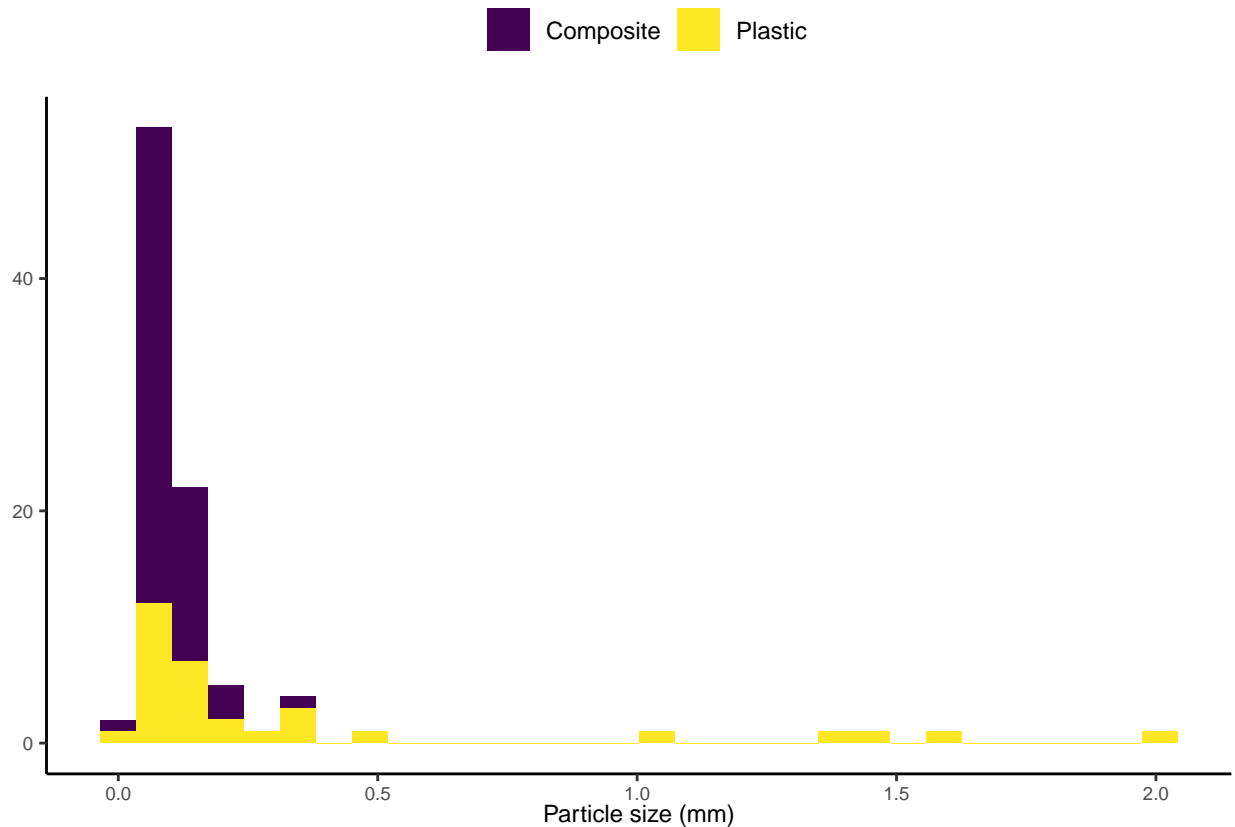
possibly an argument that identification relates to size due to clearer spectra, but only a few points scattered to the right, so not certain...box plot

```
ggplot(DFparticles, aes(group = plastic_composite, fill = plastic_composite, ##group variables as facto
                        y = size_mm,
                        factor(plastic_composite))) +
  geom_boxplot(aes()) +
  xlab(NULL)+ ##no x axis label as species already obvious
  ylab('particle size (mm)')+ ##y axis label
  theme(
    axis.title.y = element_text(margin = margin(r = 10)), ## increase space between axis labels and axi
    legend.position = "none"
  )+
  scale_fill_viridis_d()
```
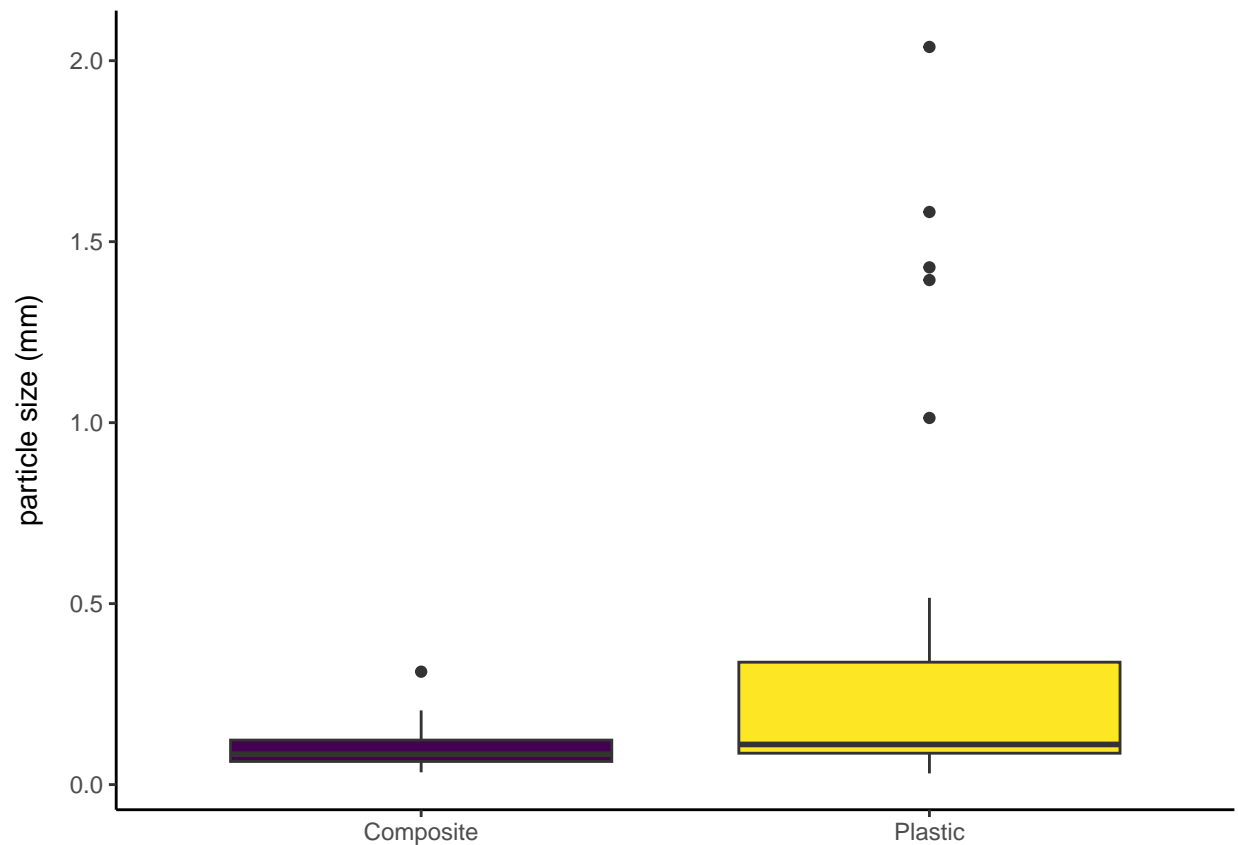
The majority are all very small but whether there is a significant difference between the groups, is difficult to tell.

```
##subset plastics from DFparticles and composites

DFparticlesplastic <- subset(DFparticles, plastic_composite == "Plastic")
DFparticlescomposite <- subset(DFparticles, plastic_composite == "Composite")

## Test for difference (non parametric two samples = Mann Whitney U)
wilcox.test(DFparticlesplastic$size_mm, DFparticlescomposite$size_mm)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  DFparticlesplastic$size_mm and DFparticlescomposite$size_mm
## W = 1350, p-value = 0.00252
## alternative hypothesis: true location shift is not equal to 0
```

The size of plastic particles significantly differs from the size of composite particles W = 1350, p<0.005

Just fish particles:

```
##subset dig only from DFparticlesplastic and composites

DFparticlesplasticdig<- subset(DFparticlesplastic, fp_type == "dig")
DFparticlescompositedig <- subset(DFparticlescomposite, fp_type == "dig")
```

```
## Test for difference (non parametric two samples = Mann Whitney U)
wilcox.test(DFparticlesplasticdig$size_mm, DFparticlescompositedig$size_mm)
```

```
## Warning in wilcox.test.default(DFparticlesplasticdig$size_mm,
## DFparticlescompositedig$size_mm): cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  DFparticlesplasticdig$size_mm and DFparticlescompositedig$size_mm
## W = 169, p-value = 0.08418
## alternative hypothesis: true location shift is not equal to 0
```

The particle size difference between plastic and composite particles found in fish is not significant

Boxplot of the fish particles:

```
ggplot(DFparticlesfish, aes(group = plastic_composite, fill = plastic_composite, ##group variables as f
                            y = size_mm,
                            factor(plastic_composite))) +
  geom_boxplot(aes()) +
  xlab(NULL)+ ##no x axis label as species already obvious
  ylab('particle size (mm)')+ ##y axis label
  theme(
    axis.title.y = element_text(margin = margin(r = 10)), ## increase space between axis labels and axi
    legend.position = "none"
  )+
  scale_fill_viridis_d()
```
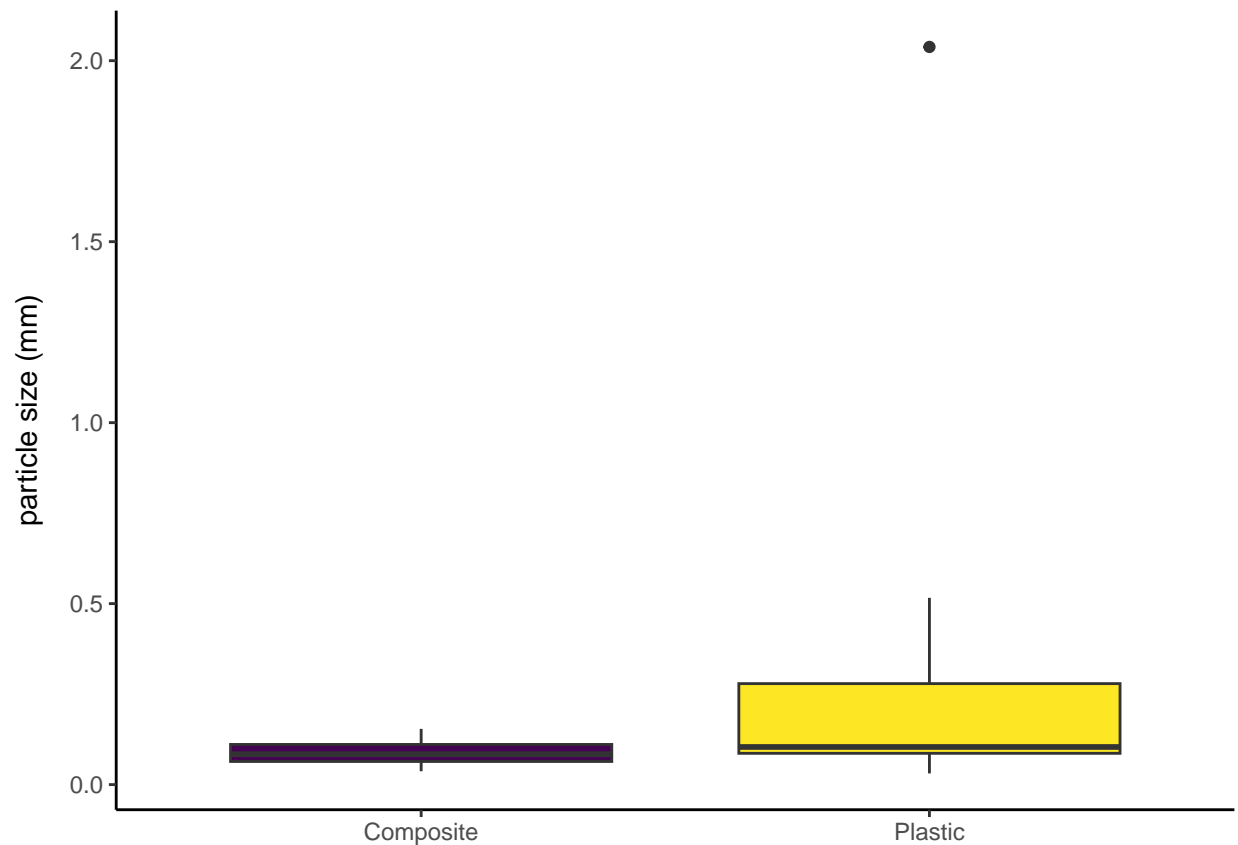
Plastics in Control Papers

```
## Plastics in controls
```

```
atmosfp <- c(26,143,52,18,102,16,73,28,2,147,13,82,3)
```

```
summary(atmosfp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00   16.00   28.00   54.23   82.00  147.00
```

```
sd(atmosfp)
```

```
## [1] 50.99698
```

```
IQR(atmosfp)
```

```
## [1] 66
```

```
procfp <- c(1,11,6,1,0,5,20,0,6,3,0,4,4,13,0,1,2,3,4,6,0,3,1,0)
```

```
summary(procfp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.750   3.000   3.917   5.250  20.000
```

```
sd(procfp)
```

```
## [1] 4.835707
```

```
IQR(procfp)
```

```
## [1] 4.5
```