

# Thesis Goal 1

Eleanor (Ella) Crotty

## Contents

Goals	1
Import data	1
Calculate eDNA index	3
Combine eDNA index with environmental data	7
Re-calculate eDNA index for 2021-2022 only	7
Plot eDNA index over time and oxygen	11
Scatterplot eDNA vs hypoxic threshold . . . . .	44
GAM of eDNA index vs oxygen	47
Heatmap of DO x Temp x eDNA	82

```
# Warnings and startup messages suppressed
library(tidyverse)
library(patchwork)
library(scales)
library(ggrepel)
library(readxl)
library(here)
library(mgcv)
library(gratia)
```

Goal 1: Compare copepod eDNA index, a measure of relative abundance using eDNA detections, to dissolved oxygen data in order to assess whether hypoxia decreases copepod abundance in OCNMS.

## Goals

- Plot eDNA index over time and oxygen
- Make scatterplots of eDNA index vs oxygen
- Do a GAM of eDNA index vs oxygen
- Compare eDNA to hypoxic thresholds

## Import data

```
envData <- read_csv(here("PMEL-Data", "EnvironmentalDataset1_copy.csv")) %>%
    filter(year != 2023) %>% # Ignoring 2023 due to gaps for now
```

```

  mutate(year = as.factor(year)) %>%
  relocate(year, .after = date)

## New names:
## Rows: 38938 Columns: 11
## -- Column specification
## ----- Delimiter: ","
## (1): source dbl (8): ...1, year, temperature, DO, salinity, potential_density,
## pres, cond dttm (2): date, sampleID
## 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.
## * `...1` -> `...1`

allReads <- read_csv(here("PMEL-Data", "FishPlusCOI_Reads_copy.csv")) %>%
  mutate(year = year(Date_local)) %>%
  relocate(year, .after = Date_local)

## Rows: 1094324 Columns: 71
## -- Column specification -----
## Delimiter: ","
## chr (39): ESV, sequence, Kingdom, Phylum, Class, Order, Family, Genus, Spec...
## dbl (27): X, pctMatch, nReads, Biological_Replicate, Technical_Replicate, D...
## lgl (3): Negative_control, Positive_control, primers_pad
## dttm (2): Date.UTC, Date_local
##
## 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.

yrReads <- allReads %>%
  filter(year != 2023) %>% # Ignoring 2023 due to gaps for now
  mutate(year = as.factor(year))
copepodNames <- read_csv(here("PMEL-Data", "OCNMS_Copepods_Krill_copy.csv"))

## Rows: 36 Columns: 2
## -- Column specification -----
## Delimiter: ","
## chr (2): Species, Notes
##
## 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.

eDNAxEnvDataSat <- read_csv(here("PMEL-Data", "eDNAxEnvDataSat_copy.csv"))

## New names:
## Rows: 80442 Columns: 81
## -- Column specification
## ----- Delimiter: ","
## (36): Kingdom, Phylum, Class, Order, Family, Genus, Species, JV243.2, B... dbl
## (35): TotalnReads, Biological_Replicate, Depth_m, Sample_volume_ml, Lat... lgl
## (5): Present, Negative_control, Positive_control, primers_pad, sampleID dttm
## (5): Date.UTC, Date_local, DateMatch, Date_local_hr, date
## 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.
## * `...1` -> `...68`

```

```

copepodFiltered <- read_csv(here("CopepodDetectionsFiltered.csv"))

## Rows: 763 Columns: 82
## -- Column specification -----
## Delimiter: ","
## chr (36): Kingdom, Phylum, Class, Order, Family, Genus, Species, JV243.2, B...
## dbl (35): TotalnReads, Biological_Replicate, Depth_m, Sample_volume_ml, Lat...
## lgl (6): Barcode_mod, Present, Negative_control, Positive_control, primers...
## dttm (5): Date.UTC, Date_local, DateMatch, Date_local_hr, date
##
## 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.

copepodFull <- read_csv(here("CopepodDetectionsFull.csv"))

## Rows: 3488 Columns: 82
## -- Column specification -----
## Delimiter: ","
## chr (36): Kingdom, Phylum, Class, Order, Family, Genus, Species, JV243.2, B...
## dbl (35): TotalnReads, Biological_Replicate, Depth_m, Sample_volume_ml, Lat...
## lgl (6): Barcode_mod, Present, Negative_control, Positive_control, primers...
## dttm (5): Date.UTC, Date_local, DateMatch, Date_local_hr, date
##
## 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.

```

## Calculate eDNA index

```

copepods <- unique(copepodFull$Species)
copepodReads <- allReads %>%
  filter(Species %in% copepods) # Just in case

index1dummy <- allReads %>% # Quick proportions
dplyr::group_by(SampleId) %>%
  mutate(Tot = sum(nReads),
        Row.sums = nReads / Tot) %>% # calculate proportions - 0 reads/0 total = NaN,
  → need to replace with 0 to make max() work
  relocate(c(Tot, Row.sums), .after = SampleId)

# First, combine PCR replicates and average
PCR_reps_combine <- function(df) {
  # print(head(df))

  # Separate out Sample_Name into three informational columns
  df_out <- df %>%
    separate(Sample_Name,
            into=c("E_no", "Cruise1", "PCR_Rep"),
            remove=F,
            fill = "right") %>% # sep is a regular expression which is annoying, but the
  → default recognizes any non-alphanumeric characters so the default works
  → here
    mutate(.)

```

```

# Define which columns to group by
id_cols1 <- colnames(df_out) # Pull out column names
id_cols1 <- id_cols1[! id_cols1 %in% c("X", "SampleId", "Sample_Name", "PCR_Rep",
→ "pctMatch", "JV_Sample_Name", "Technical_Replicate", "nReads", "Barcode.y")] # Remove
→ SampleId and Sample_Name, as well as other things that differ by PCR replicate
print(id_cols1)

df_out <- df_out %>%
  group_by_at(id_cols1) %>%
  # dplyr::mutate(SampleId = dplyr::first(SampleId)) %>% # Just ignoring sampleid for
  → now and using E-no as unique
  summarize(nReads = mean(nReads)) %>%
  relocate(nReads, .after = PI)

df_out
}

allReads_PCRcomb <- PCR_reps_combine(allReads) # Spot check E1325 was correct

## [1] "ESV"
## [2] "sequence"
## [3] "Kingdom"
## [4] "Phylum"
## [5] "Class"
## [6] "Order"
## [7] "Family"
## [8] "Genus"
## [9] "Species"
## [10] "JV243.2"
## [11] "Barcode.x"
## [12] "Barcode_mod"
## [13] "PI"
## [14] "E_no"
## [15] "Cruise1"
## [16] "Biological_Replicate"
## [17] "Negative_control"
## [18] "Positive_control"
## [19] "Cruise_ID_short"
## [20] "Cruise_ID_long"
## [21] "Field_Sample_Name"
## [22] "Cast_No."
## [23] "Rosette_position"
## [24] "Field_collection_method"
## [25] "Preservation"
## [26] "Area.within.region"
## [27] "Station"
## [28] "Depth_m"
## [29] "Sample_volume_ml"
## [30] "Personnel_responsible_for_Collecting_Sample"
## [31] "Field_Notes"
## [32] "Date.UTC"
## [33] "Date_local"
## [34] "year"

```

```

## [35] "Lat_dec"
## [36] "Lon_dec"
## [37] "Run"
## [38] "Client"
## [39] "ClientId"
## [40] "BatchId"
## [41] "Replicate"
## [42] "Amplicon"
## [43] "F_name"
## [44] "primers_fwd"
## [45] "R_name"
## [46] "primers_rev"
## [47] "PE"
## [48] "primers_pad"
## [49] "primers_errors"
## [50] "merge_maxdiffs"
## [51] "merge_maxdiffpct"
## [52] "merge_minovlen"
## [53] "ExpL"
## [54] "merge_minlen"
## [55] "merge_maxlen"
## [56] "filter_max_ee"
## [57] "denoise_alpha"
## [58] "denoise_minsize"
## [59] "tax_ID"
## [60] "tax_maxaccepts"
## [61] "tax_maxrejects"
## [62] "tax_maxhits"
## [63] "tax_qcov"
## [64] "tax_minwords"
## [65] "tax_wordlen"
## [66] "Marker"

## `summarise()` has grouped output by 'ESV', 'sequence', 'Kingdom', 'Phylum',
## 'Class', 'Order', 'Family', 'Genus', 'Species', 'JV243.2', 'Barcode.x',
## 'Barcode_mod', 'PI', 'E_no', 'Cruise1', 'Biological_Replicate',
## 'Negative_control', 'Positive_control', 'Cruise_ID_short', 'Cruise_ID_long',
## 'Field_Sample_Name', 'Cast_No.', 'Rosette_position', 'Field_collection_method',
## 'Preservation', 'Area.within.region', 'Station', 'Depth_m', 'Sample_volume_ml',
## 'Personnel_responsible_for_Collecting_Sample', 'Field_Notes', 'Date.UTC',
## 'Date_local', 'year', 'Lat_dec', 'Lon_dec', 'Run', 'Client', 'ClientId',
## 'BatchId', 'Replicate', 'Amplicon', 'F_name', 'primers_fwd', 'R_name',
## 'primers_rev', 'PE', 'primers_pad', 'primers_errors', 'merge_maxdiffs',
## 'merge_maxdiffpct', 'merge_minovlen', 'ExpL', 'merge_minlen', 'merge_maxlen',
## 'filter_max_ee', 'denoise_alpha', 'denoise_minsize', 'tax_ID',
## 'tax_maxaccepts', 'tax_maxrejects', 'tax_maxhits', 'tax_qcov', 'tax_minwords',
## 'tax_wordlen'. You can override using the `groups` argument.

# Second, combine by species

id_cols <- colnames(allReads_PCRcomb) # Pull out column names
id_cols <- id_cols[! id_cols %in% c("X", "ESV", "sequence", "nReads")] # Remove ESV +
# nreads because those are different within species

allReads_species <- allReads_PCRcomb %>%

```

```

group_by_at(id_cols) %>% # group_by_at can take a vector
summarize(TotalnReads = sum(nReads)) %>% # Removed , avgpctMatch = mean(pctMatch)
  ↪ because I had to remove pctMatch to combine PCR replicates. Sum nReads results in
  ↪ taking the sum of all ESVs within a species
relocate(TotalnReads, .after = PI)

## `summarise()` has grouped output by 'Kingdom', 'Phylum', 'Class', 'Order',
## 'Family', 'Genus', 'Species', 'JV243.2', 'Barcode.x', 'Barcode_mod', 'PI',
## 'E_no', 'Cruise1', 'Biological_Replicate', 'Negative_control',
## 'Positive_control', 'Cruise_ID_short', 'Cruise_ID_long', 'Field_Sample_Name',
## 'Cast_No.', 'Rosette_position', 'Field_collection_method', 'Preservation',
## 'Area.within.region', 'Station', 'Depth_m', 'Sample_volume_ml',
## 'Personnel_responsible_for_Collecting_Sample', 'Field_Notes', 'Date_UTC',
## 'Date_local', 'year', 'Lat_dec', 'Lon_dec', 'Run', 'Client', 'ClientId',
## 'BatchId', 'Replicate', 'Amplicon', 'F_name', 'primers_fwd', 'R_name',
## 'primers_rev', 'PE', 'primers_pad', 'primers_errors', 'merge_maxdiffs',
## 'merge_maxdiffpct', 'merge_minovlen', 'Expl', 'merge_minlen', 'merge maxlen',
## 'filter_max_ee', 'denoise_alpha', 'denoise_minsize', 'tax_ID',
## 'tax_maxaccepts', 'tax_maxrejects', 'tax_maxhits', 'tax_qcov', 'tax_minwords',
## 'tax_wordlen'. You can override using the `groups` argument.

write_csv(allReads_species, here("eDNA_Index_Hypoxia", "Data",
  ↪ "CopepodReads_Species.csv")) # Gonna want this later

all_index1 <- allReads_species %>%
  dplyr::group_by(E_no) %>% # Group by E-number, beacuse sampleID had to be removed in
  ↪ the PCR replicate combination step
  mutate(Tot = sum(TotalnReads),
    Row.sums = TotalnReads / Tot) %>% # calculate proportions - 0 reads/0 total =
    ↪ NaN, need to replace with 0 to make max() work
  relocate(c(TotalnReads, Tot, Row.sums), .after = E_no) # Move it somewhere I can see
  ↪ the damn thing

all_eDNA_index <- all_index1 %>%
  dplyr::group_by(Species) %>%
  mutate(Row.sums = case_when(Row.sums == "NaN" ~ 0,
    .default = Row.sums)) %>% # Make 0/0 = 0 and not
    ↪ NaN
  mutate(Colmax = max(Row.sums), Normalized.reads = Row.sums / Colmax) %>% #transforms
  ↪ raw number of reads to eDNA index. Creates same divide by 0 error, so:
  mutate(Normalized.reads = case_when(Normalized.reads == "NaN" ~ 0,
    .default = Normalized.reads)) %>% # Make 0/0 = 0
    ↪ and not NaN
  relocate(c(Colmax, Normalized.reads), .after = Row.sums) # Move it somewhere I can see
  ↪ the damn thing

# Filter to only copepods
copepod_eDNA_index <- all_eDNA_index %>%
  filter(Species %in% copepods) %>%
  rename(eDNA_index = `Normalized.reads`) # to make parsing this easier

write_csv(copepod_eDNA_index, here("eDNA_Index_Hypoxia", "Data",
  ↪ "Copepod_eDNA_index.csv")) # Gonna want this later

```

## Combine eDNA index with environmental data

```
# envData = EnvironmentalDataset1
# Based on eDNAxp02.Rmd from summer project
# copepod_eDNA_index is still essentially a version of allReads so I can use the same
# code I did with allReads_Species

EnvRd <- envData %>%
  mutate(DateMatch = round_date(date, unit = "10 minutes")) # Round to the nearest 10
# minutes
DetectRd <- copepod_eDNA_index %>%
  mutate(DateMatch = round_date(Date_UTC, unit = "10 minutes"), Date_local_hr =
# round_date(Date_local, unit = "hour")) # Spot check - looks good.

eDNAindxEnvData <- left_join(DetectRd, EnvRd, by = join_by(DateMatch)) %>%
  relocate(date, .after = Date_UTC) %>%
  relocate(year.x, .after = Date_UTC) %>%
  relocate(year.y, .after = Date_UTC) %>%
  filter(year.x != 2023) %>%
  filter()

investigate <- eDNAindxEnvData %>% select(Species, DateMatch, Date_UTC, Date_local_hr,
# source, temperature, DO, E_no, Rosette_position, Amplicon)

system("say Data Join Complete")

# Export the joined data
write_csv(eDNAindxEnvData, here("eDNA_Index_Hypoxia", "Data",
# "Copepod_eDNAindxEnvData.csv"))

# Make a version without the copepods that are only detected in 2023 because those plots
# are unhelpful
eDNAindxEnvData_clean <- eDNAindxEnvData %>%
  filter(Colmax > 0)

write_csv(eDNAindxEnvData_clean, here("eDNA_Index_Hypoxia", "Data",
# "Copepod_eDNAindxEnvData_clean.csv"))
```

## Re-calculate eDNA index for 2021-2022 only

```
index1dummyYr <- yrReads %>% # Quick proportions
dplyr::group_by(SampleId) %>%
  mutate(Tot = sum(nReads),
    Row.sums = nReads / Tot) %>% # calculate proportions - 0 reads/0 total = NaN,
# need to replace with 0 to make max() work
  relocate(c(Tot, Row.sums), .after = SampleId)

yrReads_PCRcomb <- PCR_reps_combine(yrReads) # Spot check E1325 was correct

## [1] "ESV"
## [2] "sequence"
## [3] "Kingdom"
```

```

## [4] "Phylum"
## [5] "Class"
## [6] "Order"
## [7] "Family"
## [8] "Genus"
## [9] "Species"
## [10] "JV243.2"
## [11] "Barcode.x"
## [12] "Barcode_mod"
## [13] "PI"
## [14] "E_no"
## [15] "Cruise1"
## [16] "Biological_Replicate"
## [17] "Negative_control"
## [18] "Positive_control"
## [19] "Cruise_ID_short"
## [20] "Cruise_ID_long"
## [21] "Field_Sample_Name"
## [22] "Cast_No."
## [23] "Rosette_position"
## [24] "Field_collection_method"
## [25] "Preservation"
## [26] "Area.within.region"
## [27] "Station"
## [28] "Depth_m"
## [29] "Sample_volume_ml"
## [30] "Personnel_responsible_for_Collecting_Sample"
## [31] "Field_Notes"
## [32] "Date.UTC"
## [33] "Date_local"
## [34] "year"
## [35] "Lat_dec"
## [36] "Lon_dec"
## [37] "Run"
## [38] "Client"
## [39] "ClientId"
## [40] "BatchId"
## [41] "Replicate"
## [42] "Amplicon"
## [43] "F_name"
## [44] "primers_fwd"
## [45] "R_name"
## [46] "primers_rev"
## [47] "PE"
## [48] "primers_pad"
## [49] "primers_errors"
## [50] "merge_maxdiffs"
## [51] "merge_maxdiffpct"
## [52] "merge_minovlen"
## [53] "ExpL"
## [54] "merge_minlen"
## [55] "merge_maxlen"
## [56] "filter_max_ee"
## [57] "denoise_alpha"

```

```

## [58] "denoise_minsize"
## [59] "tax_ID"
## [60] "tax_maxaccepts"
## [61] "tax_maxrejects"
## [62] "tax_maxhits"
## [63] "tax_qcov"
## [64] "tax_minwords"
## [65] "tax_wordlen"
## [66] "Marker"

## `summarise()` has grouped output by 'ESV', 'sequence', 'Kingdom', 'Phylum',
## 'Class', 'Order', 'Family', 'Genus', 'Species', 'JV243.2', 'Barcode.x',
## 'Barcode_mod', 'PI', 'E_no', 'Cruise1', 'Biological_Replicate',
## 'Negative_control', 'Positive_control', 'Cruise_ID_short', 'Cruise_ID_long',
## 'Field_Sample_Name', 'Cast_No.', 'Rosette_position', 'Field_collection_method',
## 'Preservation', 'Area.within.region', 'Station', 'Depth_m', 'Sample_volume_ml',
## 'Personnel_responsible_for_Collecting_Sample', 'Field_Notes', 'Date_UTC',
## 'Date_local', 'year', 'Lat_dec', 'Lon_dec', 'Run', 'Client', 'ClientId',
## 'BatchId', 'Replicate', 'Amplicon', 'F_name', 'primers_fwd', 'R_name',
## 'primers_rev', 'PE', 'primers_pad', 'primers_errors', 'merge_maxdiffs',
## 'merge_maxdiffpct', 'merge_minovlen', 'ExpL', 'merge_minlen', 'merge_maxlen',
## 'filter_max_ee', 'denoise_alpha', 'denoise_minsize', 'tax_ID',
## 'tax_maxaccepts', 'tax_maxrejects', 'tax_maxhits', 'tax_qcov', 'tax_minwords',
## 'tax_wordlen'. You can override using the `groups` argument.

# Second, combine by species

id_cols <- colnames(yrReads_PCRcomb) # Pull out column names
id_cols <- id_cols[! id_cols %in% c("X", "ESV", "sequence", "nReads")] # Remove ESV +
# nreads because those are different within species

allReads_speciesYr <- yrReads_PCRcomb %>%
  group_by_at(id_cols) %>% # group_by_at can take a vector
  summarize(TotalnReads = sum(nReads)) %>% # Removed , avgpctMatch = mean(pctMatch)
# because I had to remove pctMatch to combine PCR replicates. Sum nReads results in
# taking the sum of all ESVs within a species
  relocate(TotalnReads, .after = PI)

## `summarise()` has grouped output by 'Kingdom', 'Phylum', 'Class', 'Order',
## 'Family', 'Genus', 'Species', 'JV243.2', 'Barcode.x', 'Barcode_mod', 'PI',
## 'E_no', 'Cruise1', 'Biological_Replicate', 'Negative_control',
## 'Positive_control', 'Cruise_ID_short', 'Cruise_ID_long', 'Field_Sample_Name',
## 'Cast_No.', 'Rosette_position', 'Field_collection_method', 'Preservation',
## 'Area.within.region', 'Station', 'Depth_m', 'Sample_volume_ml',
## 'Personnel_responsible_for_Collecting_Sample', 'Field_Notes', 'Date_UTC',
## 'Date_local', 'year', 'Lat_dec', 'Lon_dec', 'Run', 'Client', 'ClientId',
## 'BatchId', 'Replicate', 'Amplicon', 'F_name', 'primers_fwd', 'R_name',
## 'primers_rev', 'PE', 'primers_pad', 'primers_errors', 'merge_maxdiffs',
## 'merge_maxdiffpct', 'merge_minovlen', 'ExpL', 'merge_minlen', 'merge_maxlen',
## 'filter_max_ee', 'denoise_alpha', 'denoise_minsize', 'tax_ID',
## 'tax_maxaccepts', 'tax_maxrejects', 'tax_maxhits', 'tax_qcov', 'tax_minwords',
## 'tax_wordlen'. You can override using the `groups` argument.

write_csv(allReads_speciesYr, here("eDNA_Index_Hypoxia", "Data",
# "CopepodReads_Species_no23.csv")) # Gonna want this later

```

```

all_index1Yr <- allReads_speciesYr %>%
  dplyr::group_by(E_no) %>% # Group by E-number, beacuse sampleID had to be removed in
  → the PCR replicate combination step
  mutate(Tot = sum(TotalnReads),
    Row.sums = TotalnReads / Tot) %>% # calculate proportions - 0 reads/0 total =
    → NaN, need to replace with 0 to make max() work
  relocate(c(TotalnReads, Tot, Row.sums), .after = E_no) # Move it somewhere I can see
  → the damn thing

yr_eDNA_index <- all_index1Yr %>%
  dplyr::group_by(Species) %>%
  mutate(Row.sums = case_when(Row.sums == "NaN" ~ 0,
    .default = Row.sums)) %>% # Make 0/0 = 0 and not
    → NaN
  mutate(Colmax = max(Row.sums), Normalized.reads = Row.sums / Colmax) %>% #transforms
  → raw number of reads to eDNA index. Creates same divide by 0 error, so:
  mutate(Normalized.reads = case_when(Normalized.reads == "NaN" ~ 0,
    .default = Normalized.reads)) %>% # Make 0/0 = 0
    → and not NaN
  relocate(c(Colmax, Normalized.reads), .after = Row.sums) # Move it somewhere I can see
  → the damn thing

# Filter to only copepods
copepod_eDNA_indexYr <- yr_eDNA_index %>%
  filter(Species %in% copepods) %>%
  rename(eDNA_index = `Normalized.reads`) # to make parsing this easier

write_csv(copepod_eDNA_indexYr, here("eDNA_Index_Hypoxia", "Data",
  → "Copepod_eDNA_index_no23.csv")) # Gonna want this later

# envData = EnvironmentalDataset1
# Based on eDNAxp02.Rmd from summer project
# copepod_eDNA_index is still essentially a version of allReads so I can use the same
  → code I did with allReads_Species

DetectRdYr <- copepod_eDNA_indexYr %>%
  mutate(DateMatch = round_date(Date_UTC, unit = "10 minutes"), Date_local_hr =
  → round_date(Date_local, unit = "hour")) # Spot check - looks good.

eDNAindxEnvDataYr <- left_join(DetectRdYr, EnvRd, by = join_by(DateMatch)) %>%
  relocate(date, .after = Date_UTC) %>%
  relocate(year.x, .after = Date_UTC) %>%
  relocate(year.y, .after = Date_UTC) %>%
  filter(year.x != 2023) %>%
  filter()

investigateYr <- eDNAindxEnvDataYr %>% select(Species, DateMatch, Date_UTC,
  → Date_local_hr, source, temperature, DO, E_no, Rosette_position, Amplicon)

system("say Data Join Without 2023 Complete")

# Export the joined data

```

```

write_csv(eDNAindxEnvDataYr, here("eDNA_Index_Hypoxia", "Data",
  ↵  "Copepod_eDNAindxEnvData_no23.csv"))

# Make a version without the copepods that are only detected in 2023 because those plots
# are unhelpful
eDNAindxEnvData_cleanYr <- eDNAindxEnvDataYr %>%
  filter(Colmax > 0)

write_csv(eDNAindxEnvData_cleanYr, here("eDNA_Index_Hypoxia", "Data",
  ↵  "Copepod_eDNAindxEnvData_clean_no23.csv"))

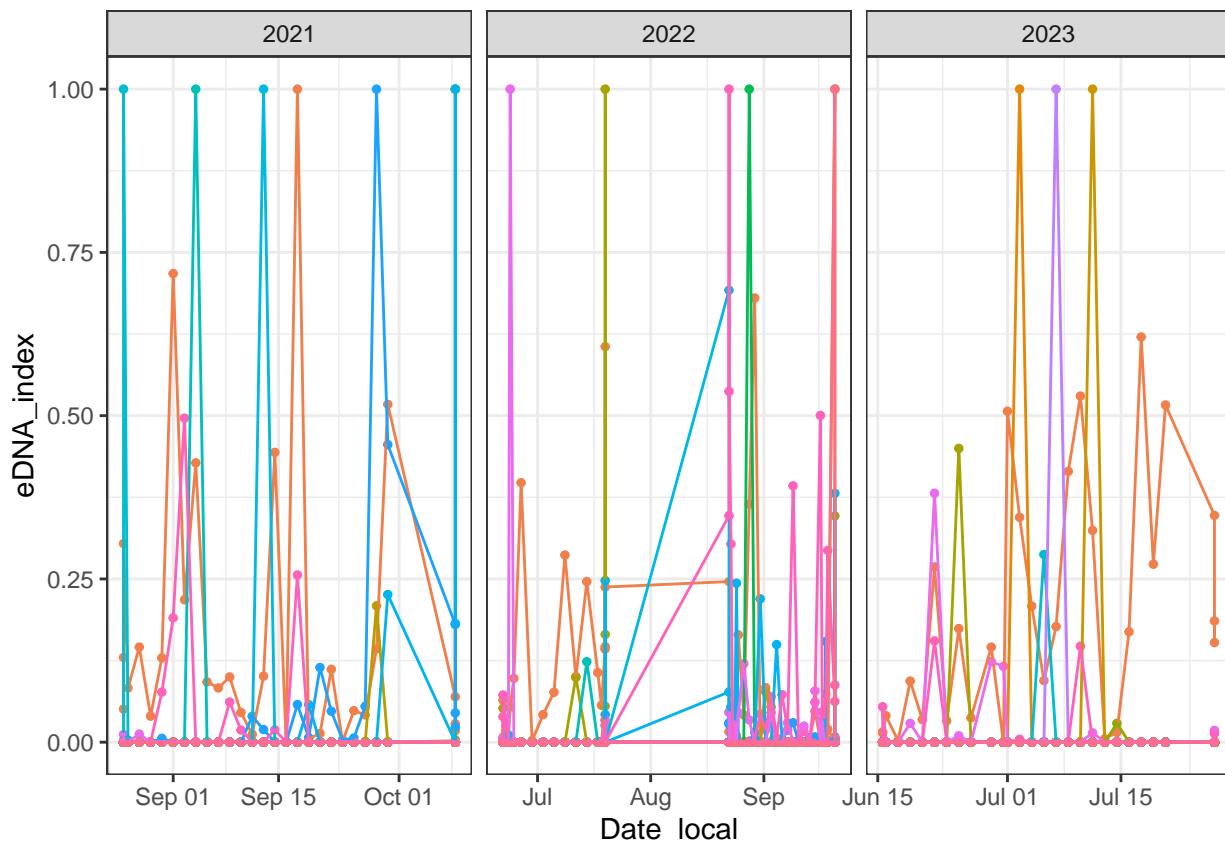
```

## Plot eDNA index over time and oxygen

```

# Explore eDNA index by species
ggplot(copepod_eDNA_index, aes(x = Date_local, y = eDNA_index, color = Species)) +
  geom_point(show.legend = F, size = 1) +
  facet_wrap(facets = vars(year(Date_local)), scales = "free_x") +
  geom_line(show.legend = F) +
  theme_bw()

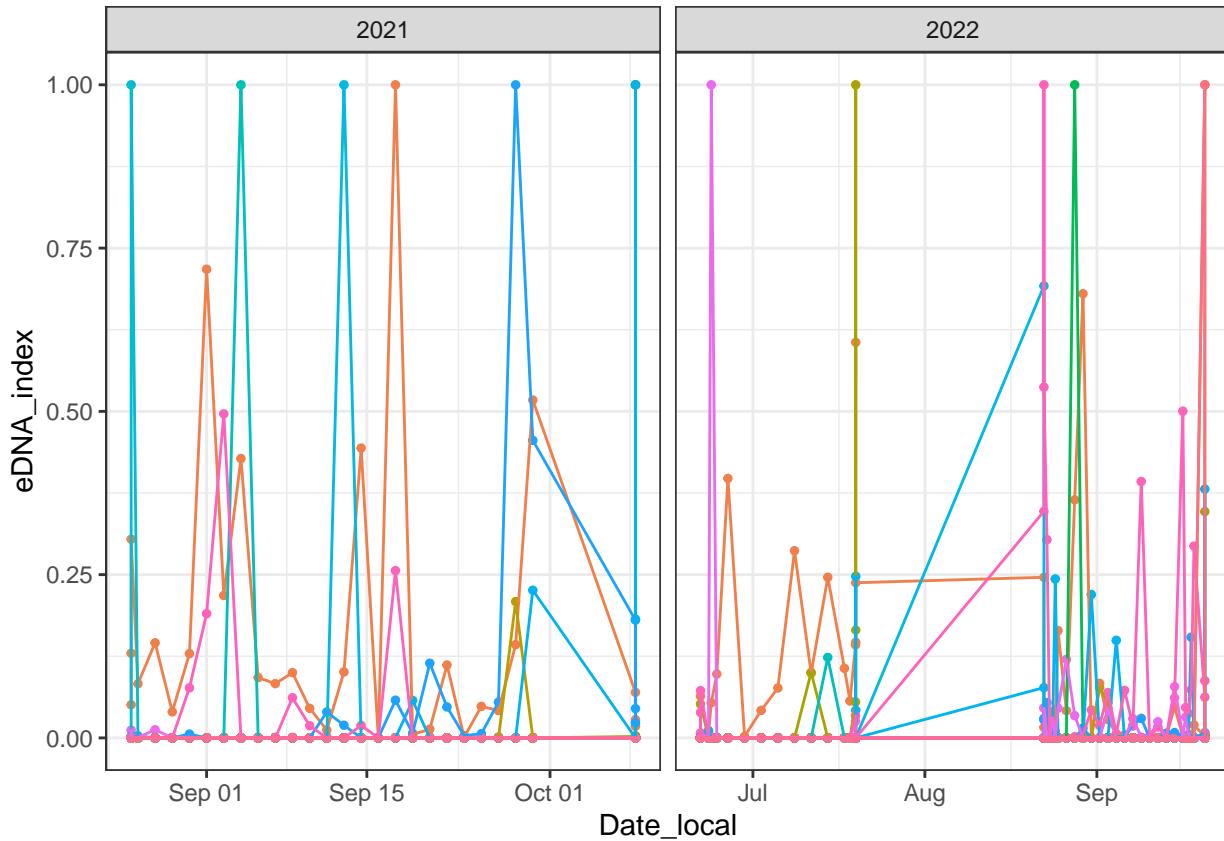
```



```

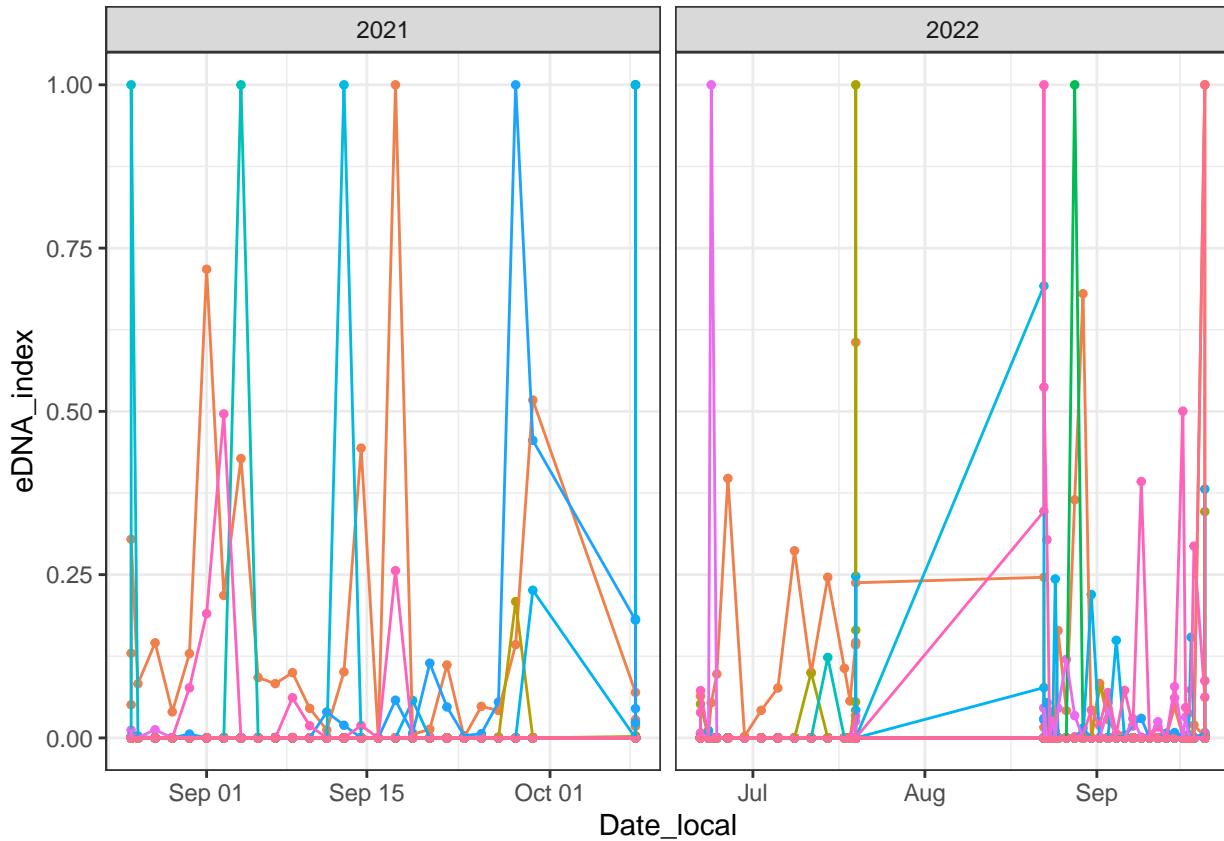
ggplot(eDNAindxEnvData, aes(x = Date_local, y = eDNA_index, color = Species)) +
  geom_point(show.legend = F, size = 1) +
  facet_wrap(facets = vars(year(Date_local)), scales = "free_x") +
  geom_line(show.legend = F) +
  theme_bw()

```



# Seems legit!

```
ggplot(eDNAindxEnvDataYr, aes(x = Date_local, y = eDNA_index, color = Species)) +
  geom_point(show.legend = F, size = 1) +
  facet_wrap(facets = vars(year(Date_local)), scales = "free_x") +
  geom_line(show.legend = F) +
  theme_bw()
```



```

source(here("eDNA_Index_Hypoxia", "eDNA_EnvGraphFunction.R"))

## ggbreak v0.1.2
##
## If you use ggbreak in published research, please cite the following
## paper:
## S Xu, M Chen, T Feng, L Zhan, L Zhou, G Yu. Use ggbreak to effectively
## utilize plotting space to deal with large datasets and outliers.
## Frontiers in Genetics. 2021, 12:774846. doi: 10.3389/fgene.2021.774846
## Warning in geom_rect(data = sampleHighlight, inherit.aes = FALSE, mapping =
## aes(xmin = x1b, : Ignoring unknown parameters: `stroke`
## Warning in geom_rect(data = sampleHighlight, inherit.aes = FALSE, mapping =
## aes(xmin = x2b, : Ignoring unknown parameters: `stroke`
## Warning in geom_rect(data = sampleHighlight, inherit.aes = FALSE, mapping =
## aes(xmin = x3b, : Ignoring unknown parameters: `stroke`

eDNAGraph(eDNAindxEnvData_clean,
          envCond = "DO",
          envCondName = "Oxygen",
          filepath = here("eDNA_Index_Hypoxia", "Plots", "eDNAXDO"),
          ylab = "Dissolved Oxygen (mg/L)",
          widthpx = 3000, # make it longer
          threshold = F,
          thresholdLvl = 2
)

```

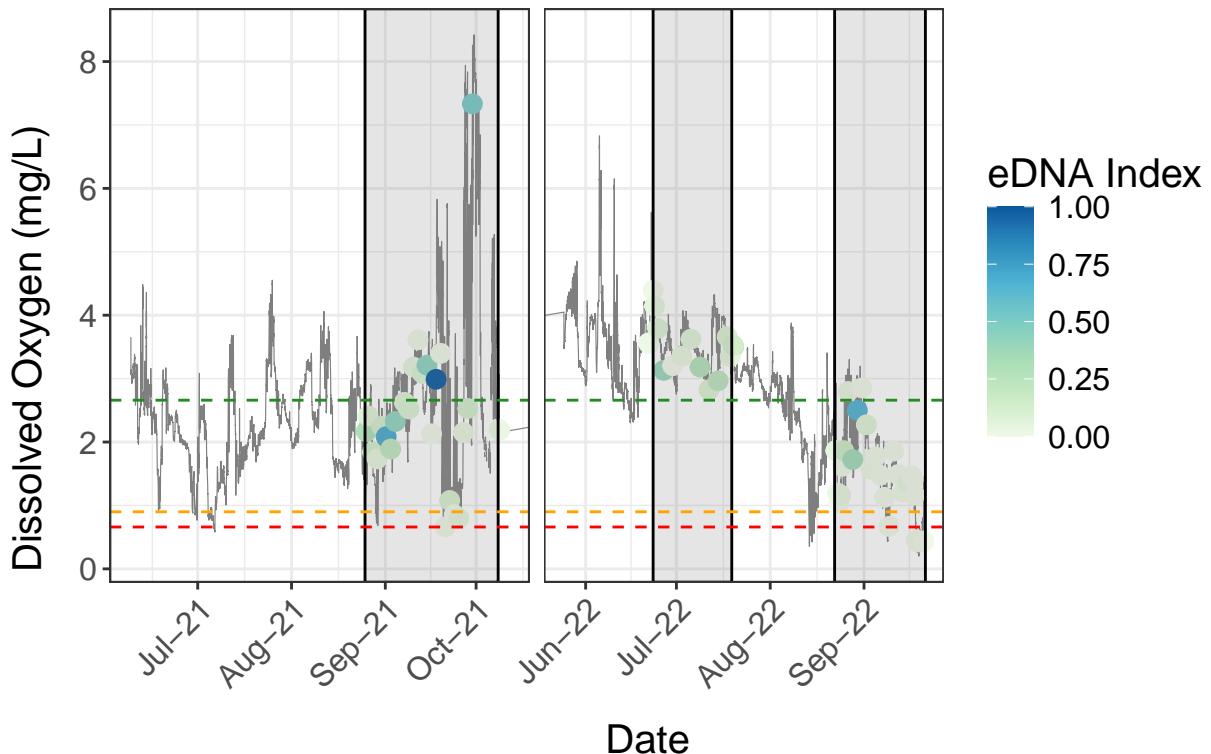
```

## [1] "HEADS UP: Date/time must be called exactly date and be in POSIXct, and envCond must be entered as a character vector"
## [1] "If you don't want a threshold line, set threshold = F instead of setting a thresholdLvl"
## [1] "Also for some reason you have to press 1 to confirm this function. Don't worry about it."
## [1] "Acartia longiremis"
## [1] "Acartia longiremis eDNA Index vs Oxygen"

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

## Acartia longiremis eDNA Index vs Oxygen

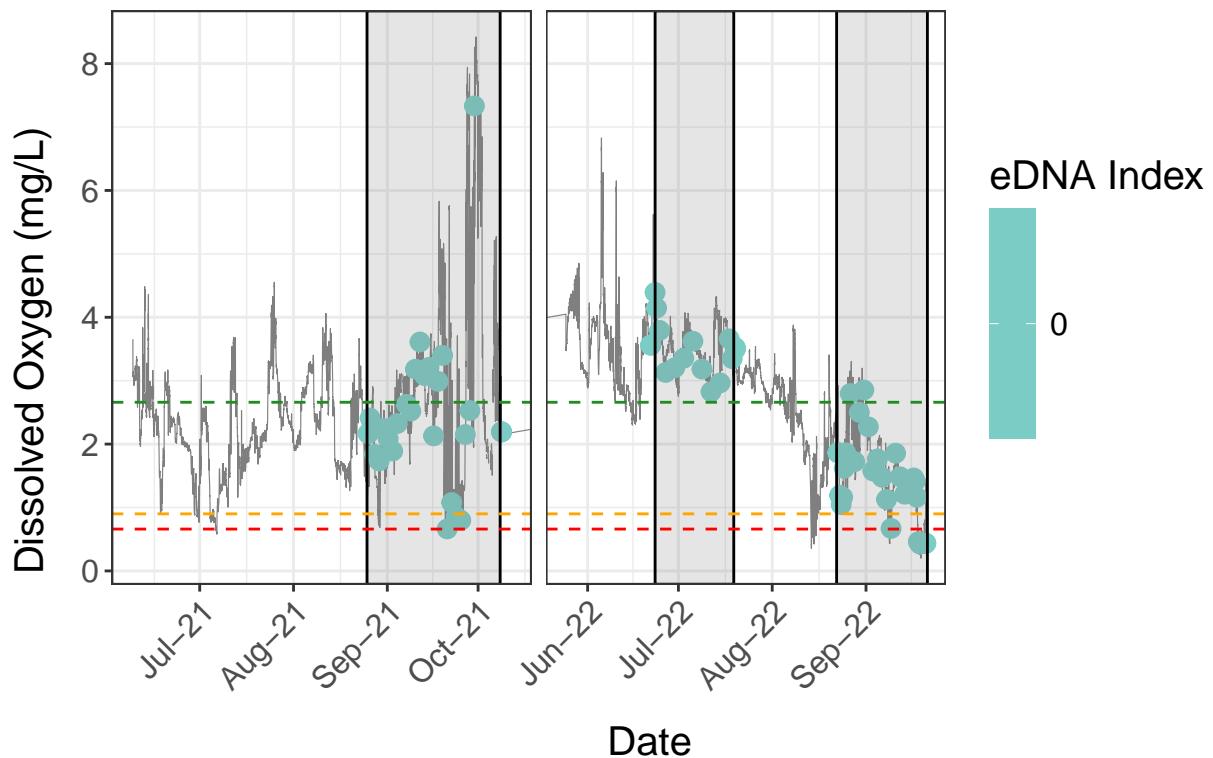


```

## [1] "Ameira sp. FHL 1"
## [1] "Ameira sp. FHL 1 eDNA Index vs Oxygen"

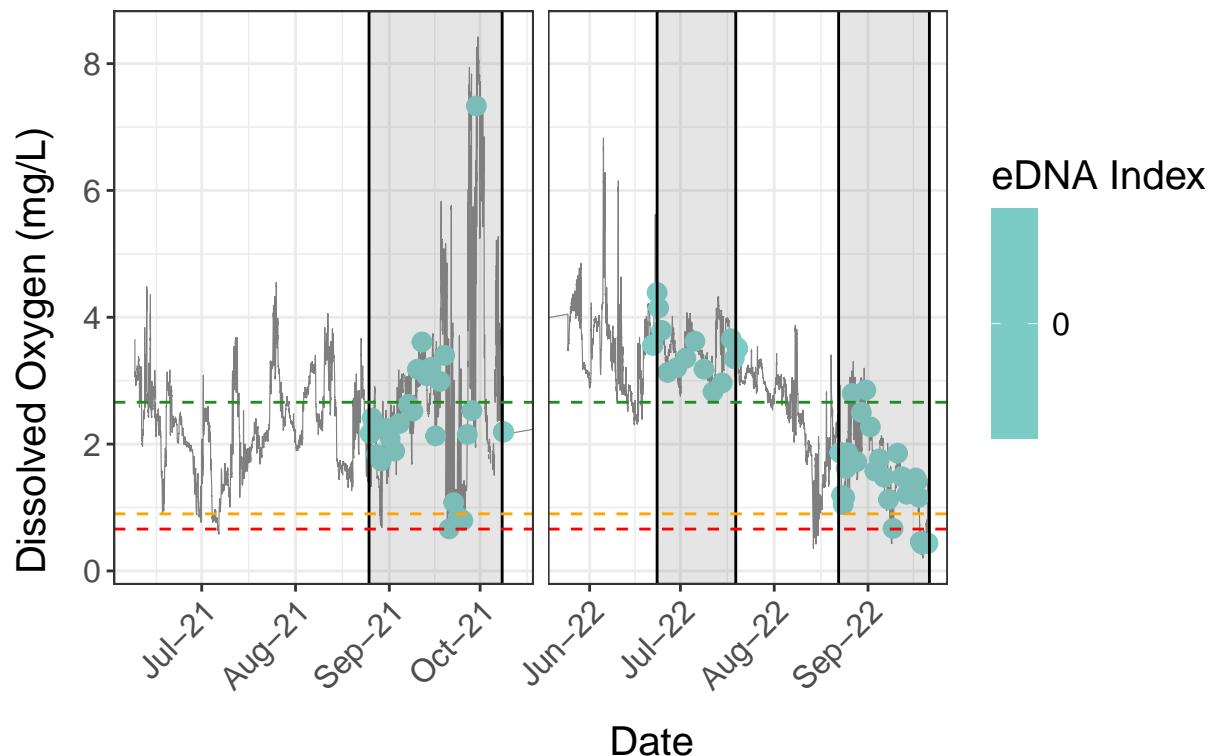
```

## Ameira sp. FHL 1 eDNA Index vs Oxygen



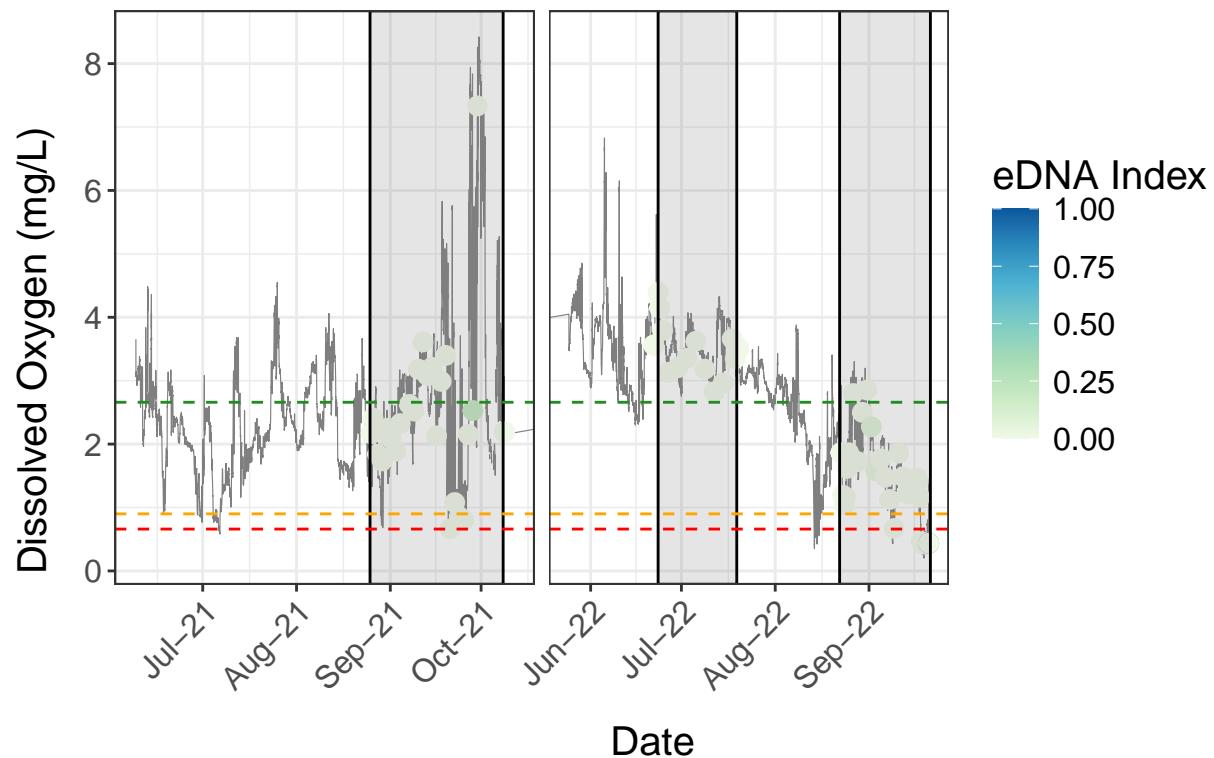
```
## [1] "Calanus marshallae"  
## [1] "Calanus marshallae eDNA Index vs Oxygen"
```

## Calanus marshallae eDNA Index vs Oxygen



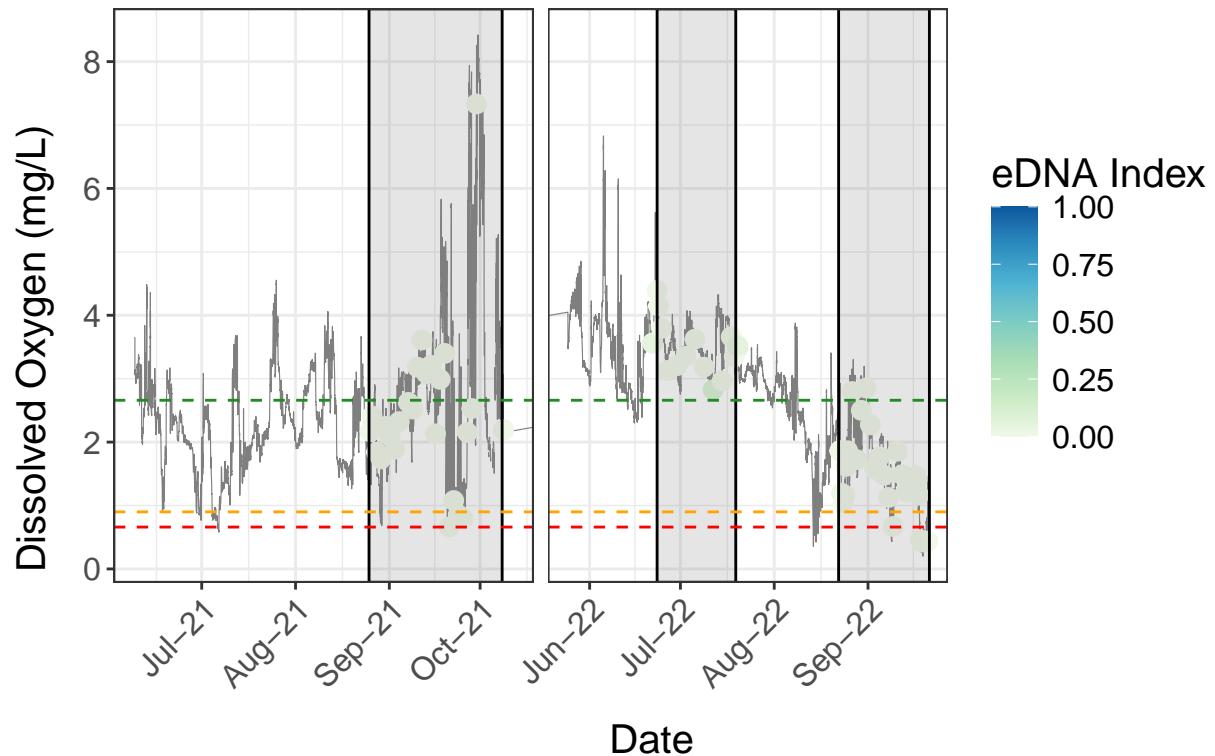
```
## [1] "Calanus pacificus"  
## [1] "Calanus pacificus eDNA Index vs Oxygen"
```

## Calanus pacificus eDNA Index vs Oxygen



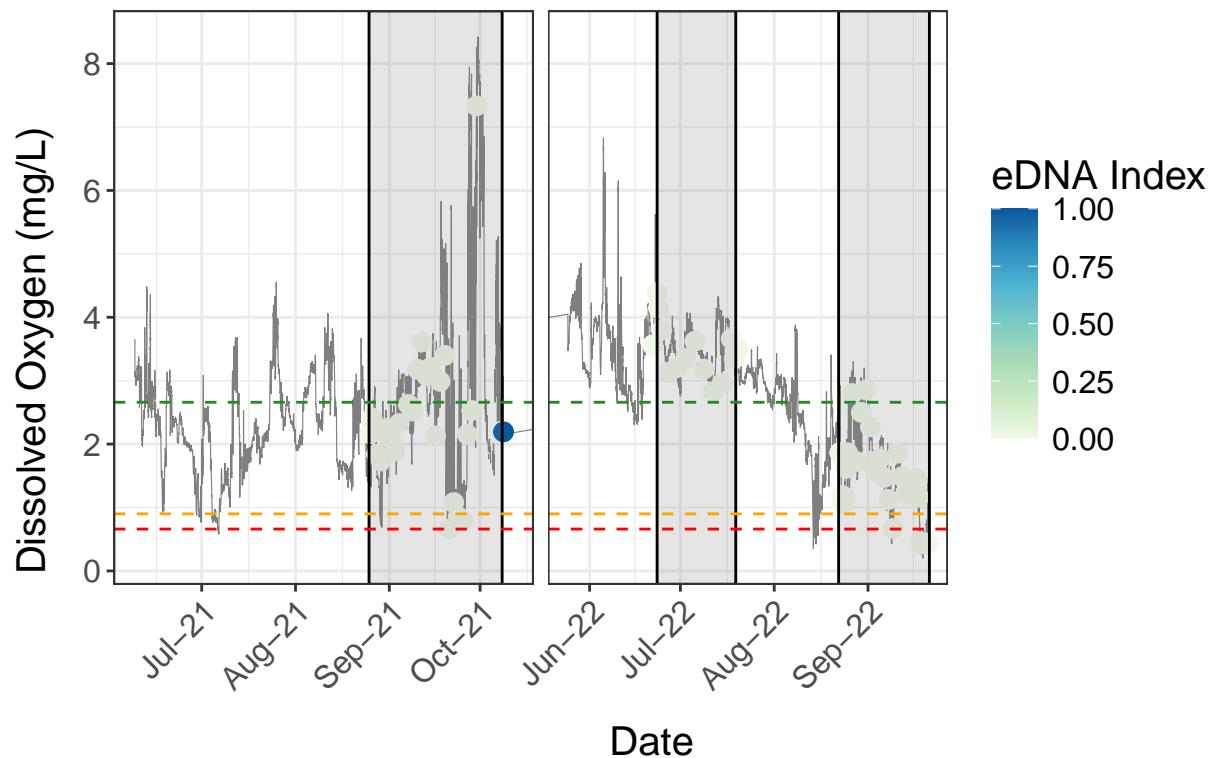
```
## [1] "Centropages abdominalis"  
## [1] "Centropages abdominalis eDNA Index vs Oxygen"
```

## Centropages abdominalis eDNA Index vs Oxygen



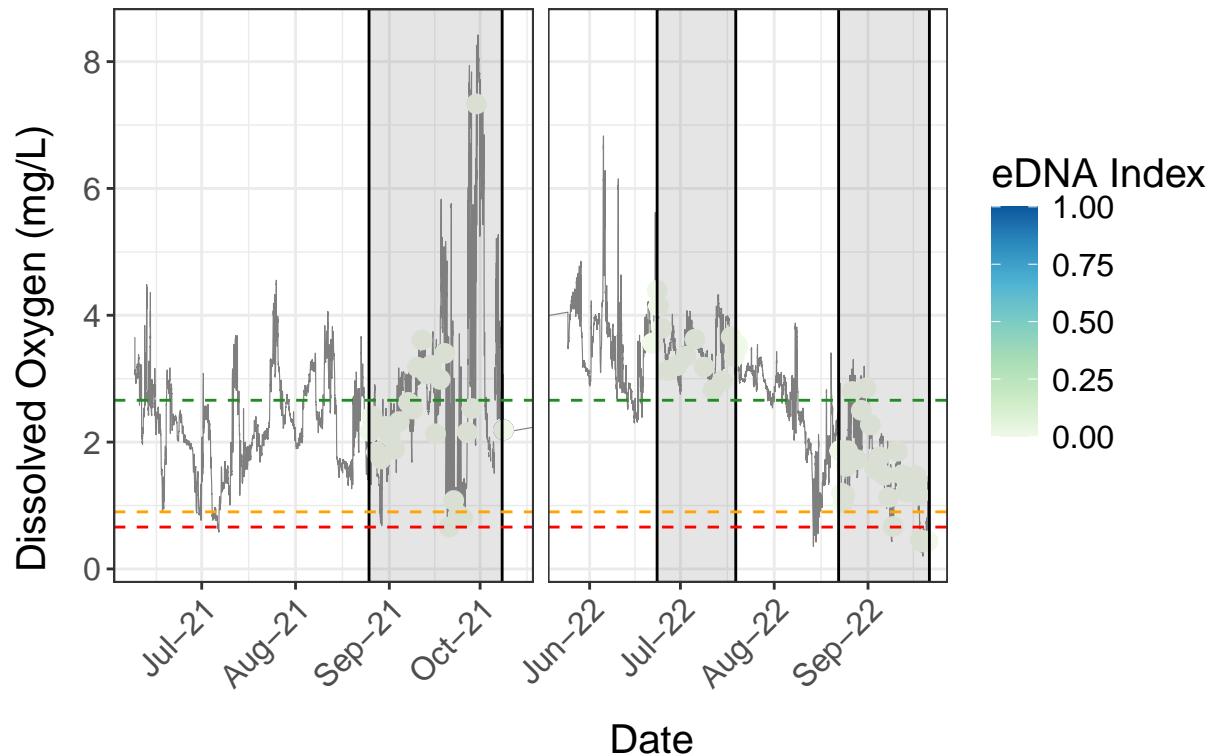
```
## [1] "Clausocalanus parapergens"  
## [1] "Clausocalanus parapergens eDNA Index vs Oxygen"
```

## *Clausocalanus parapergens* eDNA Index vs Oxygen



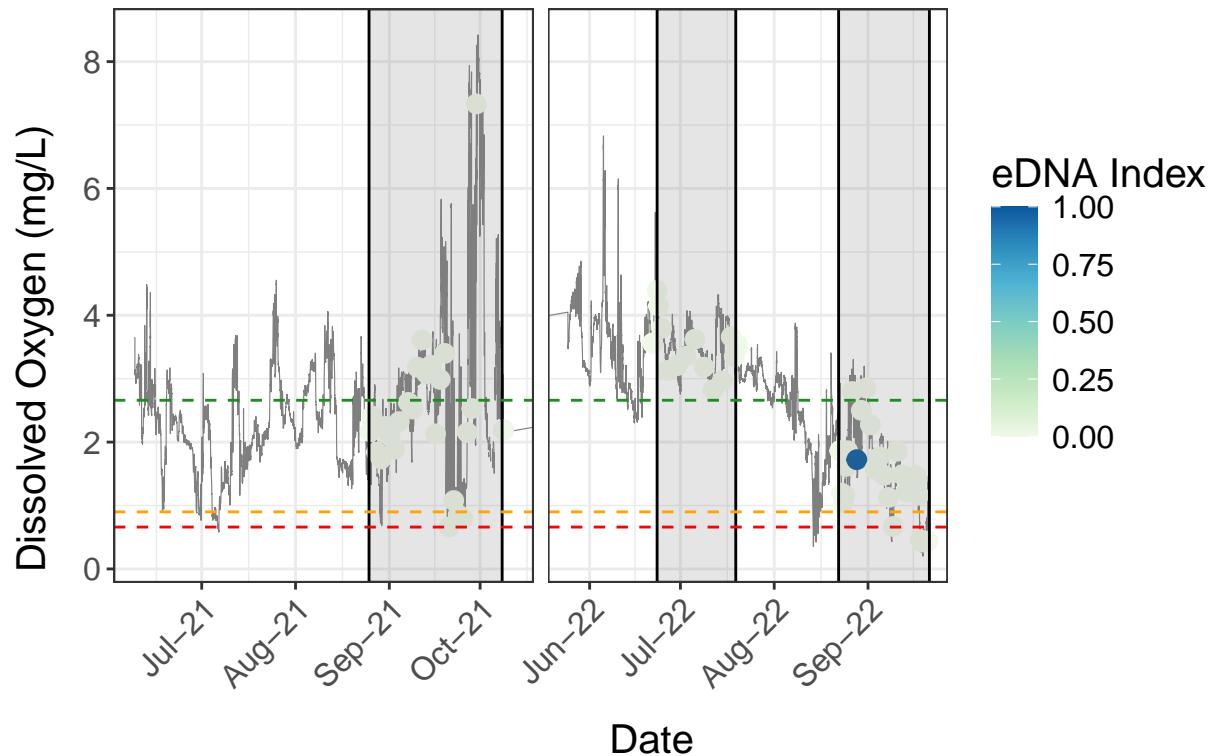
```
## [1] "Clausocalanus pergens"  
## [1] "Clausocalanus pergens eDNA Index vs Oxygen"
```

## *Clausocalanus pergens* eDNA Index vs Oxygen



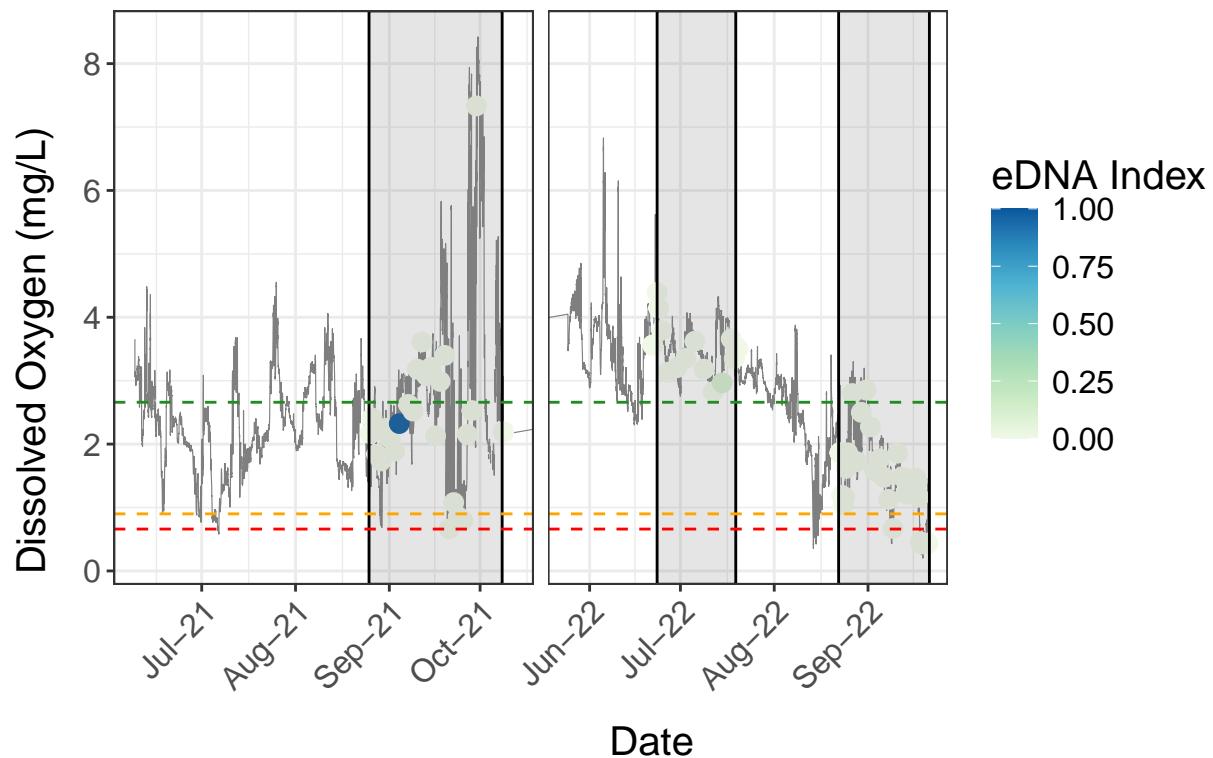
```
## [1] "Diacyclops incolotaenia"  
## [1] "Diacyclops incolotaenia eDNA Index vs Oxygen"
```

## Diacyclops incolaenia eDNA Index vs Oxygen



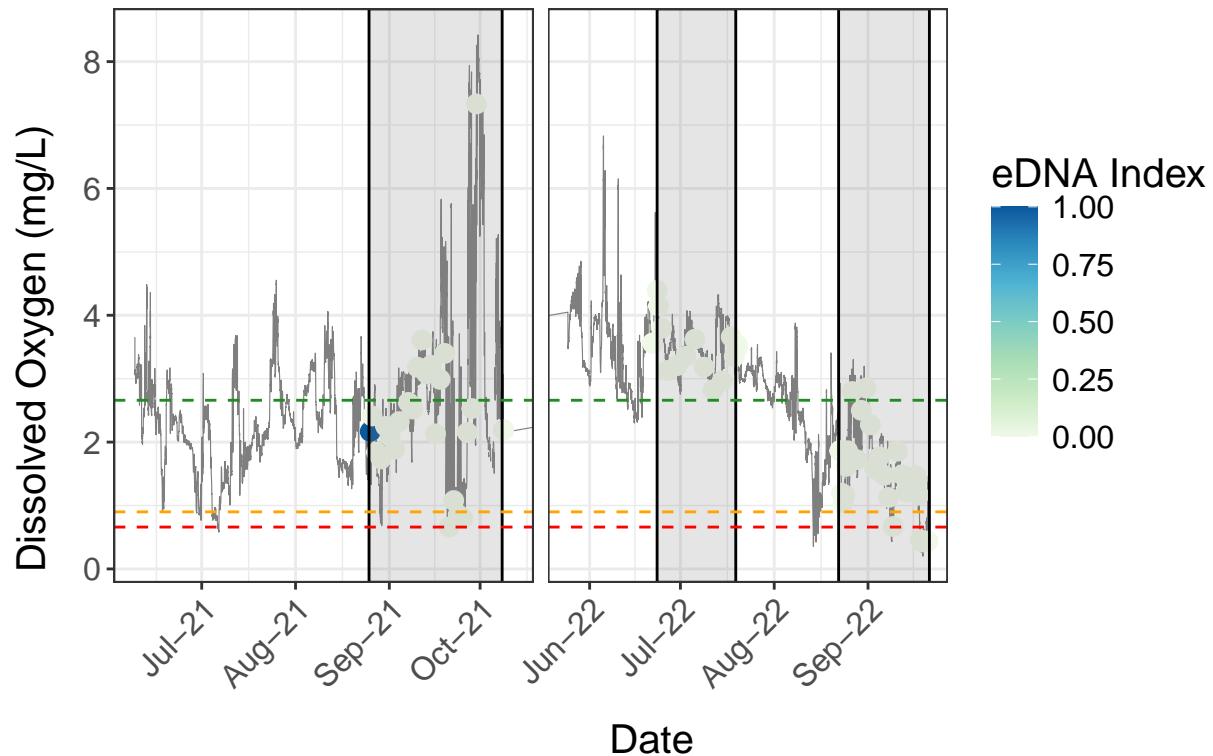
```
## [1] "Lucicutia flavigornis"  
## [1] "Lucicutia flavigornis eDNA Index vs Oxygen"
```

## *Lucicutia flavigornis* eDNA Index vs Oxygen



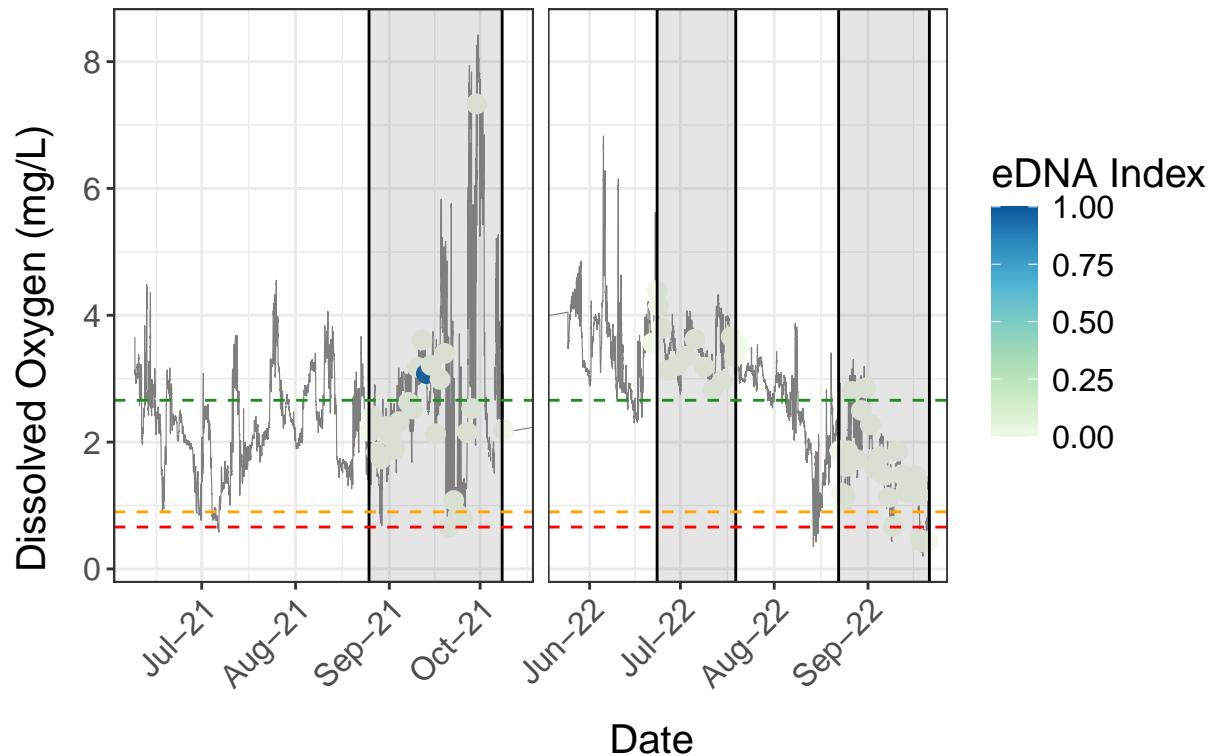
```
## [1] "Metridia lucens"  
## [1] "Metridia lucens eDNA Index vs Oxygen"
```

## Metridia lucens eDNA Index vs Oxygen



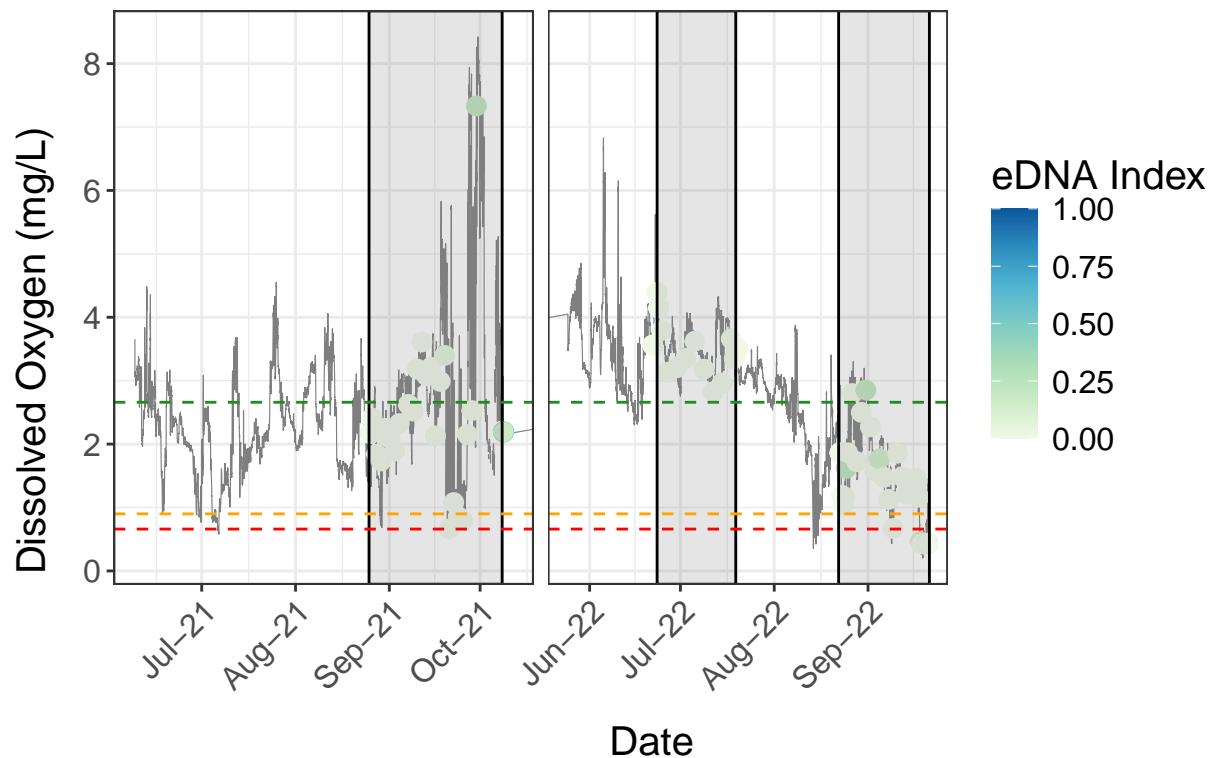
```
## [1] "Metridia pacifica"  
## [1] "Metridia pacifica eDNA Index vs Oxygen"
```

## Metridia pacifica eDNA Index vs Oxygen



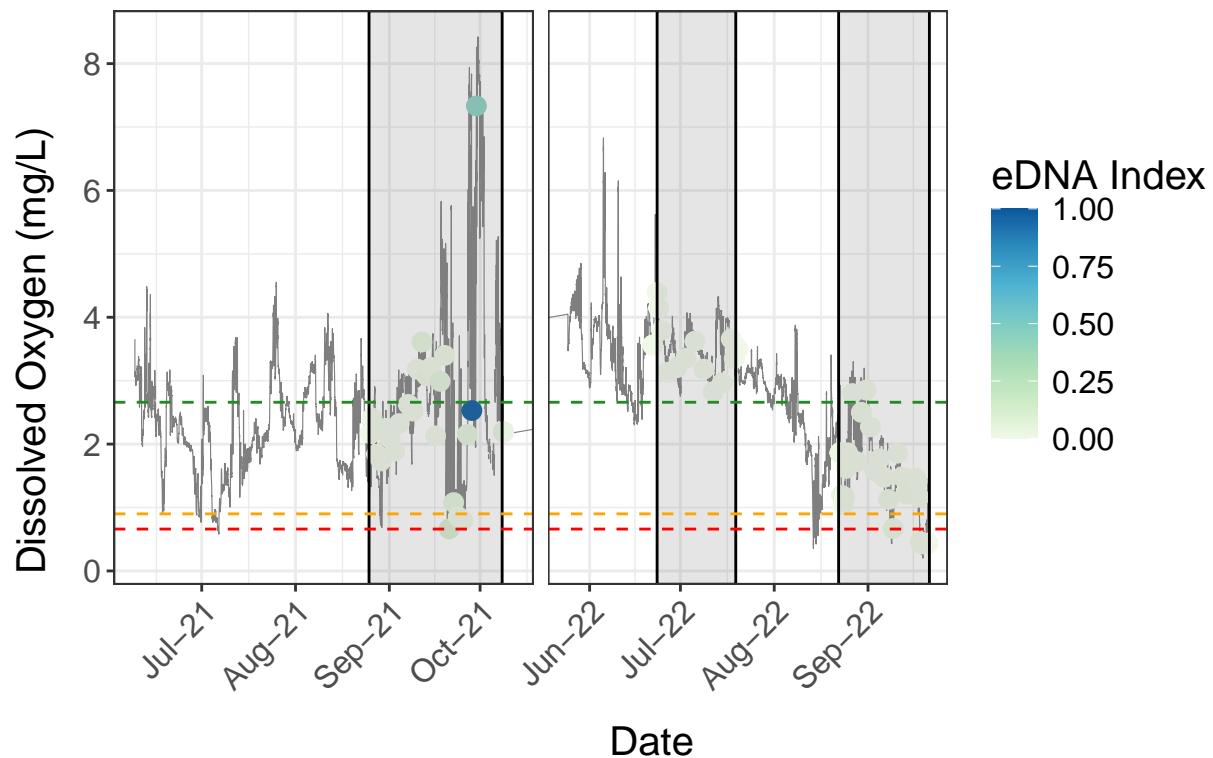
```
## [1] "Oithona similis"  
## [1] "Oithona similis eDNA Index vs Oxygen"
```

## Oithona similis eDNA Index vs Oxygen



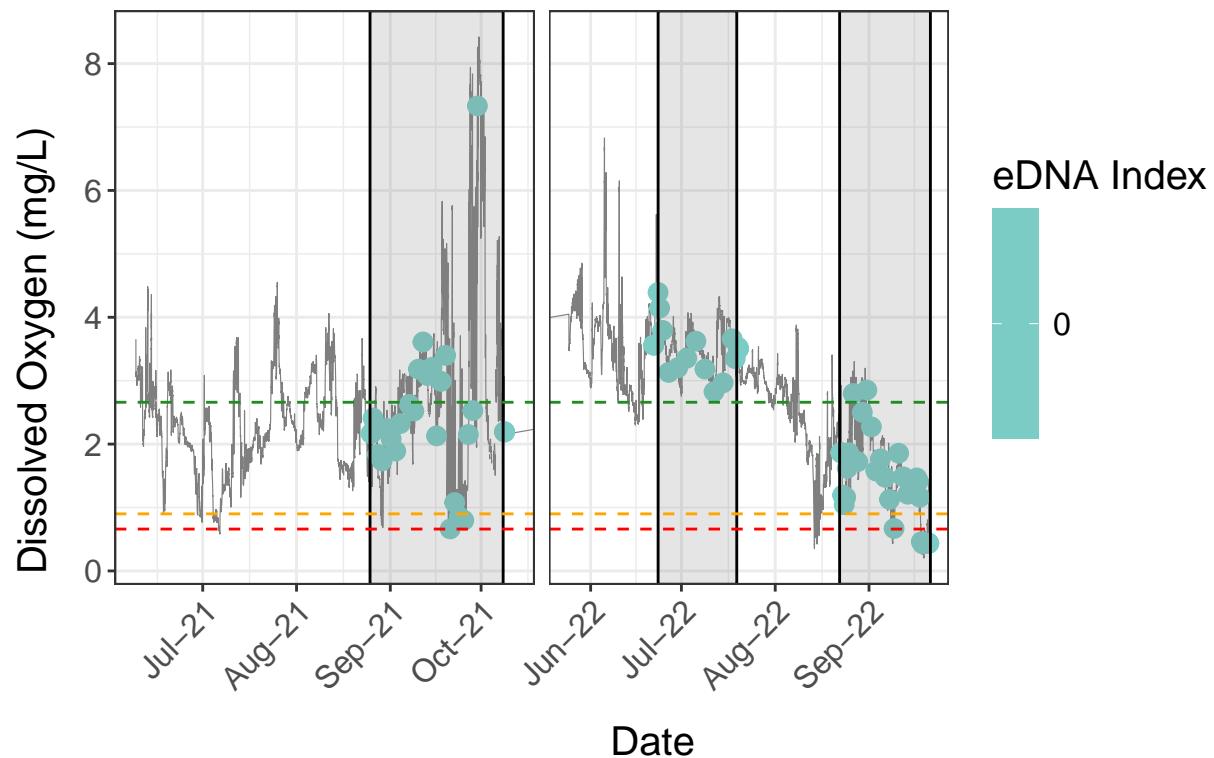
```
## [1] "Paracalanus sp. C AC-2013"
## [1] "Paracalanus sp. C AC-2013 eDNA Index vs Oxygen"
```

## Paracalanus sp. C AC-2013 eDNA Index vs Oxygen



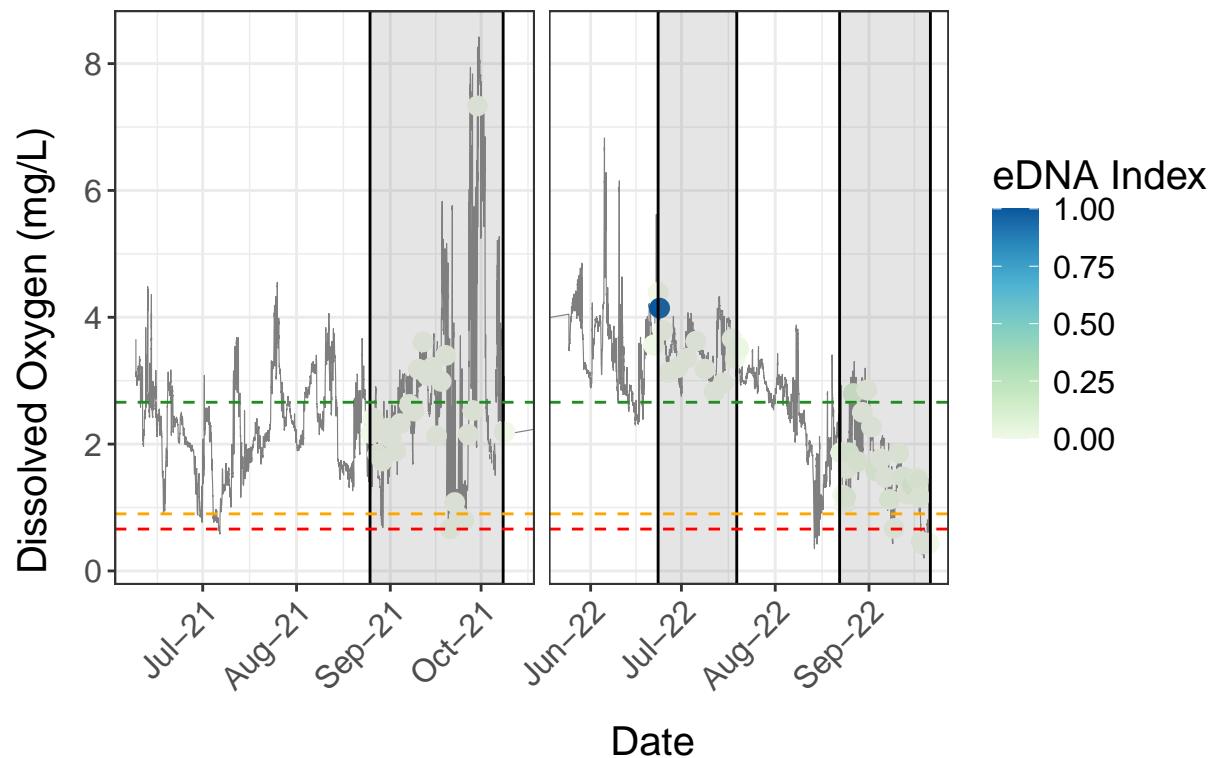
```
## [1] "Pseudobradya minor"  
## [1] "Pseudobradya minor eDNA Index vs Oxygen"
```

## Pseudobradya minor eDNA Index vs Oxygen



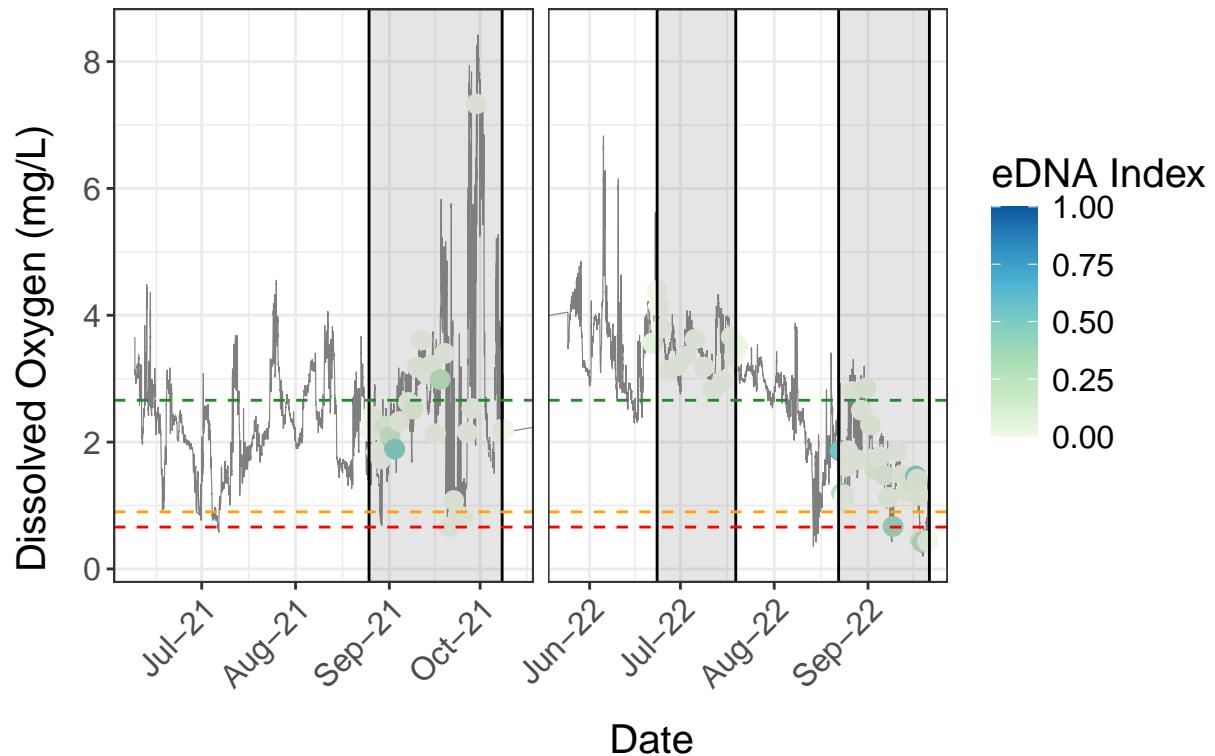
```
## [1] "Pseudocalanus newmani"  
## [1] "Pseudocalanus newmani eDNA Index vs Oxygen"
```

## Pseudocalanus newmani eDNA Index vs Oxygen



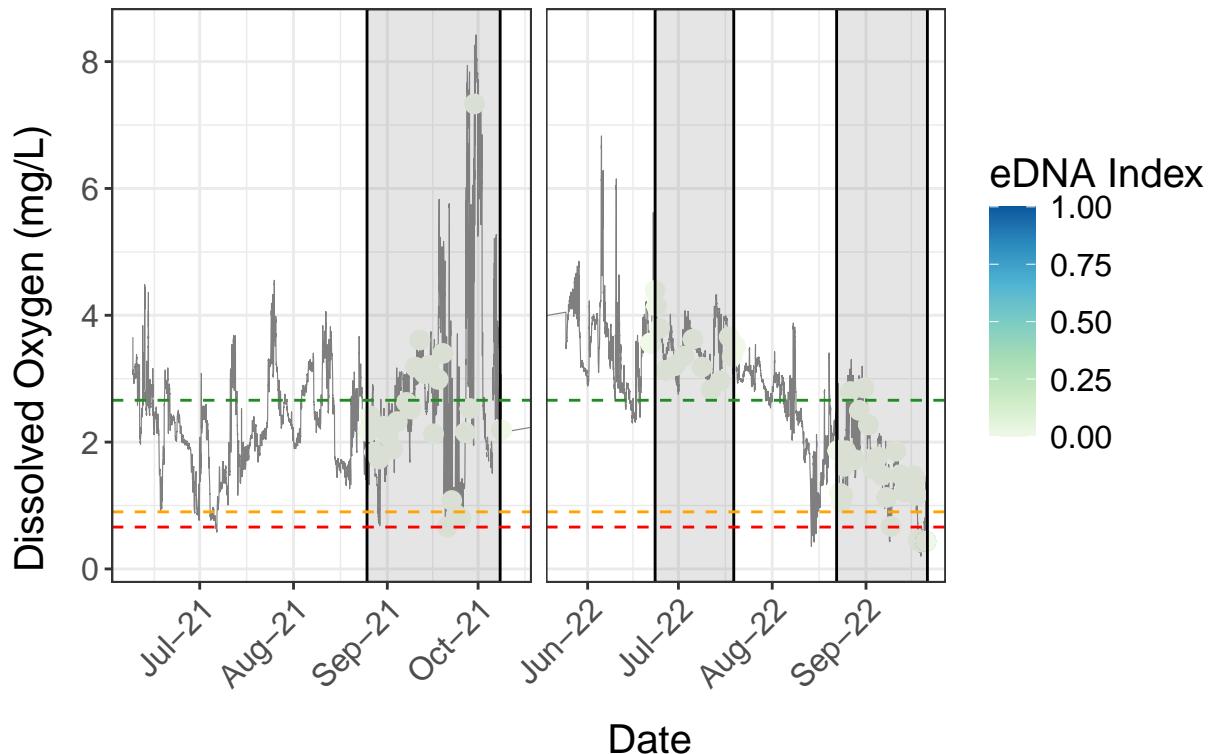
```
## [1] "Thermocyclops inversus"
## [1] "Thermocyclops inversus eDNA Index vs Oxygen"
```

## Thermocyclops inversus eDNA Index vs Oxygen



```
## [1] "Triconia minuta"  
## [1] "Triconia minuta eDNA Index vs Oxygen"
```

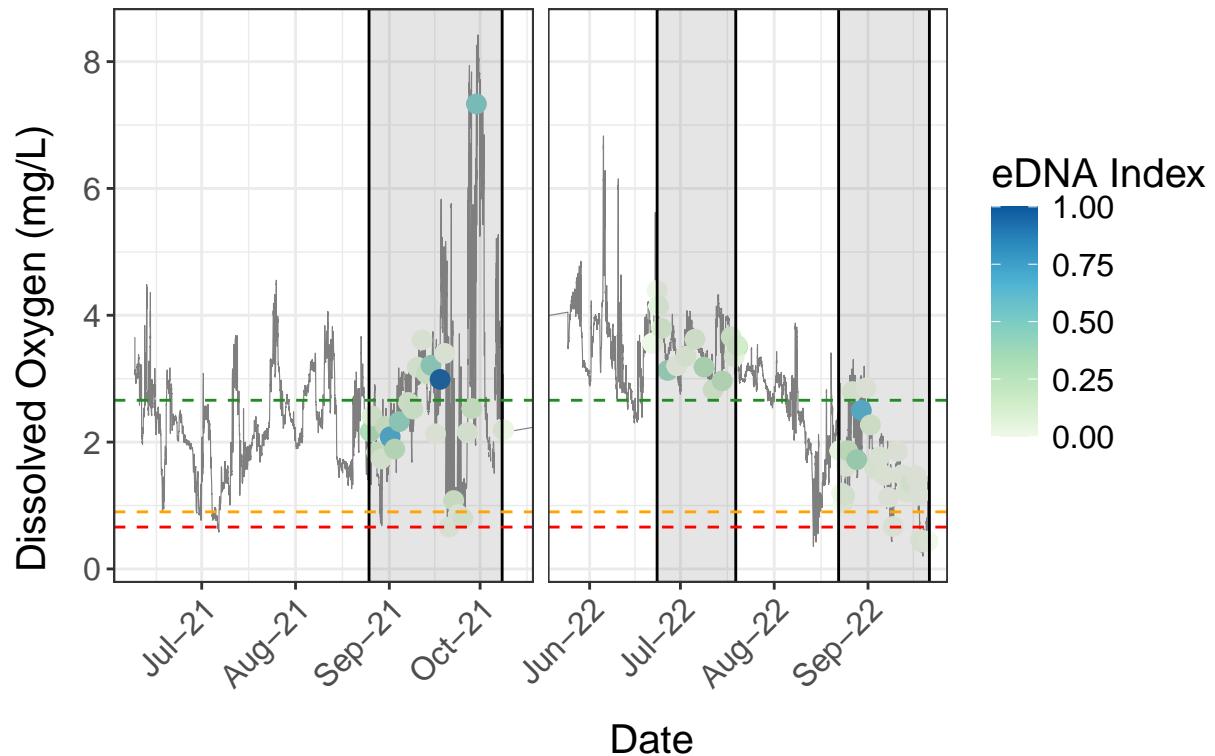
## Triconia minuta eDNA Index vs Oxygen



```
eDNAGraph(eDNAindxEnvData_cleanYr, # ignoring 2023
            envCond = "DO",
            envCondName = "Oxygen",
            filepath = here("eDNA_Index_Hypoxia", "Plots", "eDNAXDO_no23"),
            ylab = "Dissolved Oxygen (mg/L)",
            widthpx = 3000, # make it longer
            threshold = F,
            thresholdLvl = 2
            )

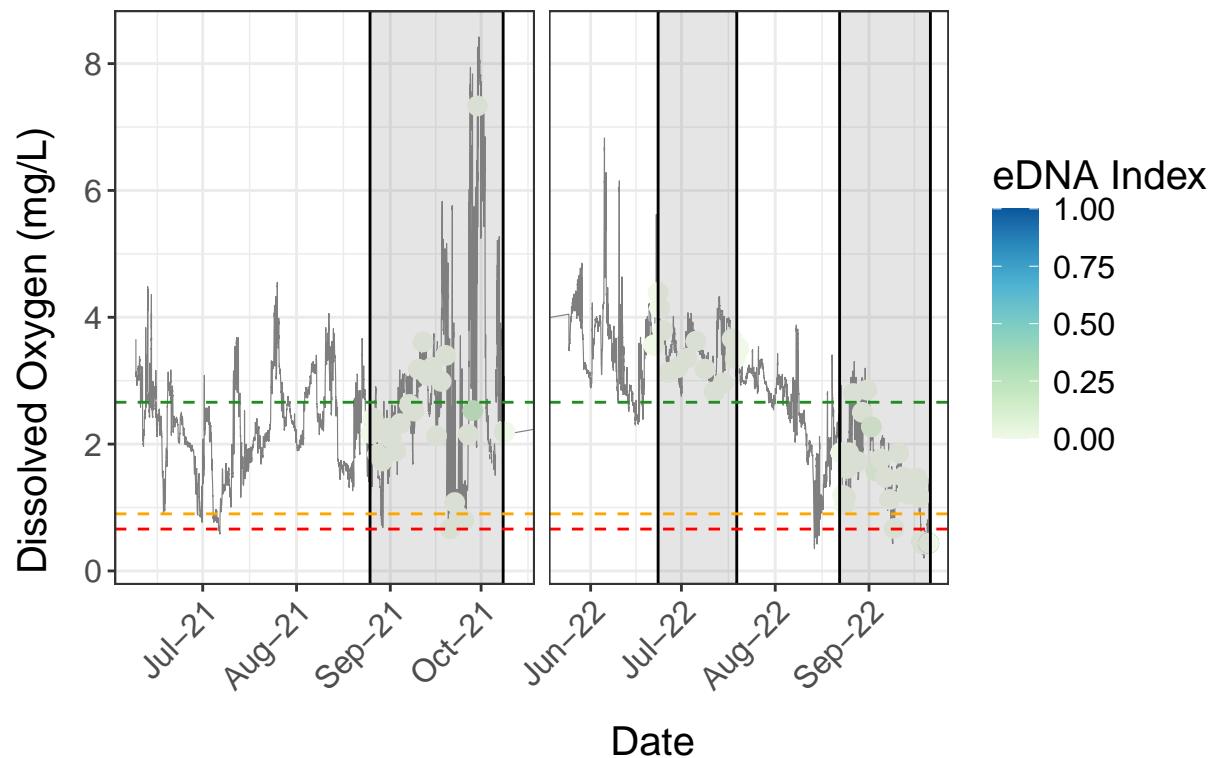
## [1] "HEADS UP: Date/time must be called exactly date and be in POSIXct, and envCond must be entered as a string"
## [1] "If you don't want a threshold line, set threshold = F instead of setting a thresholdLvl"
## [1] "Also for some reason you have to press 1 to confirm this function. Don't worry about it."
## [1] "Acartia longiremis"
## [1] "Acartia longiremis eDNA Index vs Oxygen"
```

## Acartia longiremis eDNA Index vs Oxygen



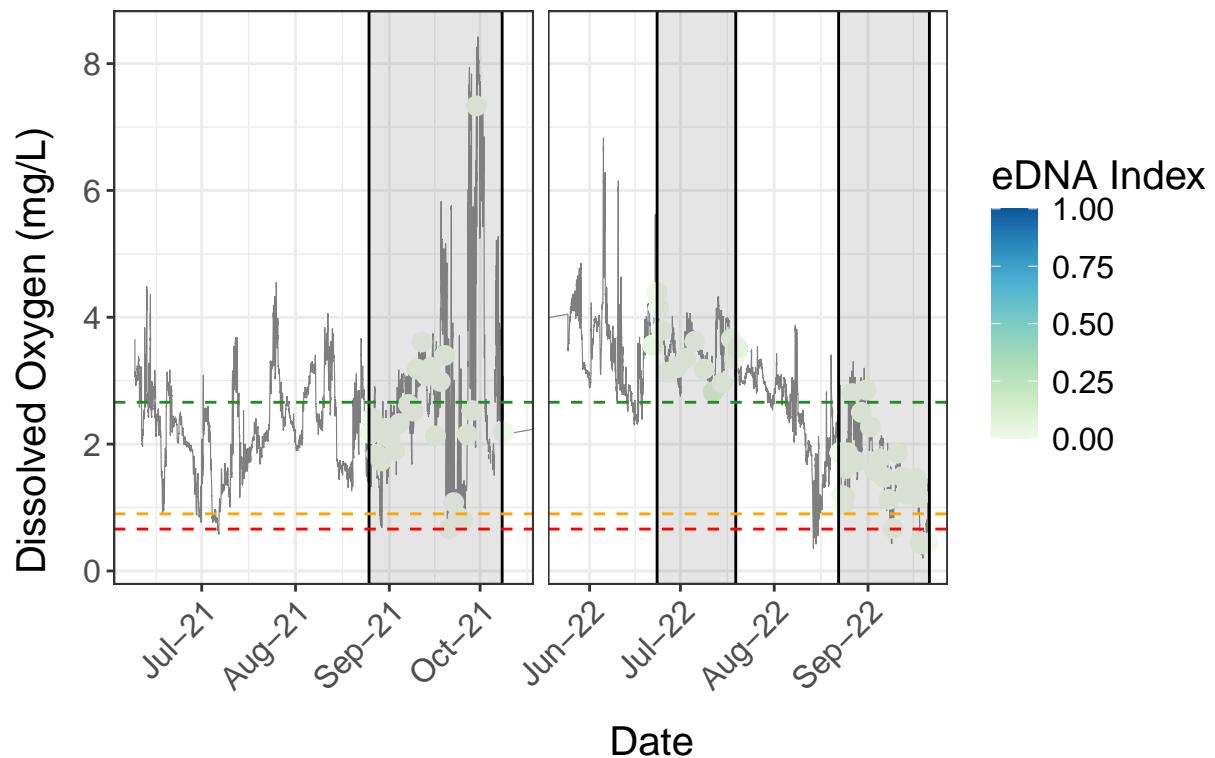
```
## [1] "Calanus pacificus"  
## [1] "Calanus pacificus eDNA Index vs Oxygen"
```

## Calanus pacificus eDNA Index vs Oxygen



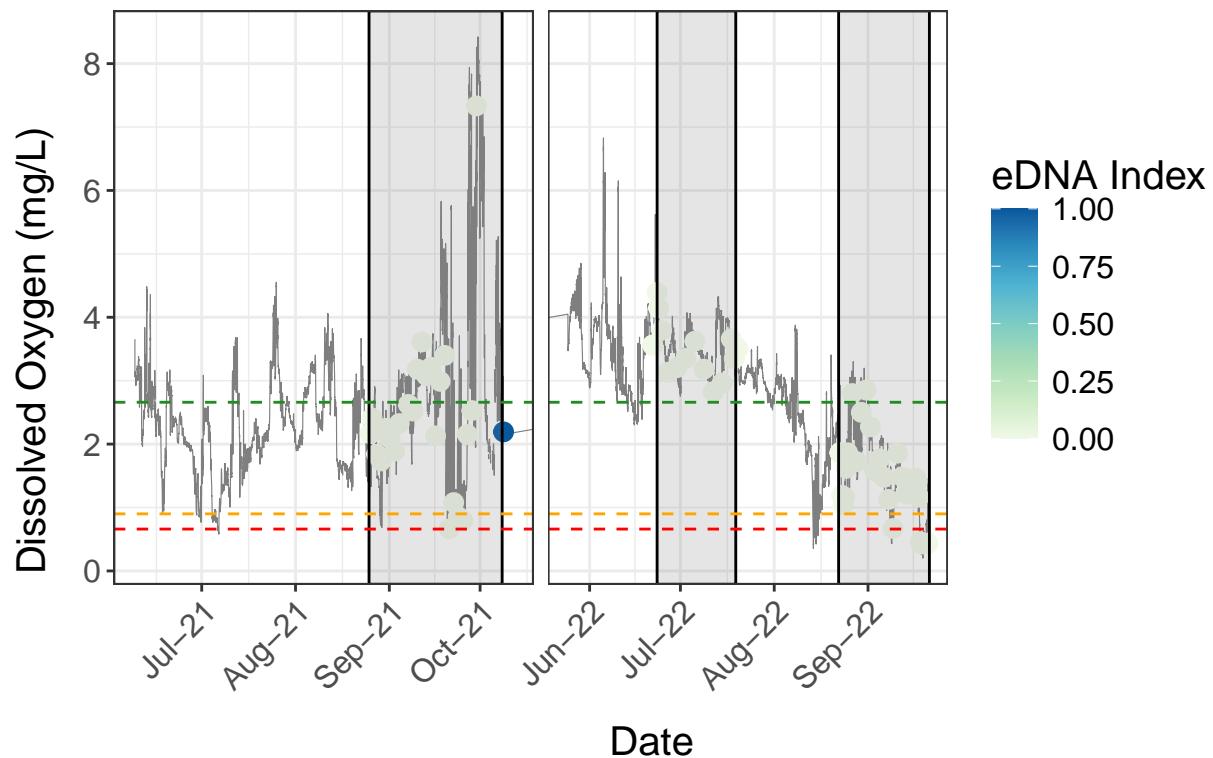
```
## [1] "Centropages abdominalis"  
## [1] "Centropages abdominalis eDNA Index vs Oxygen"
```

## Centropages abdominalis eDNA Index vs Oxygen



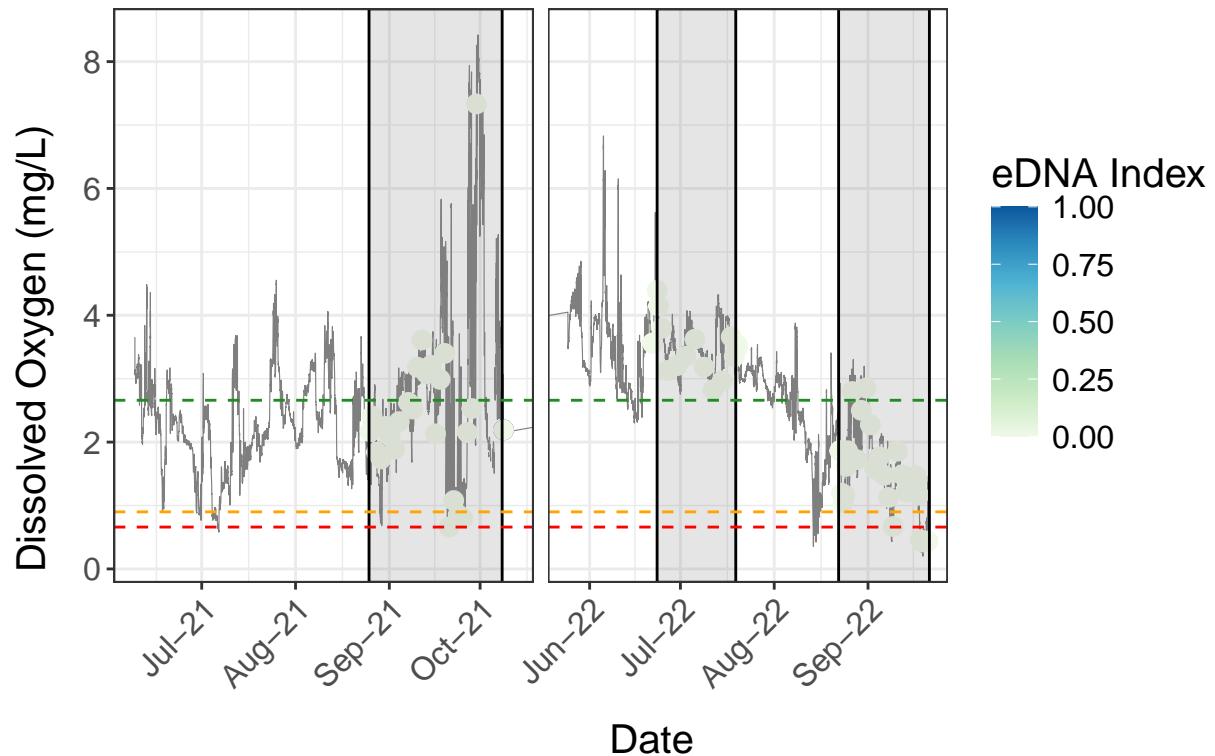
```
## [1] "Clausocalanus parapergens"  
## [1] "Clausocalanus parapergens eDNA Index vs Oxygen"
```

## *Clausocalanus parapergens* eDNA Index vs Oxygen



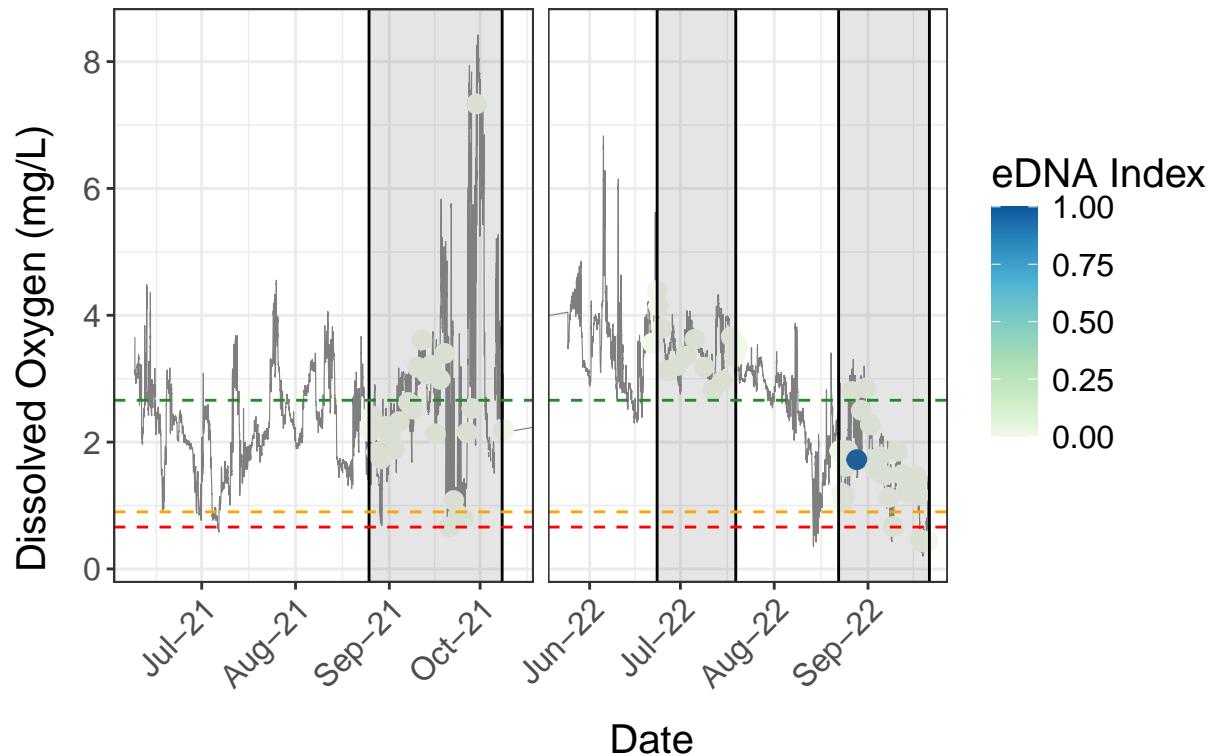
```
## [1] "Clausocalanus pergens"  
## [1] "Clausocalanus pergens eDNA Index vs Oxygen"
```

## *Clausocalanus pergens* eDNA Index vs Oxygen



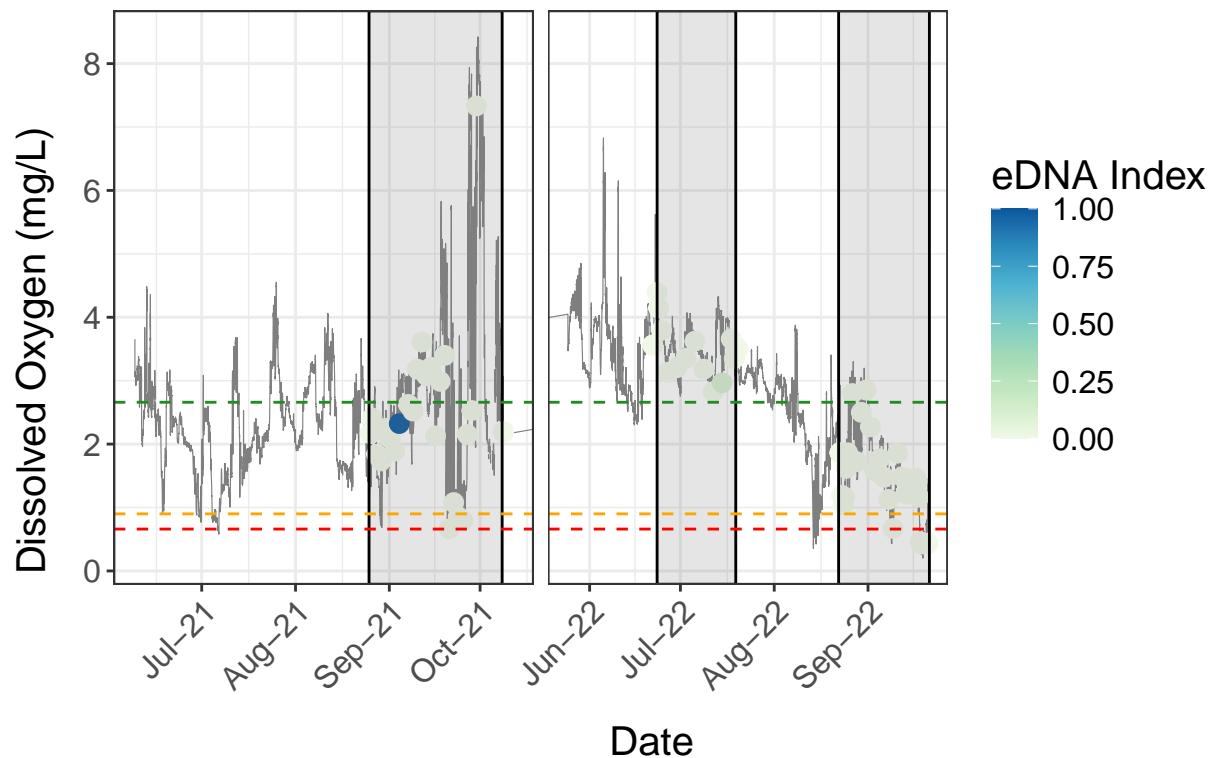
```
## [1] "Diacyclops incolotaenia"  
## [1] "Diacyclops incolotaenia eDNA Index vs Oxygen"
```

## Diacyclops incolaenia eDNA Index vs Oxygen



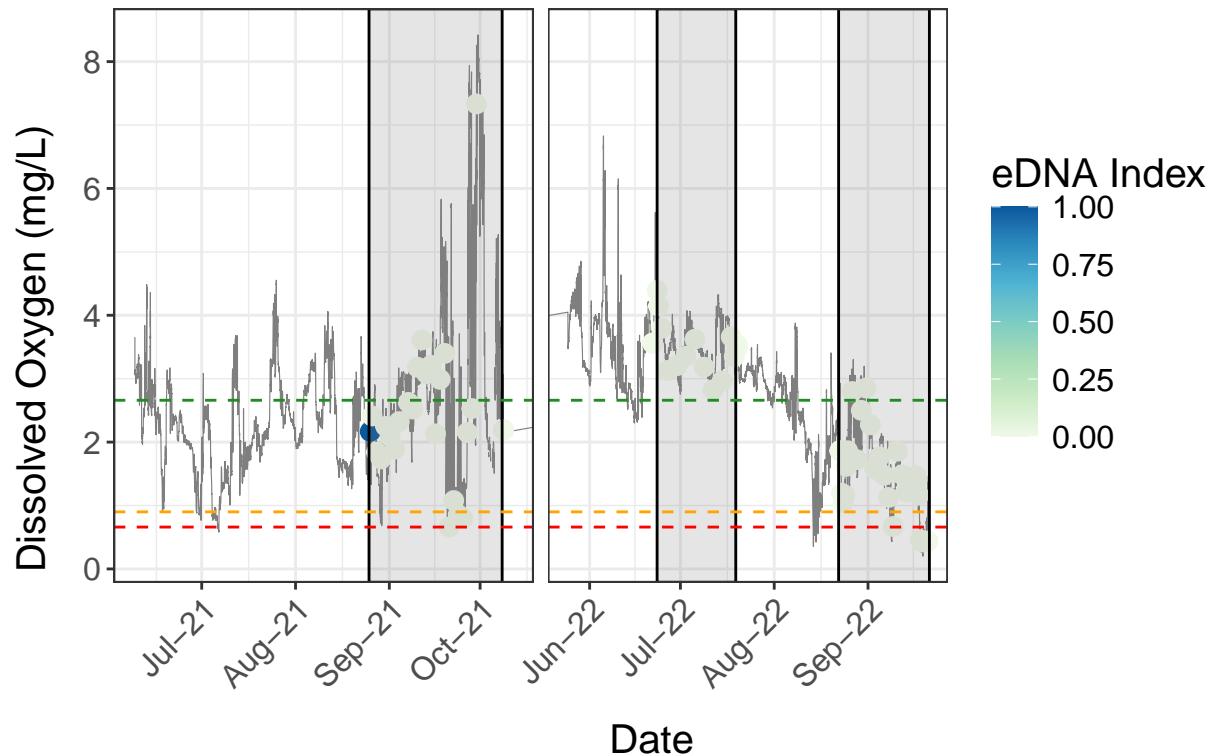
```
## [1] "Lucicutia flavigornis"  
## [1] "Lucicutia flavigornis eDNA Index vs Oxygen"
```

## *Lucicutia flavigornis* eDNA Index vs Oxygen



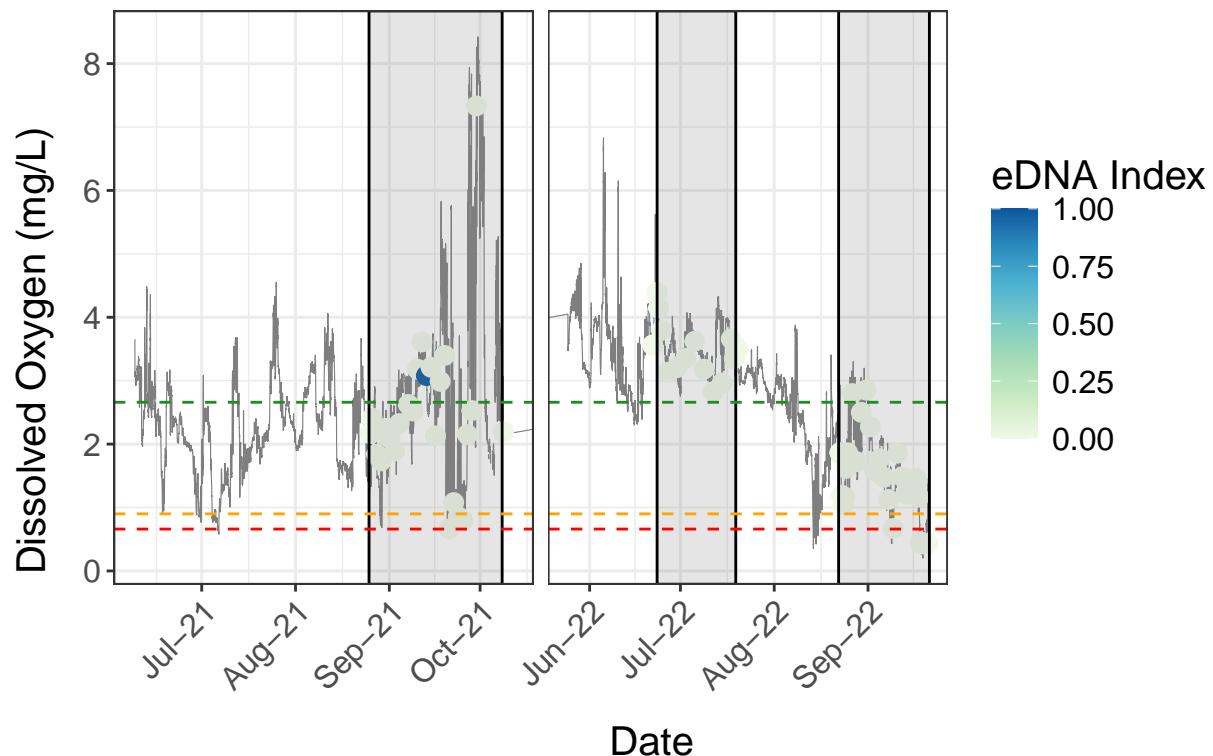
```
## [1] "Metridia lucens"  
## [1] "Metridia lucens eDNA Index vs Oxygen"
```

## Metridia lucens eDNA Index vs Oxygen



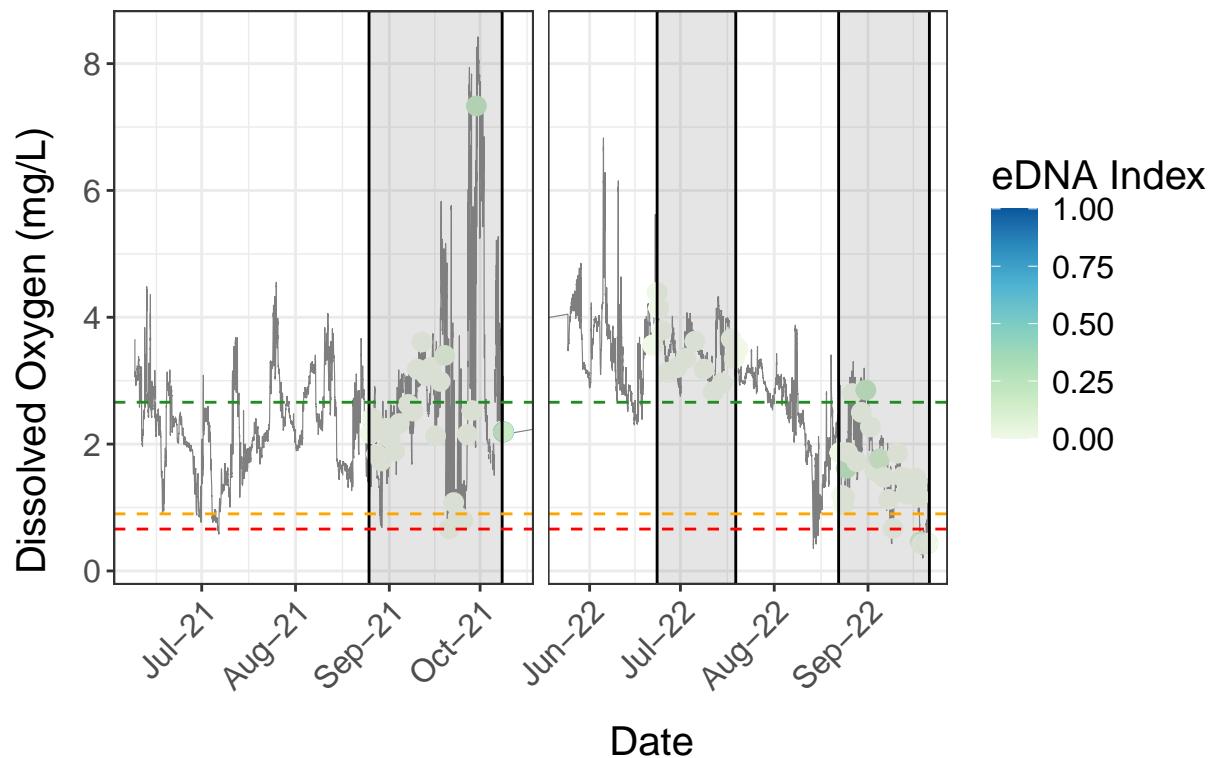
```
## [1] "Metridia pacifica"  
## [1] "Metridia pacifica eDNA Index vs Oxygen"
```

## Metridia pacifica eDNA Index vs Oxygen



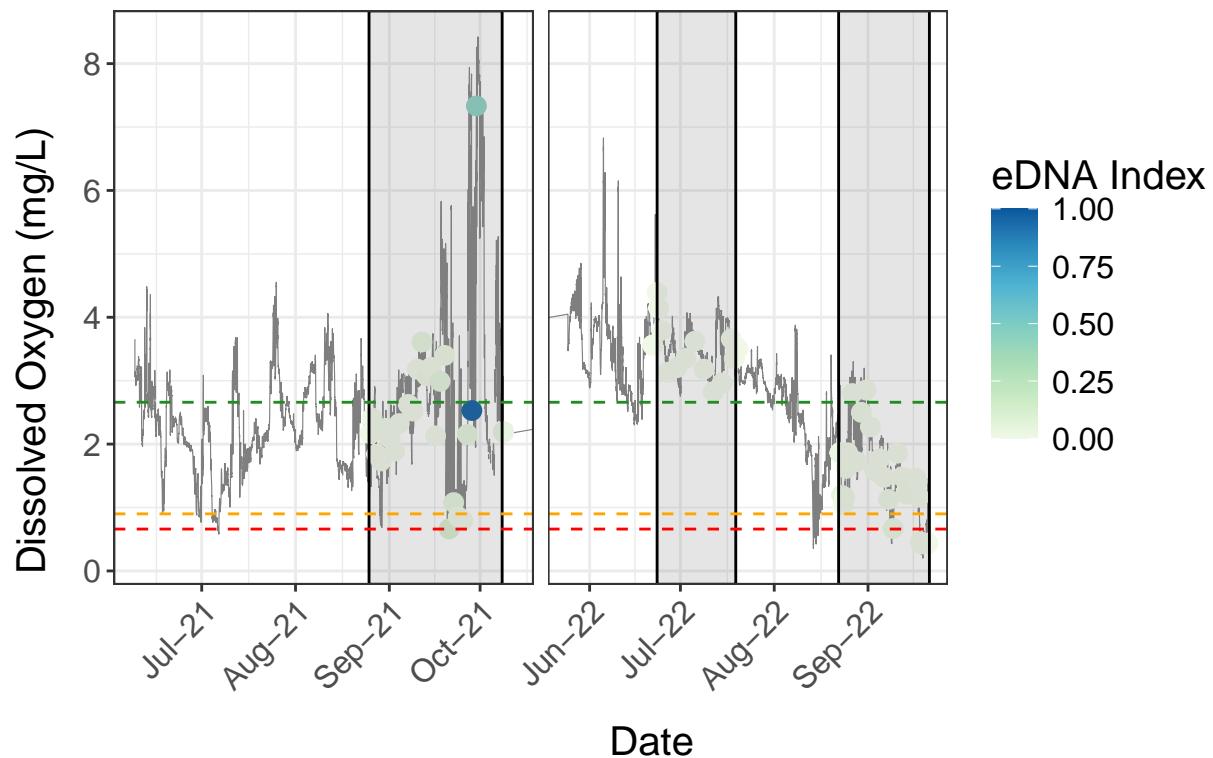
```
## [1] "Oithona similis"  
## [1] "Oithona similis eDNA Index vs Oxygen"
```

## Oithona similis eDNA Index vs Oxygen



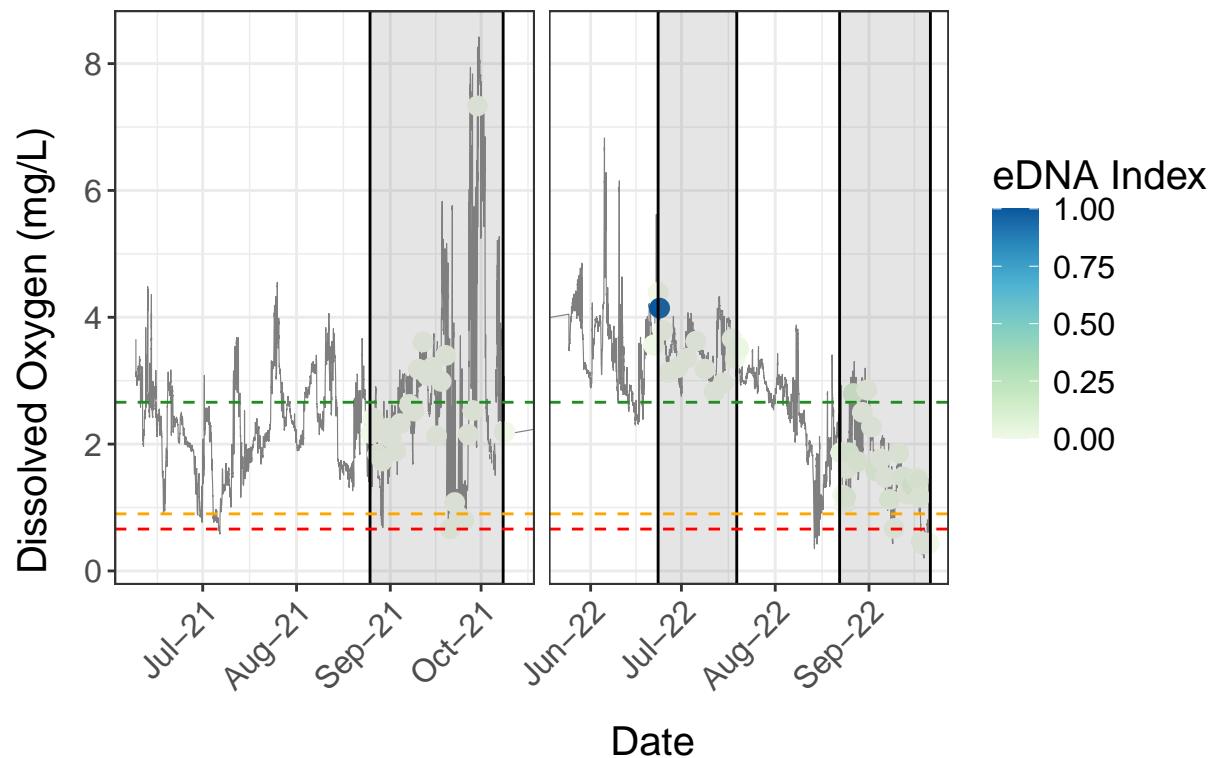
```
## [1] "Paracalanus sp. C AC-2013"
## [1] "Paracalanus sp. C AC-2013 eDNA Index vs Oxygen"
```

## Paracalanus sp. C AC-2013 eDNA Index vs Oxygen



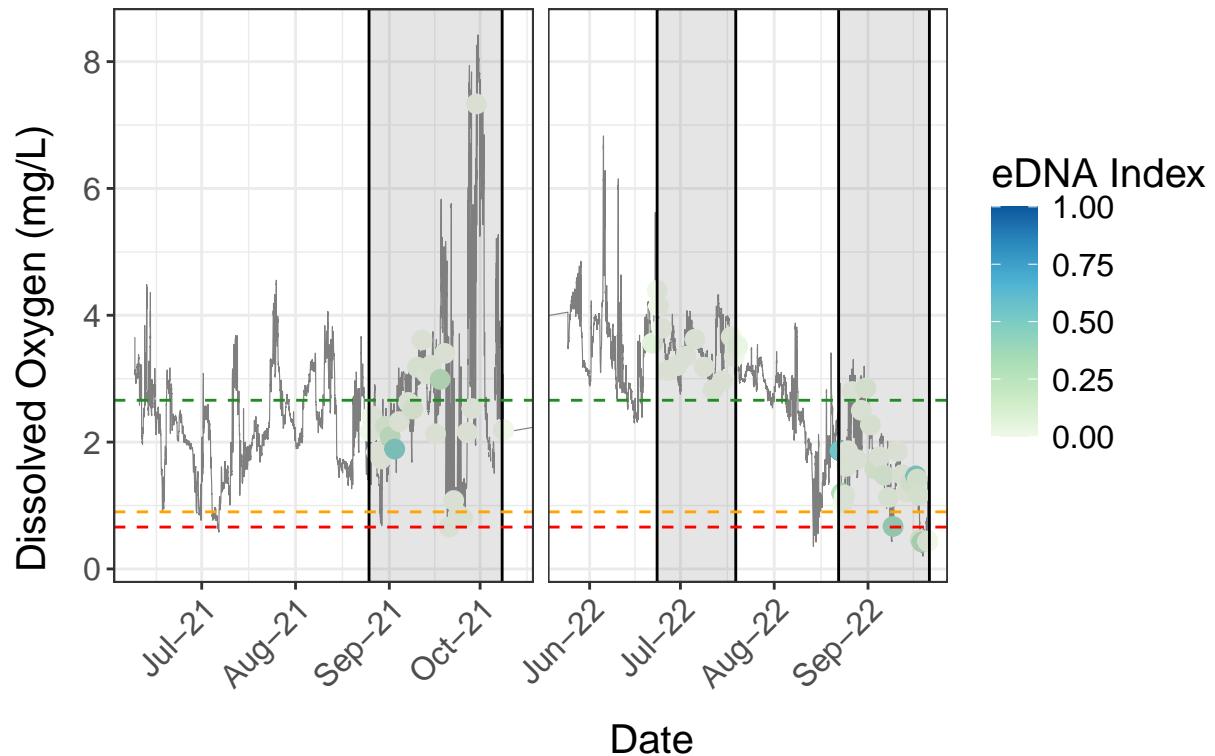
```
## [1] "Pseudocalanus newmani"  
## [1] "Pseudocalanus newmani eDNA Index vs Oxygen"
```

## Pseudocalanus newmani eDNA Index vs Oxygen



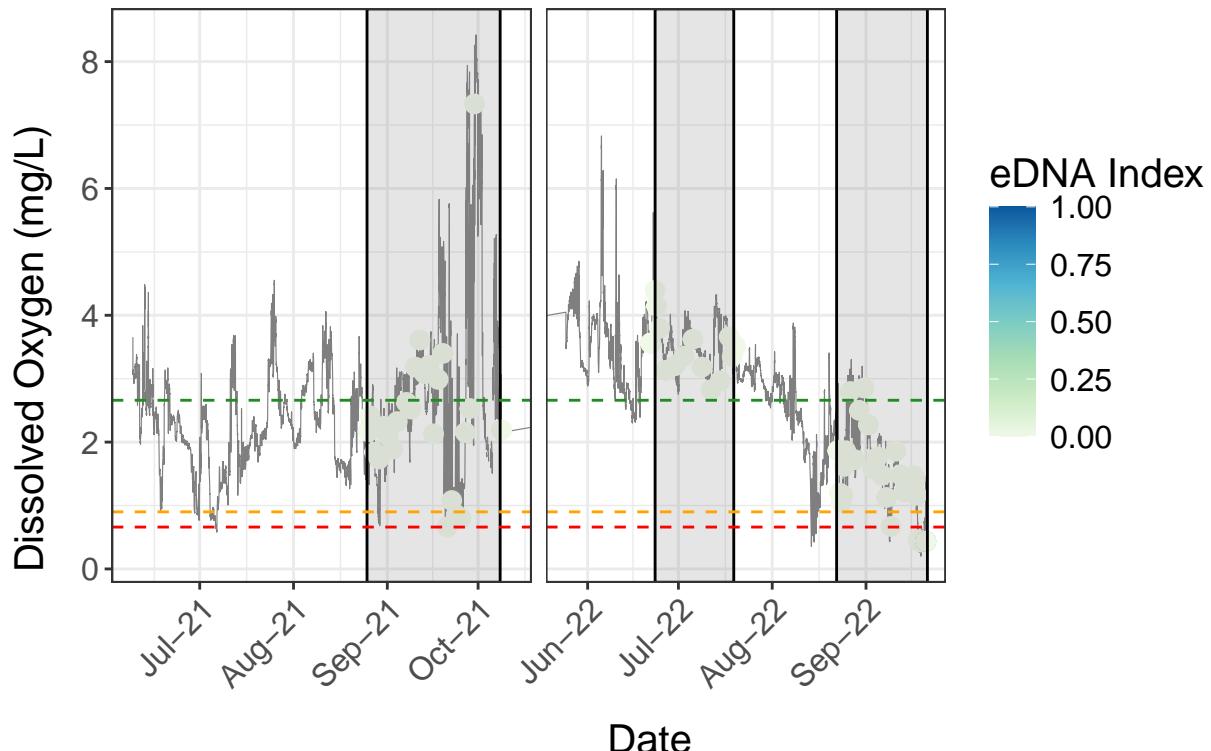
```
## [1] "Thermocyclops inversus"
## [1] "Thermocyclops inversus eDNA Index vs Oxygen"
```

## Thermocyclops inversus eDNA Index vs Oxygen



```
## [1] "Triconia minuta"  
## [1] "Triconia minuta eDNA Index vs Oxygen"
```

## Triconia minuta eDNA Index vs Oxygen



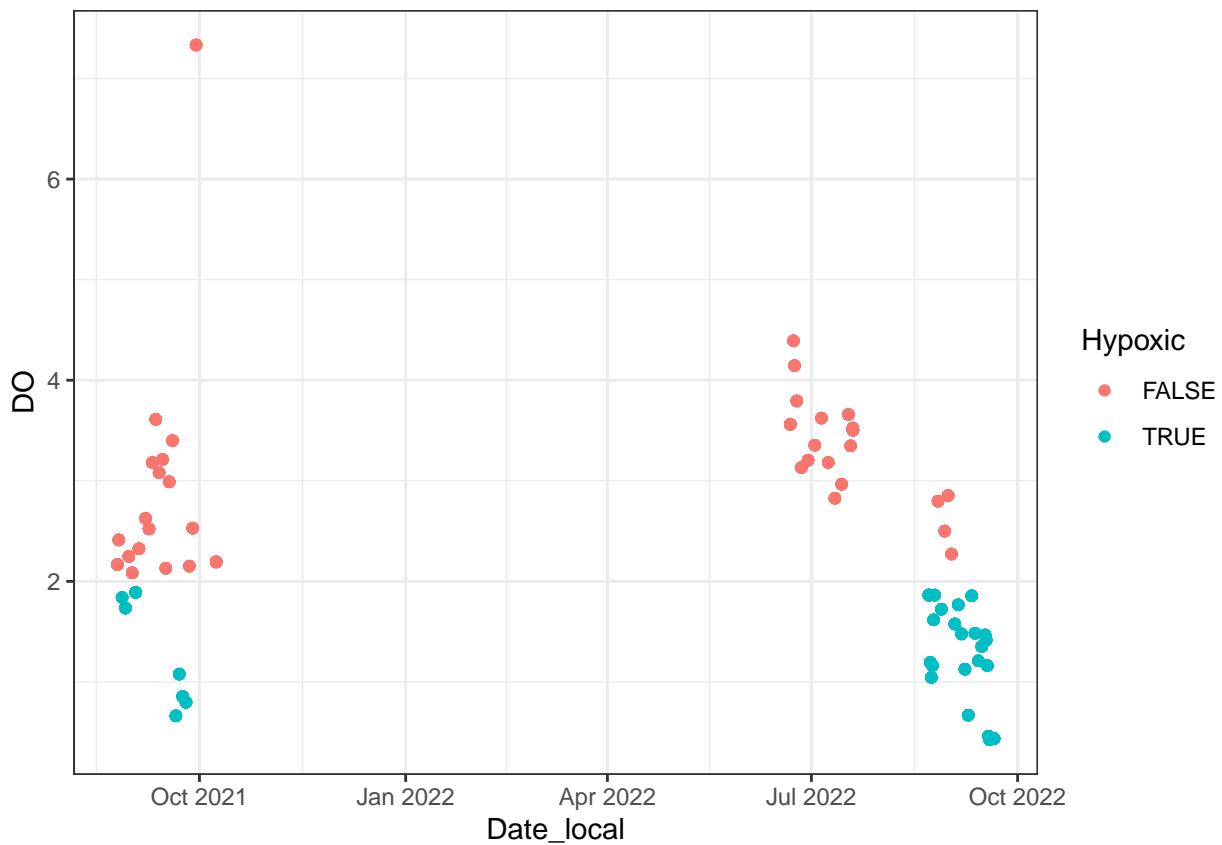
```
system("say Graphs complete")
```

### Scatterplot eDNA vs hypoxic threshold

```
eDNAindxEnvData_cleanYr <- eDNAindxEnvData_cleanYr %>%
  mutate(Hypoxic = case_when(DO < 2 ~ T, .default = F))

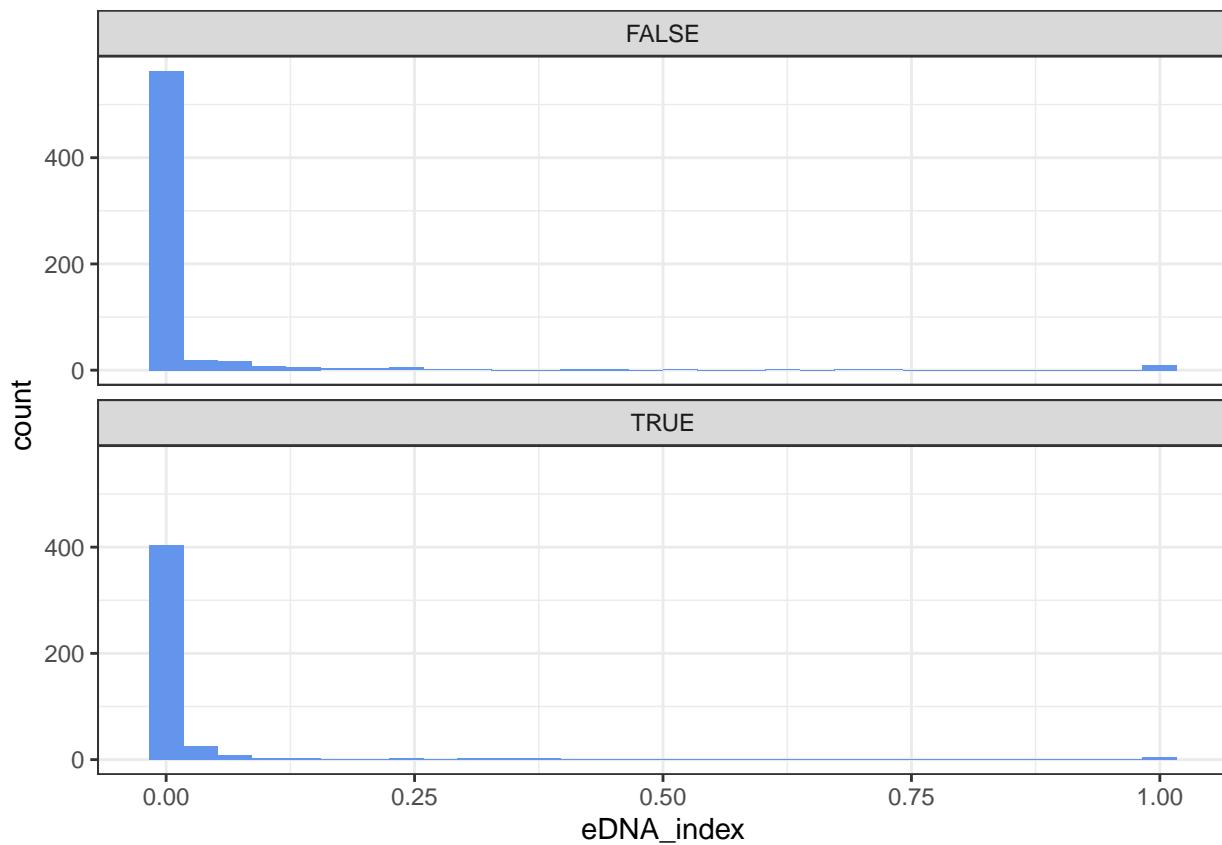
ggplot(eDNAindxEnvData_cleanYr, aes(x = Date_local, y = DO, color = Hypoxic)) +
  theme_bw() +
  geom_point()

## Warning: Removed 14 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



```
ggplot(eDNAindxEnvData_cleanYr, aes(x = eDNA_index)) +
  geom_histogram(fill = "cornflowerblue") +
  theme_bw() +
  facet_wrap(facets = vars(Hypoxic), ncol = 1)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



## GAM of eDNA index vs oxygen

```
dfsplit <- split(eDNAindxEnvData_cleanYr, eDNAindxEnvData_cleanYr$Species) # Split by
→ species

# Because the distribution of eDNA index is restricted to 0-1, Zack suggested using a
→ beta regression flavor of GAM. A linear predictor controls the mean, and it estimates
→ variance
# copied from documentation example, family = betar(link="logit")
# Picking a link function: can be "logit", "probit", "cloglog" and "cauchit". The
→ default is logit, I'm going with that for now.
# This will convert all the exactly zero values to 0.0000000000001ish, will that be a
→ problem since we have so many 0s? They're theoretically replaced with highly
→ improbable values. The documentation suggests manually resetting the 0s in a way that
→ reflects the sampling setup.

# Documentation says:
# These models are useful for proportions data which can not be modelled as binomial.
→ Note the assumption that data are in (0,1), despite the fact that for some parameter
→ values 0 and 1 are perfectly legitimate observations. The restriction is needed to
→ keep the log likelihood bounded for all parameter values. Any data exactly at 0 or 1
→ are reset to be just above 0 or just below 1 using the eps argument (in fact any
→ observation <eps is reset to eps and any observation >1-eps is reset to 1-eps). Note
→ the effect of this resetting. If mu phi>1 mu phi>1 then impossible 0s are replaced
→ with highly improbable eps values. If the inequality is reversed then 0s with
→ infinite probability density are replaced with eps values having high finite
→ probability density. The equivalent condition for 1s is (1-mu)phi>1(1-mu)phi>1.
→ Clearly all types of resetting are somewhat unsatisfactory, and care is needed if
→ data contain 0s or 1s (often it makes sense to manually reset the 0s and 1s in a
→ manner that somehow reflects the sampling setup).

for (i in 1:length(dfsplit)) { # For each species:
  species <- dfsplit[[i]]$Species[1] # Species name
  title <- paste(species, sep = " ", "eDNA Index vs Dissolved Oxygen") # Plot title,
→ changed for eDNA
  print(title)

  # Plot the GAM + save it
  # I don't think I can make geom_smooth do a beta regression --
  print(ggplot(dfsplit[[i]], aes(x = DO, y = eDNA_index)) + # plot this species
    geom_point(color = "orange2", alpha = 0.7) +
    geom_smooth(method = 'gam', se = F, color = "cornflowerblue") + # visualize GAM
    ggtitle(title) +
    theme_bw())
  ggsave(filename = paste(species, sep = "_", "eDNA_DO_GAM.png"), path =
    here("eDNA_Index_Hypoxia", "Plots", "GAM_Initial"), width = 2500, height = 2000,
    units = "px")

  # Make the GAM as an object and print some quality metrics
  ind_DO_GAM <- gam(eDNA_index ~ s(DO), bs = "cr", data = dfsplit[[i]], family =
    betar(link="logit"))
  print(summary(ind_DO_GAM))
```

```

print("GAM quality")
print(paste("AIC: ", sep = "", AIC(ind_DO_GAM))) # quality metric
print(summary(ind_DO_GAM)$sp.criterion)
print(gam.check(ind_DO_GAM, k.rep = 1000)) # Bunch of graphs and also text output,
→ which is used to check how the GAM is doing

## how to interpret
# edf = effective degrees of freedom
# GCV = generalized cross validation, an estimate of the mean square prediction
→ error. comparing between models, lower GCV is better
# Lower AIC is also better
# Higher R^2 is better

# system(paste(paste(paste("say", "Loop"), i), "Complete"))
}

## [1] "Acartia longiremis eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate

##
## Family: Beta regression(0.731)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.2644    0.1362  -9.281  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.001  1.001  6.302  0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0917  Deviance explained = -16.9%
## -REML = -239.29  Scale est. = 1           n = 78
## [1] "GAM quality"

```

```

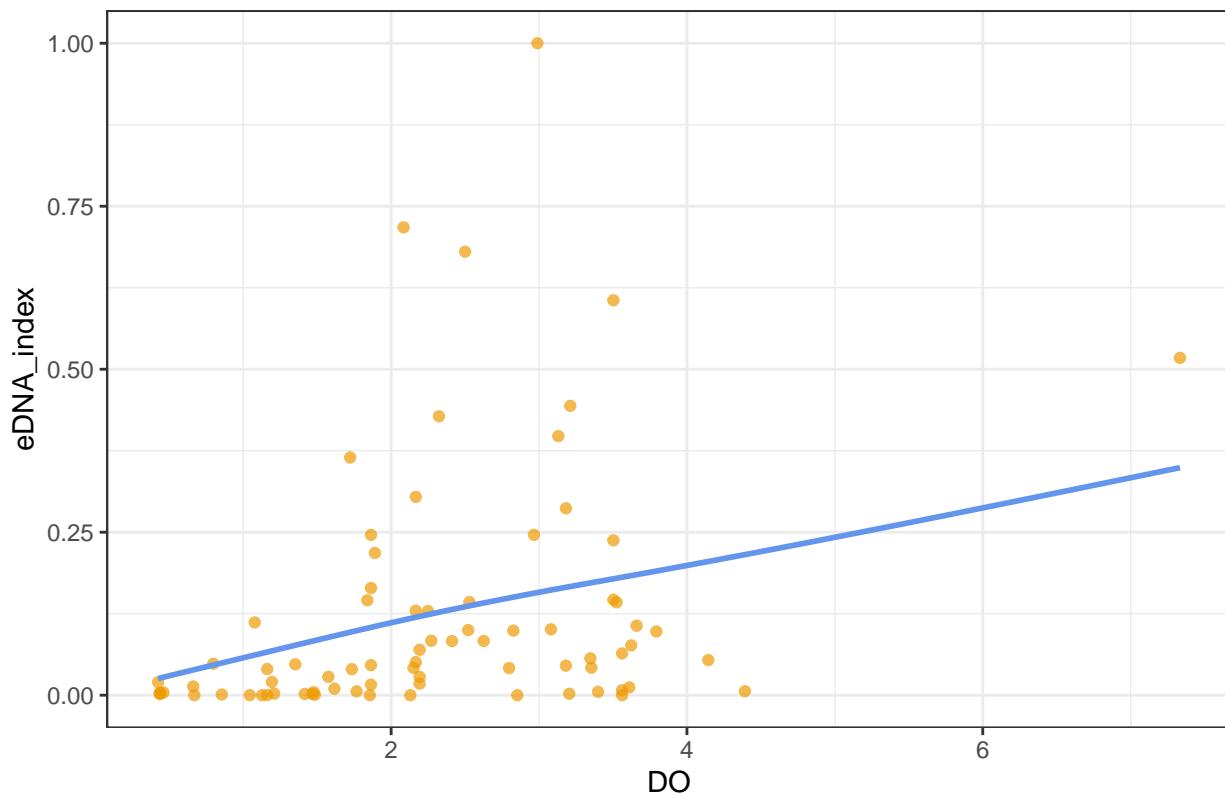
## [1] "AIC: -476.893118657352"
##      GCV.Cp
## -239.2911

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

Acartia longiremis eDNA Index vs Dissolved Oxygen

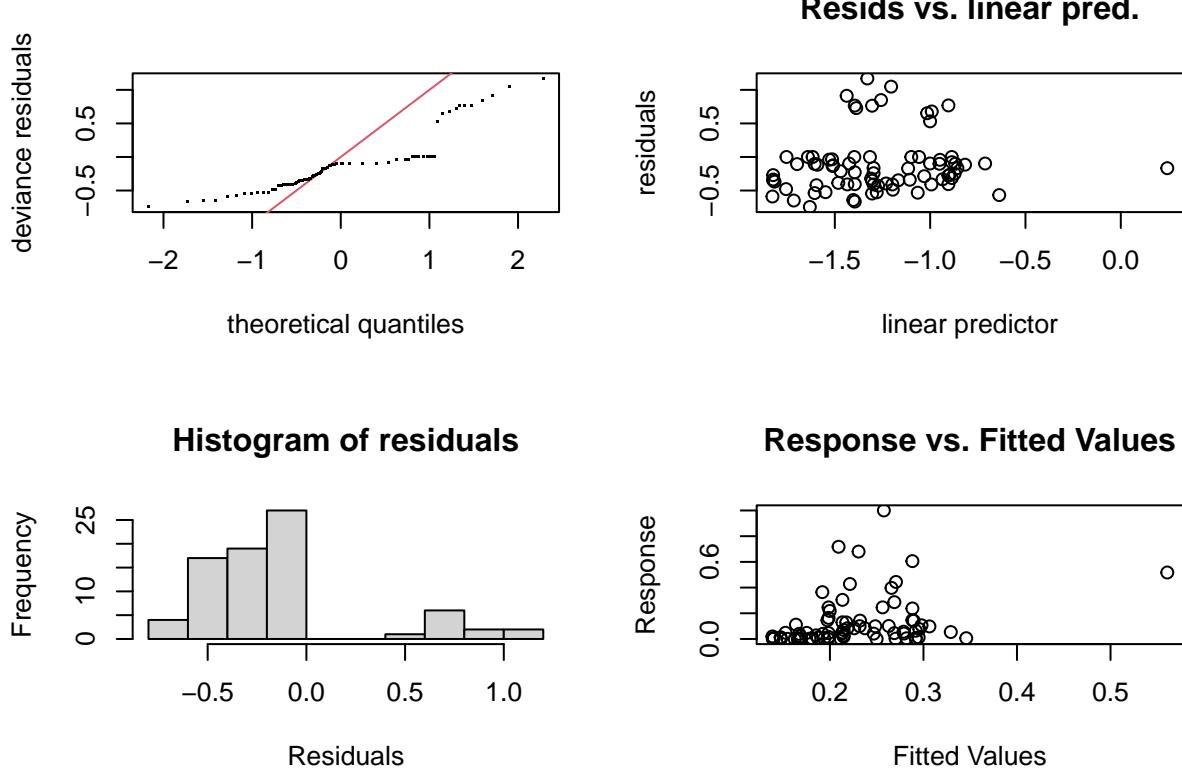


```

##
## Method: REML   Optimizer: outer newton
## full convergence after 8 iterations.
## Gradient range [-0.0001081796,-5.696983e-05]
## (score -239.2911 & scale 1).
## Hessian positive definite, eigenvalue range [0.0001081907,40.00607].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    1.07    0.9
## $mfrow
## [1] 2 2
##
## [1] "Calanus pacificus eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
##
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
##
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.301)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)

```

```

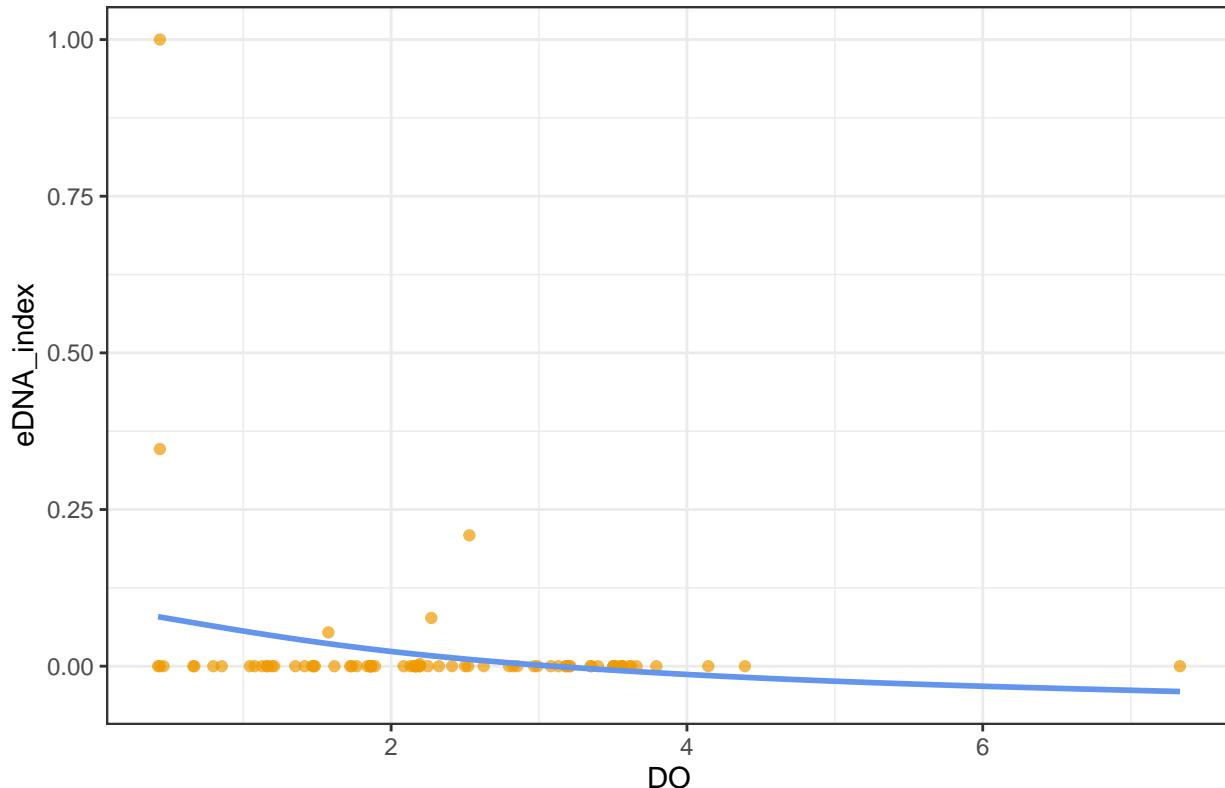
## 
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -2.1595    0.1253 -17.23   <2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Approximate significance of smooth terms:
##             edf Ref.df Chi.sq p-value    
## s(DO) 1.002 1.004 0.368  0.545    
## 
## R-sq.(adj) =  0.0114  Deviance explained = -4.44%    
## -REML = -1949.6  Scale est. = 1           n = 78    
## [1] "GAM quality"
## [1] "AIC: -3897.8481226595"
##      GCV.Cp
## -1949.645

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

Calanus pacificus eDNA Index vs Dissolved Oxygen



```

## 
## Method: REML  Optimizer: outer newton
## full convergence after 7 iterations.
## Gradient range [-0.0008483339,7.038015e-05]
## (score -1949.645 & scale 1).

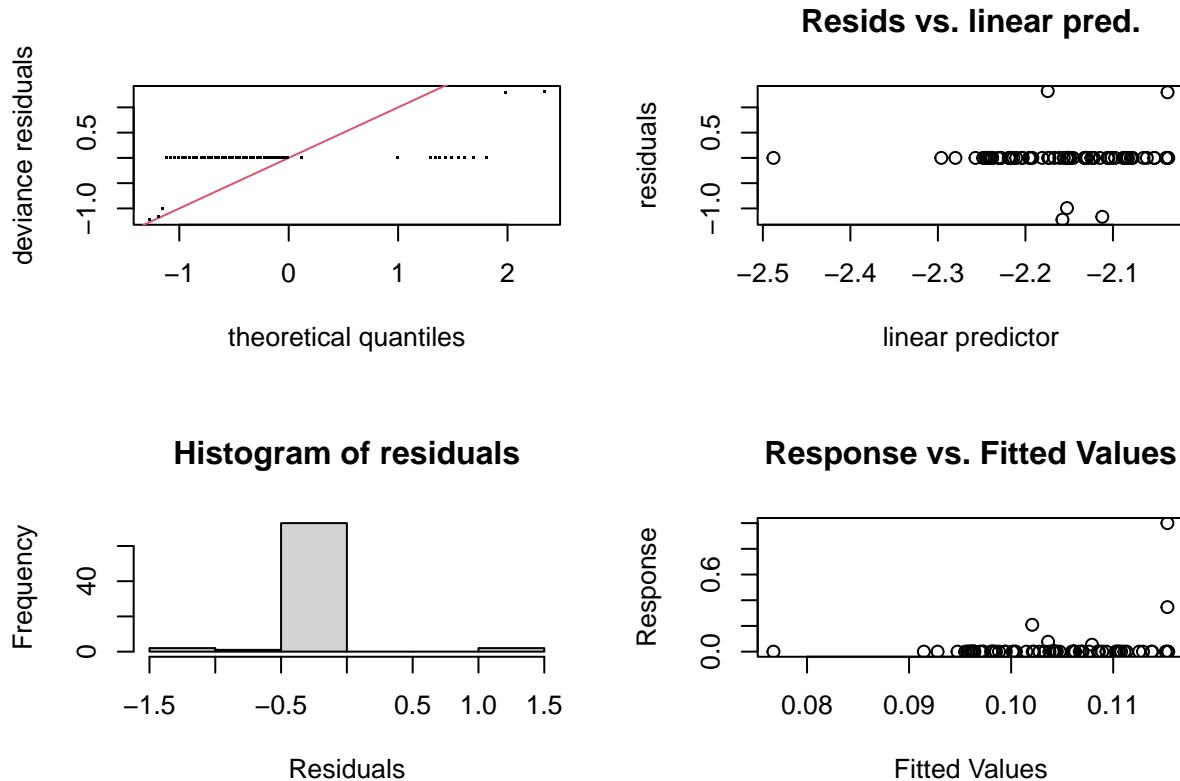
```

```

## Hessian positive definite, eigenvalue range [0.0008476686,17.58934].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO)  9    1    1.01    0.27
## $mfrw
## [1] 2 2
##
## [1] "Centropages abdominalis eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate

```

```

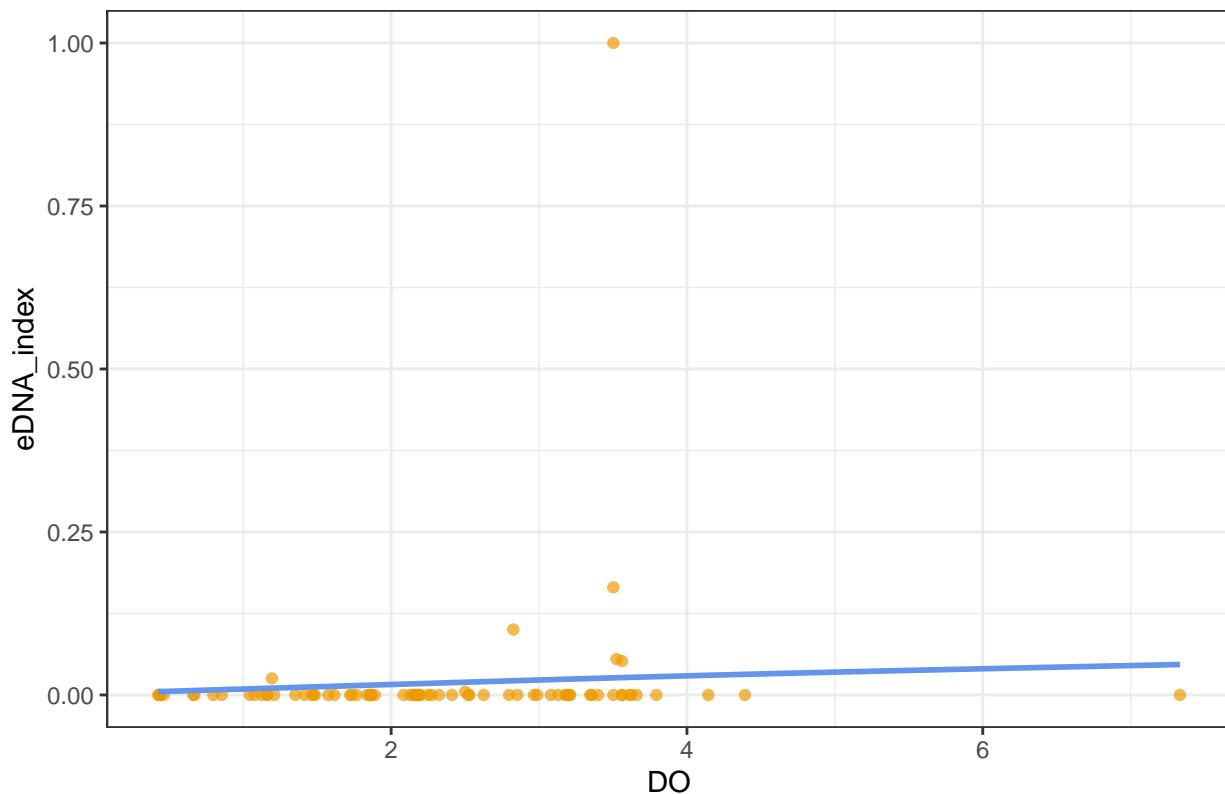
##
## Family: Beta regression(0.304)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.1583    0.1253 -17.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.002 1.004 0.256  0.614
##
## R-sq.(adj) = -0.00019  Deviance explained = -5.44%
## -REML = -1923.8  Scale est. = 1           n = 78
## [1] "GAM quality"
## [1] "AIC: -3846.28711010422"
##     GCV.Cp
## -1923.806

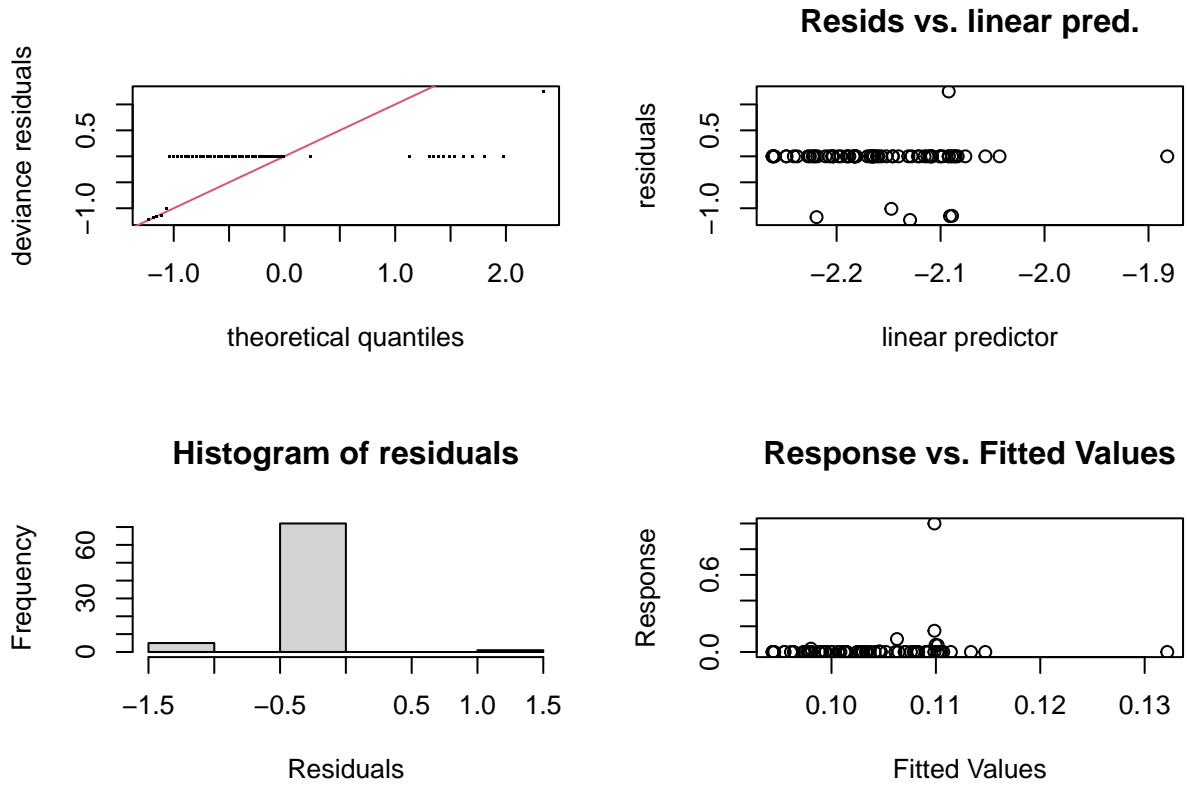
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Centropages abdominalis eDNA Index vs Dissolved Oxygen





```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    1.01    0.85
## $mfrow
## [1] 2 2
##
## [1] "Clausocalanus parapergens eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
##
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
##
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.298)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)

```

```

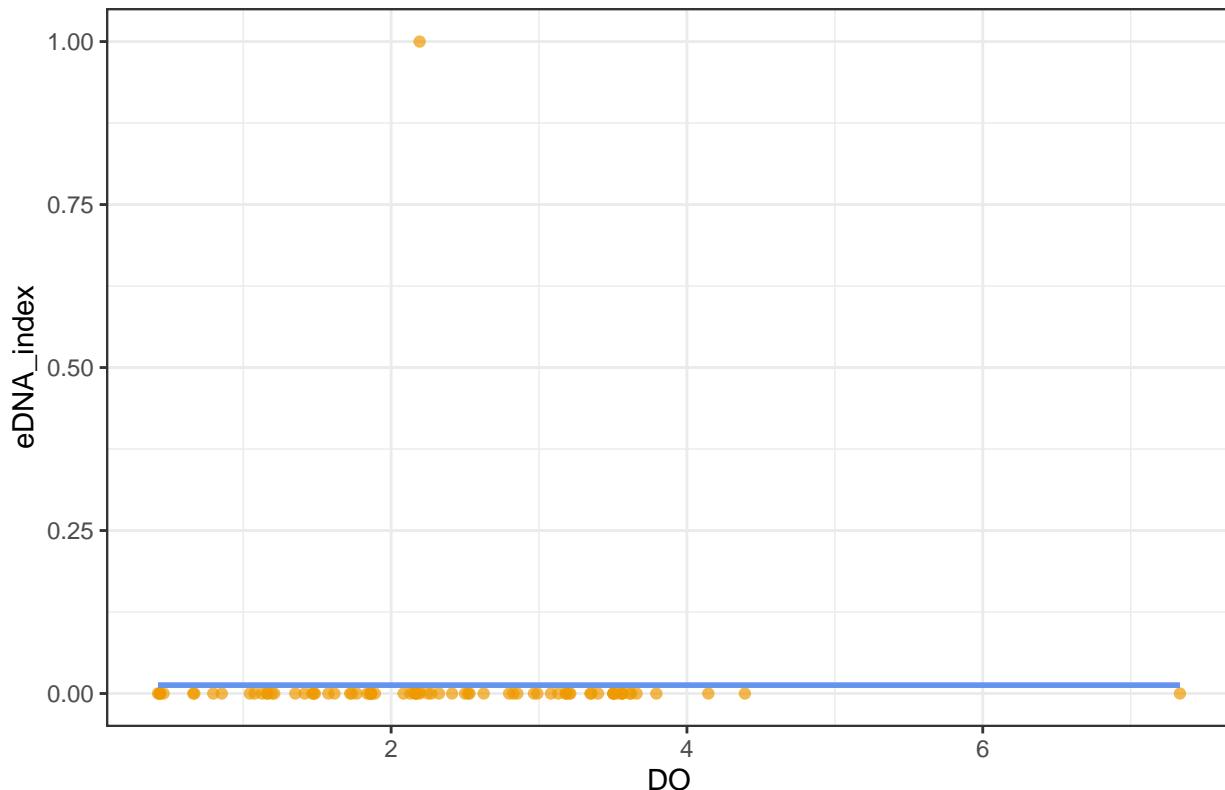
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.190     0.125  -17.53 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.001 1.003 0.001 0.983
##
## R-sq.(adj) = -0.0131 Deviance explained = -7.46%
## -REML = -2039.3 Scale est. = 1 n = 78
## [1] "GAM quality"
## [1] "AIC: -4077.2639763543"
## GCV.Cp
## -2039.295

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

Clausocalanus parapergens eDNA Index vs Dissolved Oxygen



```

##
## Method: REML Optimizer: outer newton
## full convergence after 8 iterations.
## Gradient range [-0.0006019291,8.804339e-05]
## (score -2039.295 & scale 1).

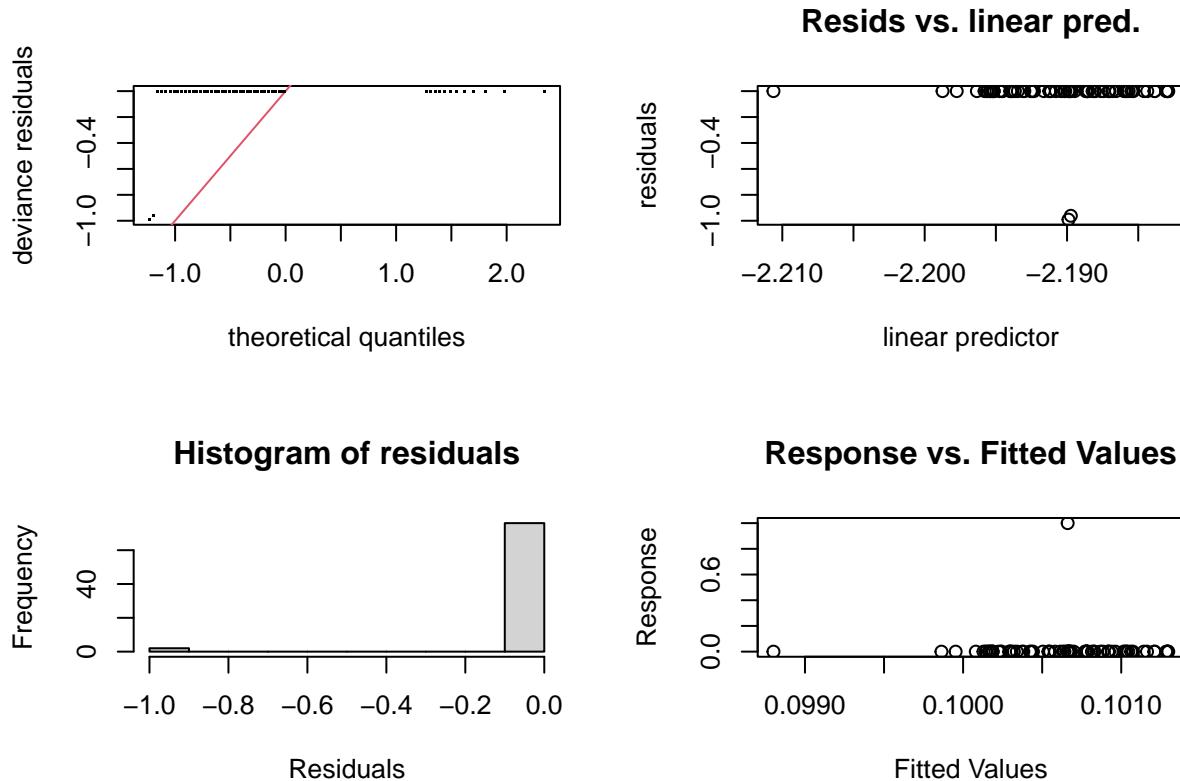
```

```

## Hessian positive definite, eigenvalue range [0.0006013575,16.99307].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    1.01    0.34
## $mfrw
## [1] 2 2
##
## [1] "Clausocalanus pergens eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate

```

```

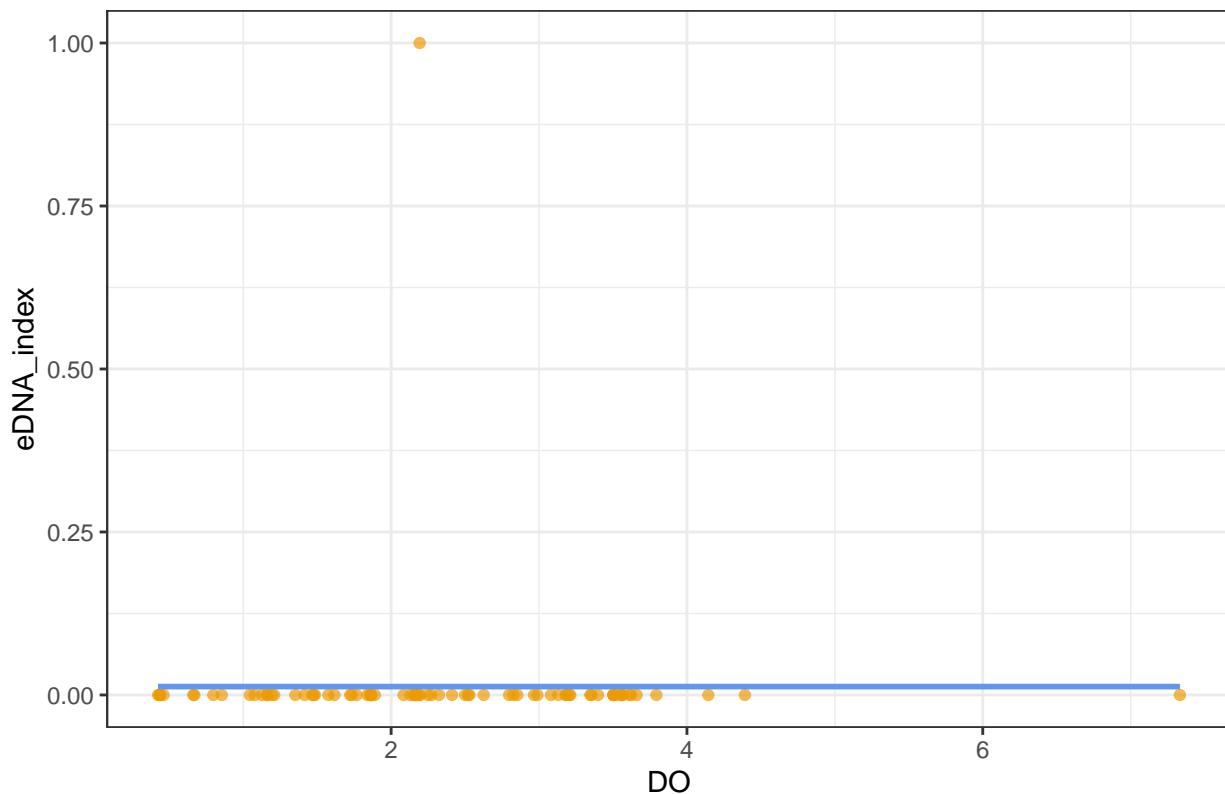
##
## Family: Beta regression(0.294)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.2003    0.1249 -17.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.002 1.005    0    0.996
##
## R-sq.(adj) = -0.0131  Deviance explained = -4.22%
## -REML = -2087.9  Scale est. = 1           n = 78
## [1] "GAM quality"
## [1] "AIC: -4174.45882641852"
## GCV.Cp
## -2087.9

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

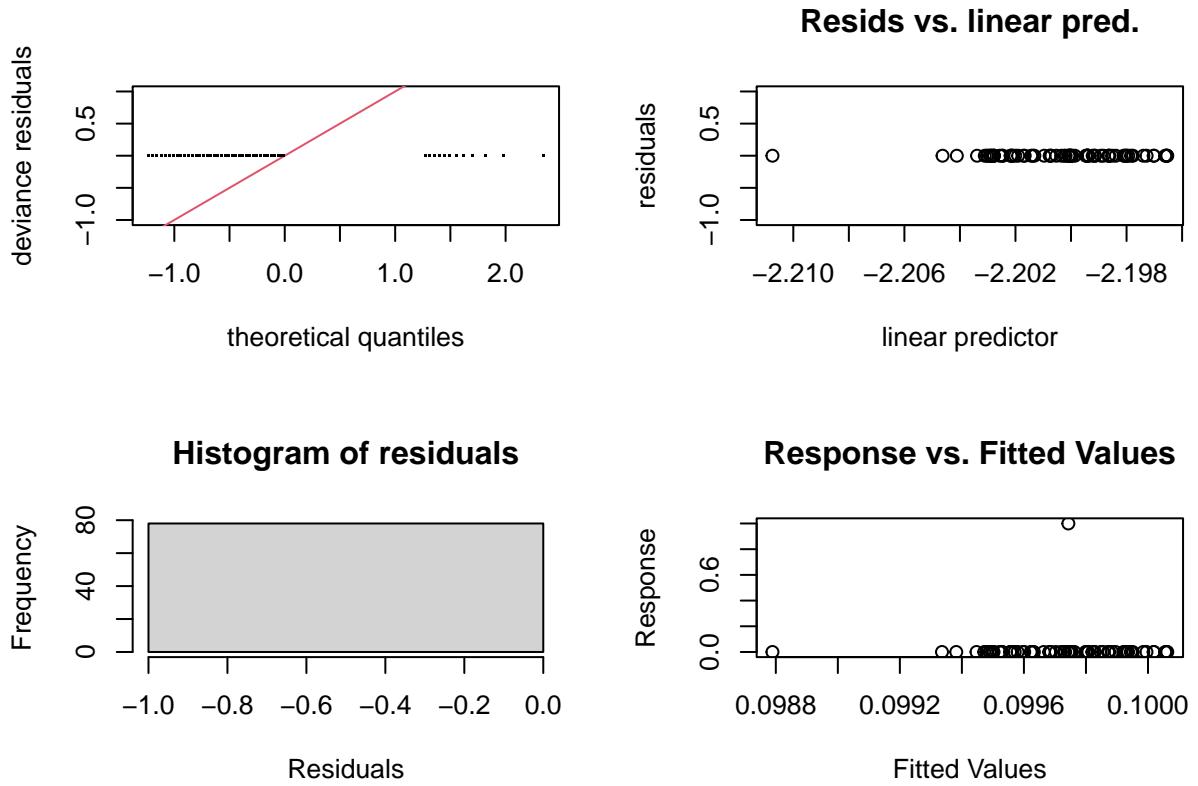
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Clausocalanus pergens eDNA Index vs Dissolved Oxygen



```
##  
## Method: REML    Optimizer: outer newton  
## full convergence after 8 iterations.  
## Gradient range [-0.001171398,0.0001876884]  
## (score -2087.9 & scale 1).  
## Hessian positive definite, eigenvalue range [0.001169388,16.82959].  
## Model rank = 10 / 10  
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):  
## saturated likelihood may be inaccurate
```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO)  9     1    NaN <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## $mfrow
## [1] 2 2
##
## [1] "Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.294)
## Link function: logit
##

```

```

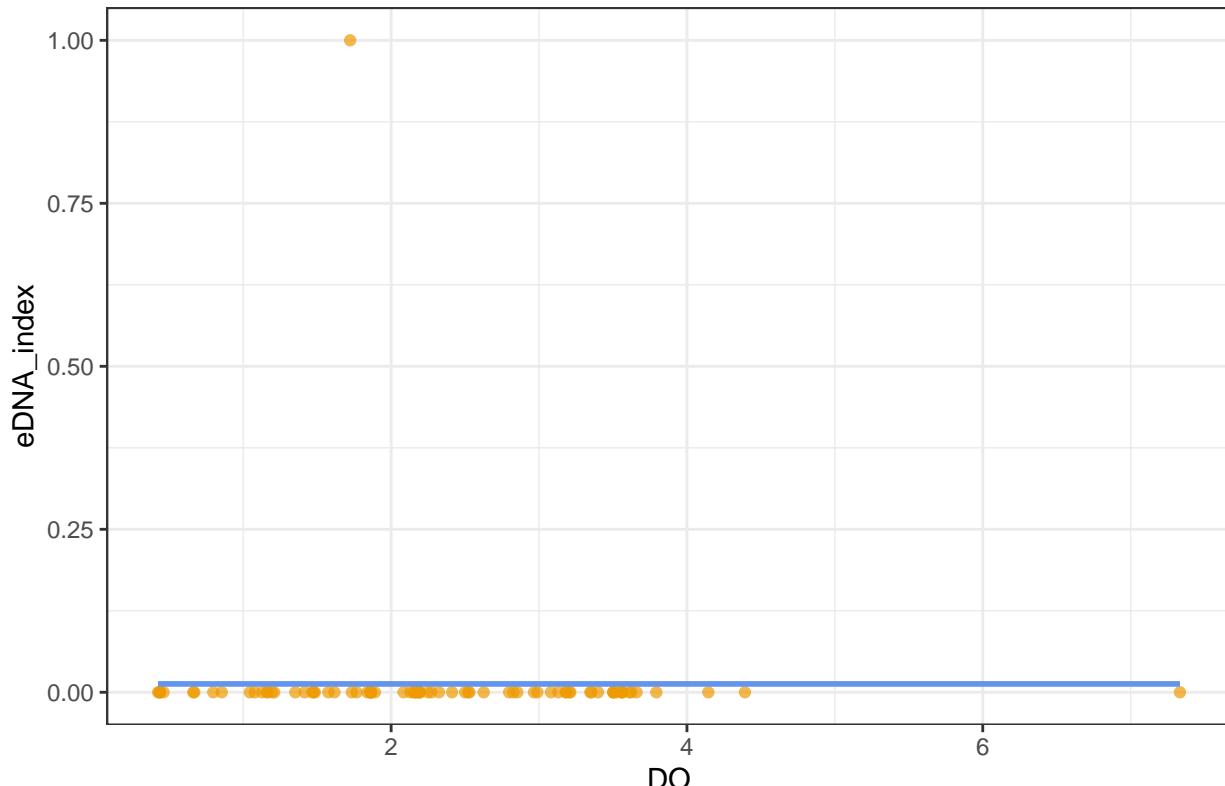
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.2005    0.1249 -17.62  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df Chi.sq p-value
## s(DO) 1.002 1.005  0.01    0.93
##
## R-sq.(adj) = -0.0121  Deviance explained = -4.22%
## -REML = -2087.9  Scale est. = 1          n = 78
## [1] "GAM quality"
## [1] "AIC: -4174.46920781147"
## GCV.Cp
## -2087.911

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen



```

##
## Method: REML   Optimizer: outer newton
## full convergence after 8 iterations.

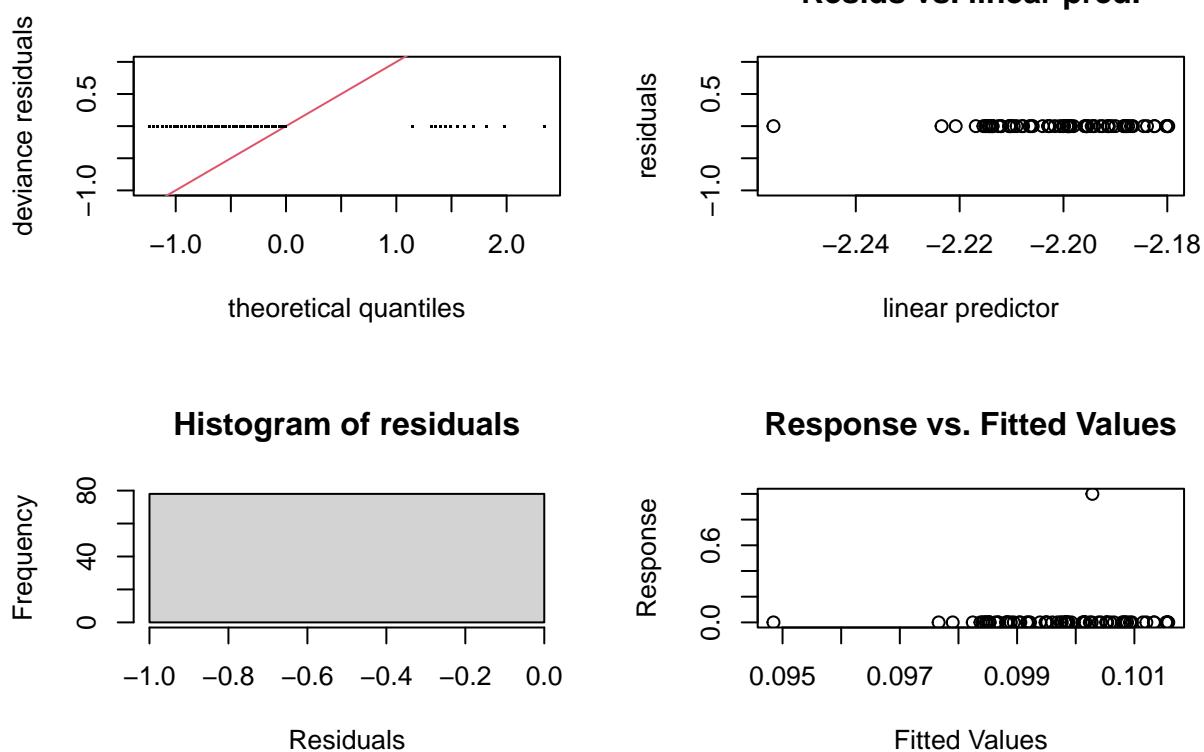
```

```

## Gradient range [-0.001156205,0.0001866311]
## (score -2087.911 & scale 1).
## Hessian positive definite, eigenvalue range [0.001154345,16.83226].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##          k'  edf k-index p-value
## s(DO)    9     1      NaN  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## $mfrow
## [1] 2 2
##
## [1] "Lucicutia flavigornis eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range

```

```

## (`geom_point()`).

## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate

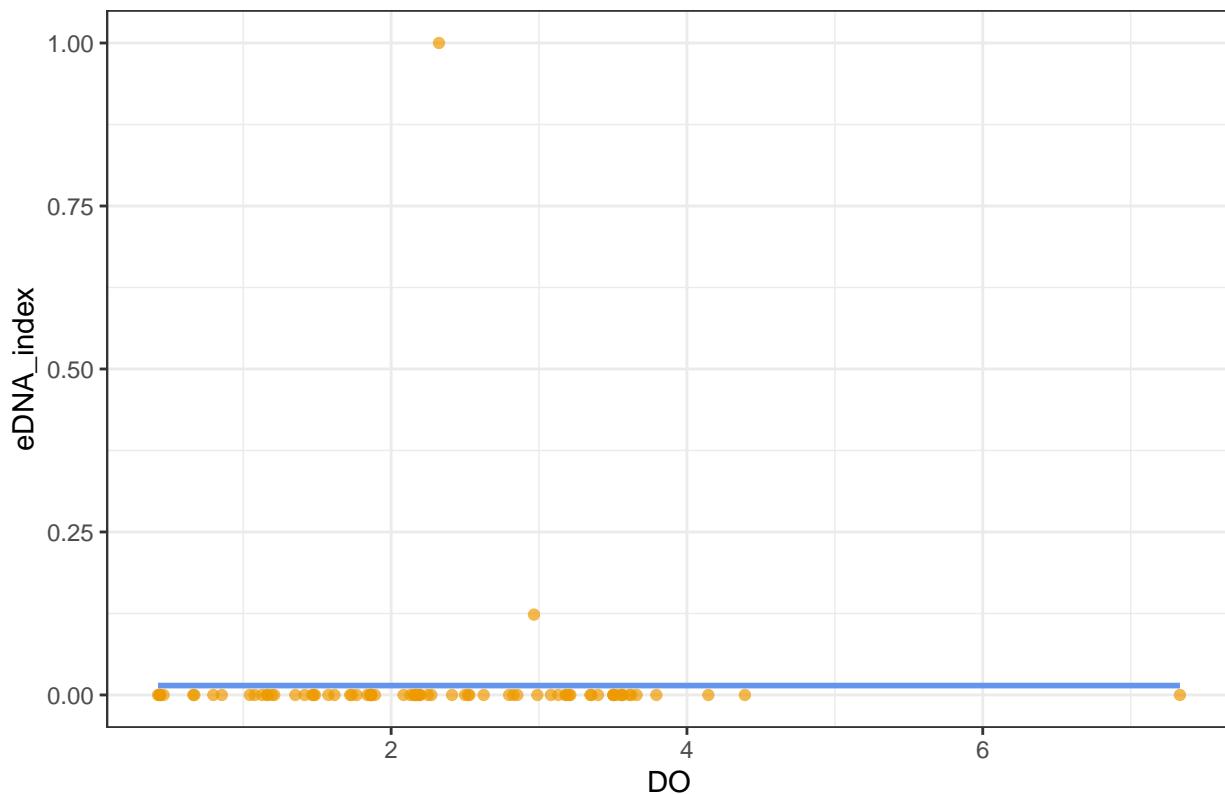
##
## Family: Beta regression(0.295)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.1922    0.1249 -17.55  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.001 1.003 0.003  0.964
##
## R-sq.(adj) = -0.0131  Deviance explained = -5.39%
## -REML = -2059.5  Scale est. = 1           n = 78
## [1] "GAM quality"
## [1] "AIC: -4117.70204337752"
## GCV.Cp
## -2059.515

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

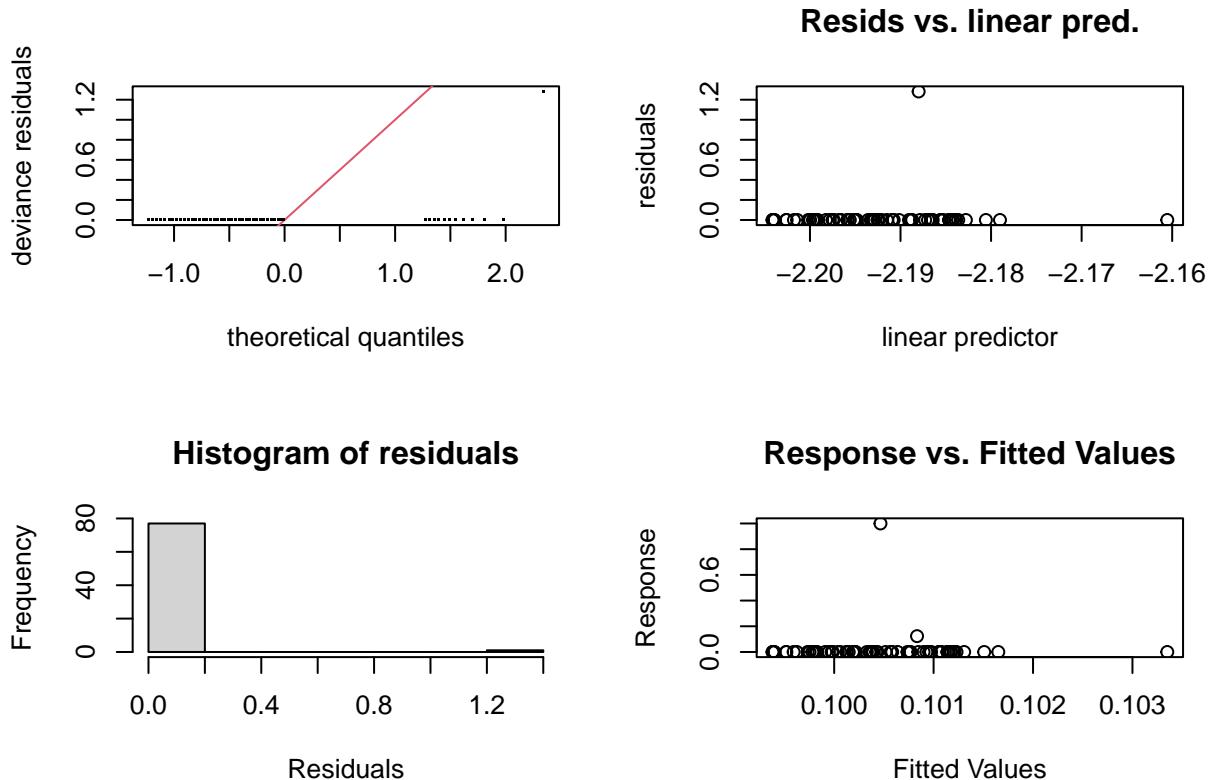
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Lucicutia flavigornis eDNA Index vs Dissolved Oxygen



```
##  
## Method: REML    Optimizer: outer newton  
## full convergence after 8 iterations.  
## Gradient range [-0.000658527,0.0001039363]  
## (score -2059.515 & scale 1).  
## Hessian positive definite, eigenvalue range [0.0006578594,16.96007].  
## Model rank = 10 / 10  
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):  
## saturated likelihood may be inaccurate
```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    1.01    0.56
## $mfrow
## [1] 2 2
##
## [1] "Metridia lucens eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.294)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)

```

```

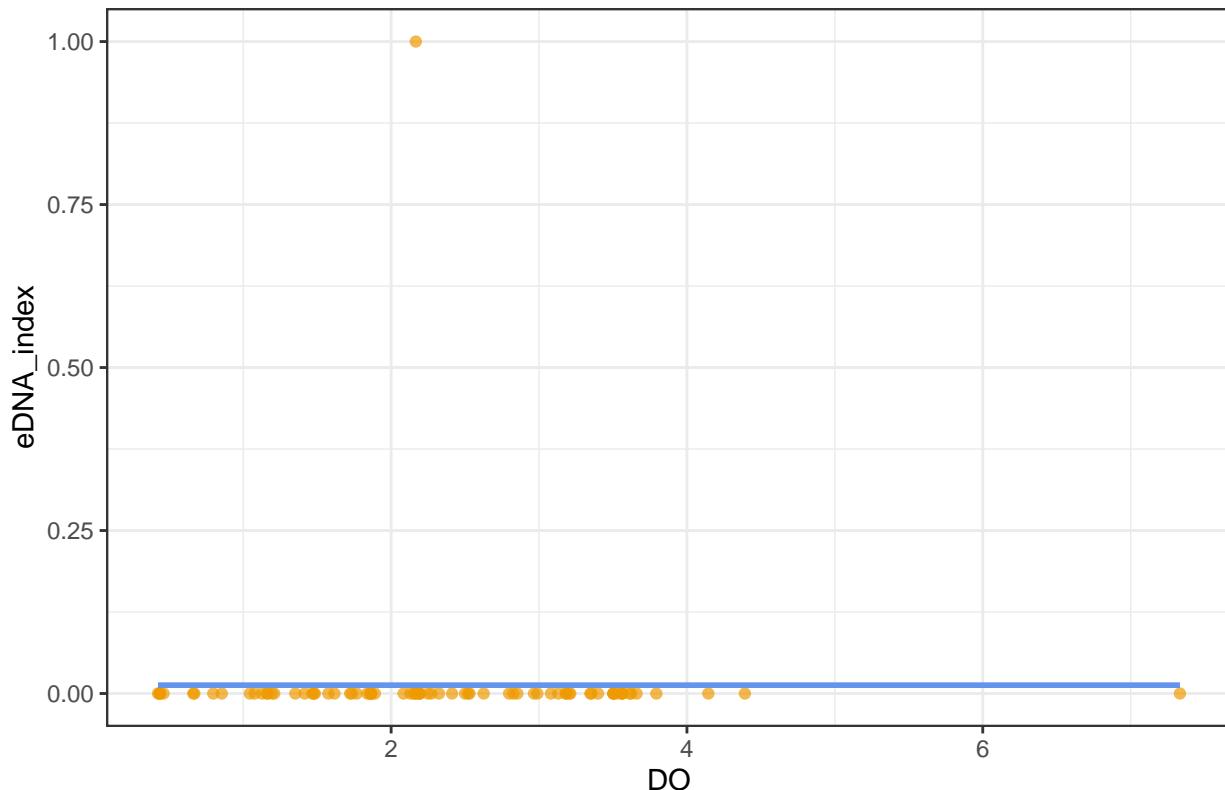
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.2003    0.1249 -17.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df Chi.sq p-value
## s(DO) 1.002 1.005 0.001 0.995
##
## R-sq.(adj) = -0.0131  Deviance explained = -4.22%
## -REML = -2087.9  Scale est. = 1          n = 78
## [1] "GAM quality"
## [1] "AIC: -4174.45901539693"
## GCV.Cp
## -2087.901

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

Metridia lucens eDNA Index vs Dissolved Oxygen



```

##
## Method: REML  Optimizer: outer newton
## full convergence after 8 iterations.
## Gradient range [-0.001172453,0.0001878609]
## (score -2087.901 & scale 1).

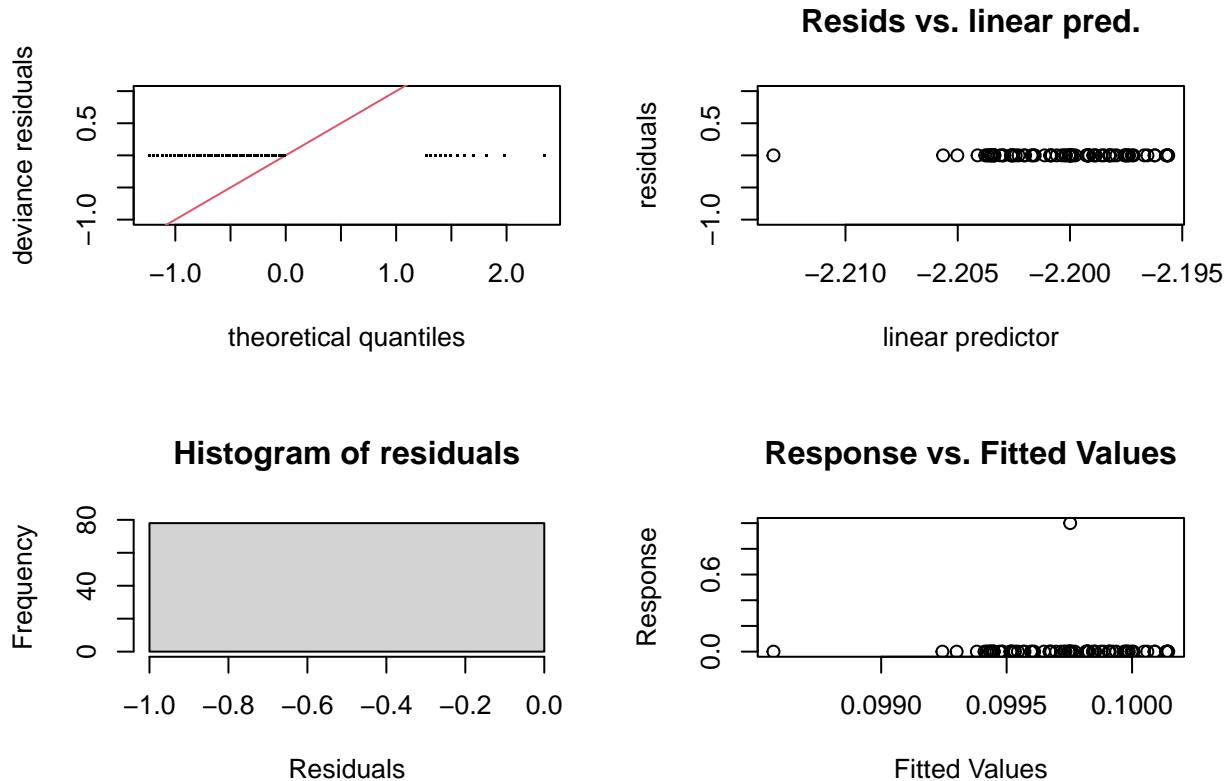
```

```

## Hessian positive definite, eigenvalue range [0.001170443,16.82967].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO)  9    1     NaN <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## $mfrow
## [1] 2 2
##
## [1] "Metridia pacifica eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood

```

```

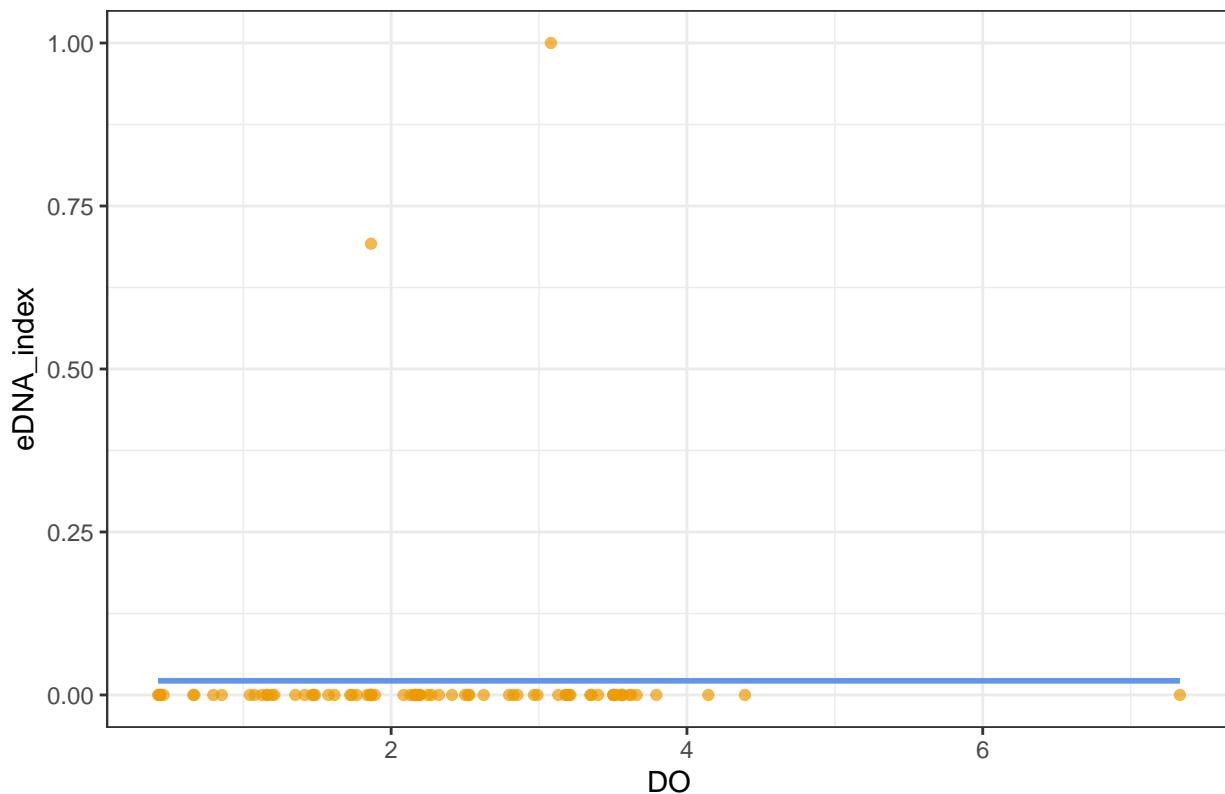
## may be inaccurate
##
## Family: Beta regression(0.29)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.1744    0.1251 -17.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.002 1.003  0.01    0.93
##
## R-sq.(adj) = -0.0126  Deviance explained = -3.39%
## -REML = -2058.6  Scale est. = 1      n = 78
## [1] "GAM quality"
## [1] "AIC: -4115.89551050575"
## GCV.Cp
## -2058.618

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

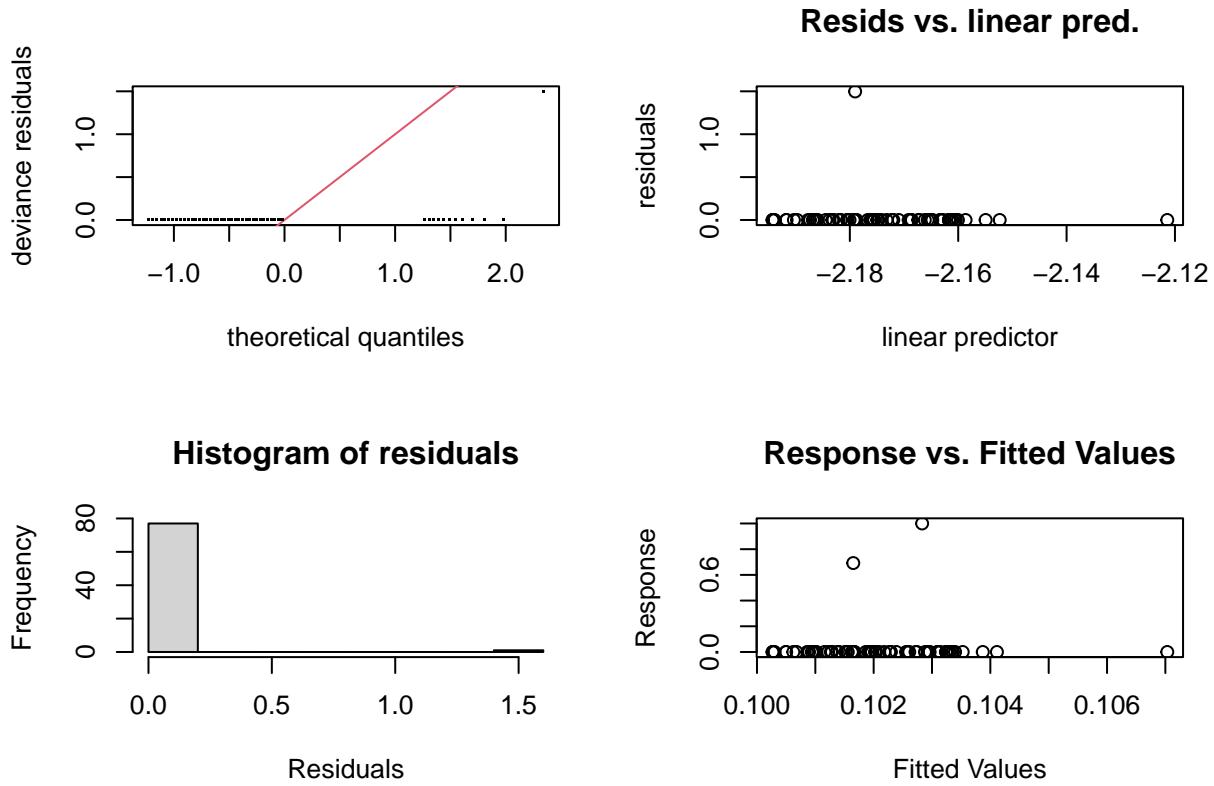
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Metridia pacifica eDNA Index vs Dissolved Oxygen



```
##  
## Method: REML    Optimizer: outer newton  
## full convergence after 8 iterations.  
## Gradient range [-0.0007662289,0.000129087]  
## (score -2058.618 & scale 1).  
## Hessian positive definite, eigenvalue range [0.0007653417,17.25282].  
## Model rank = 10 / 10  
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):  
## saturated likelihood may be inaccurate
```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    1.01    0.23
## $mfrow
## [1] 2 2
##
## [1] "Oithona similis eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
##
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
##
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.309)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)

```

```

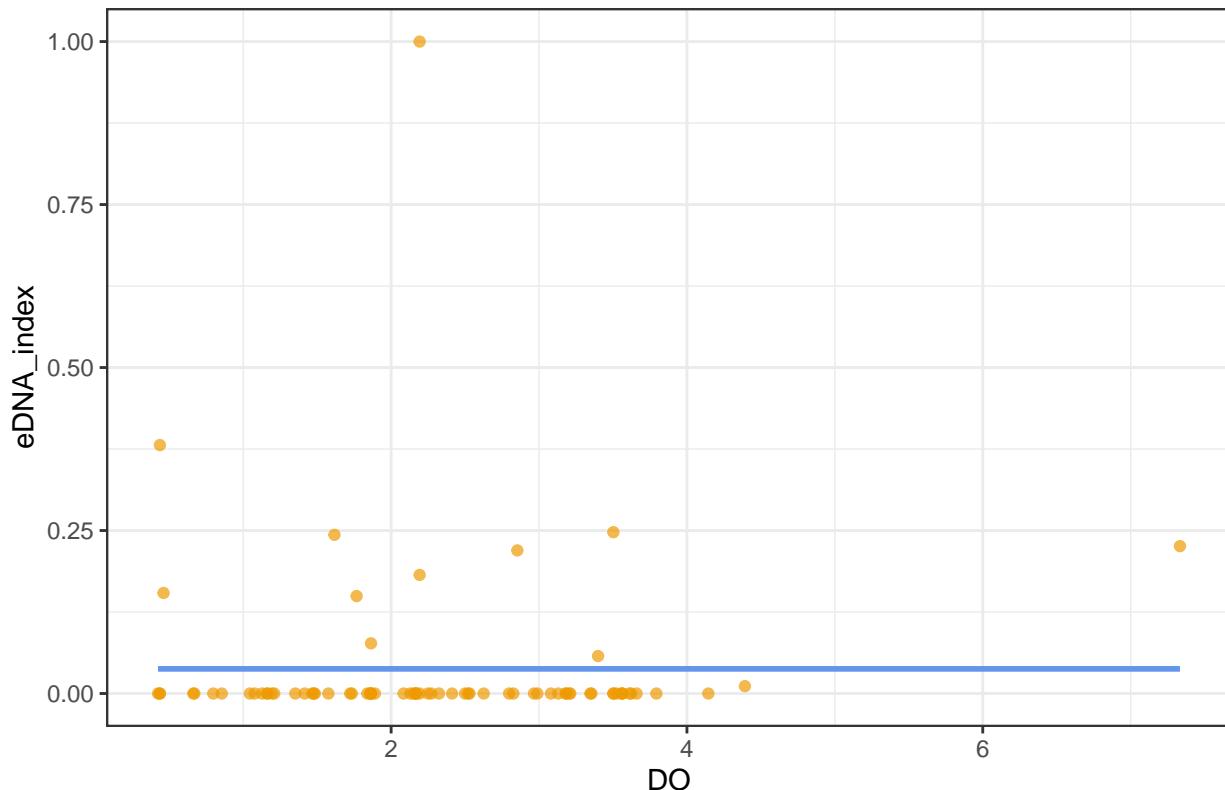
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.100     0.126  -16.67 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.146 1.278 0.403 0.733
##
## R-sq.(adj) = -0.011  Deviance explained = -2.28%
## -REML = -1776.2  Scale est. = 1      n = 78
## [1] "GAM quality"
## [1] "AIC: -3550.54966738607"
## GCV.Cp
## -1776.241

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Oithona similis eDNA Index vs Dissolved Oxygen



```

##
## Method: REML   Optimizer: outer newton
## full convergence after 6 iterations.
## Gradient range [-0.0008310726,-0.0005373641]
## (score -1776.241 & scale 1).

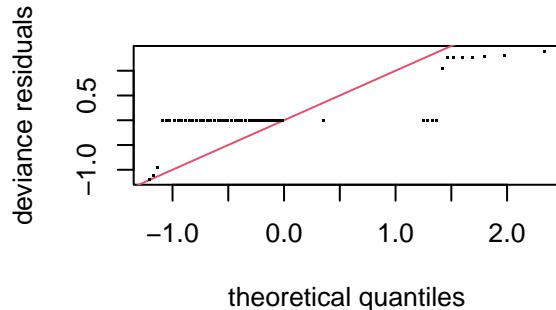
```

```

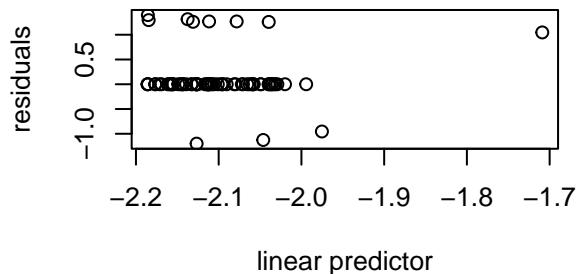
## Hessian positive definite, eigenvalue range [0.006838523,18.67712].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

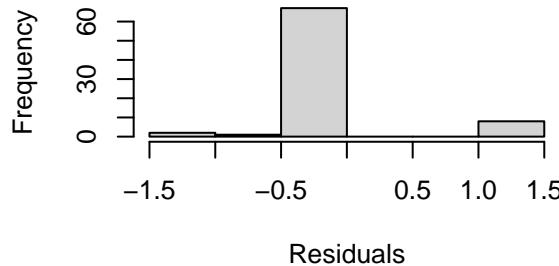
```



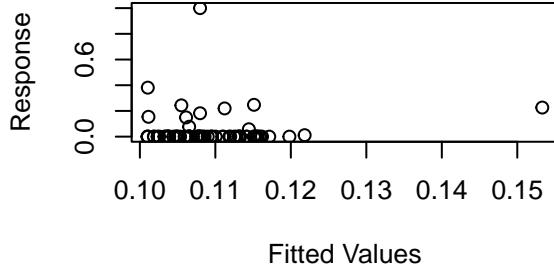
**Resids vs. linear pred.**



**Histogram of residuals**



**Response vs. Fitted Values**



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##          k'  edf k-index p-value
## s(DO) 9.00 1.15    1.13    0.96
## $mfrow
## [1] 2 2
##
## [1] "Paracalanus sp. C AC-2013 eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate

```

```

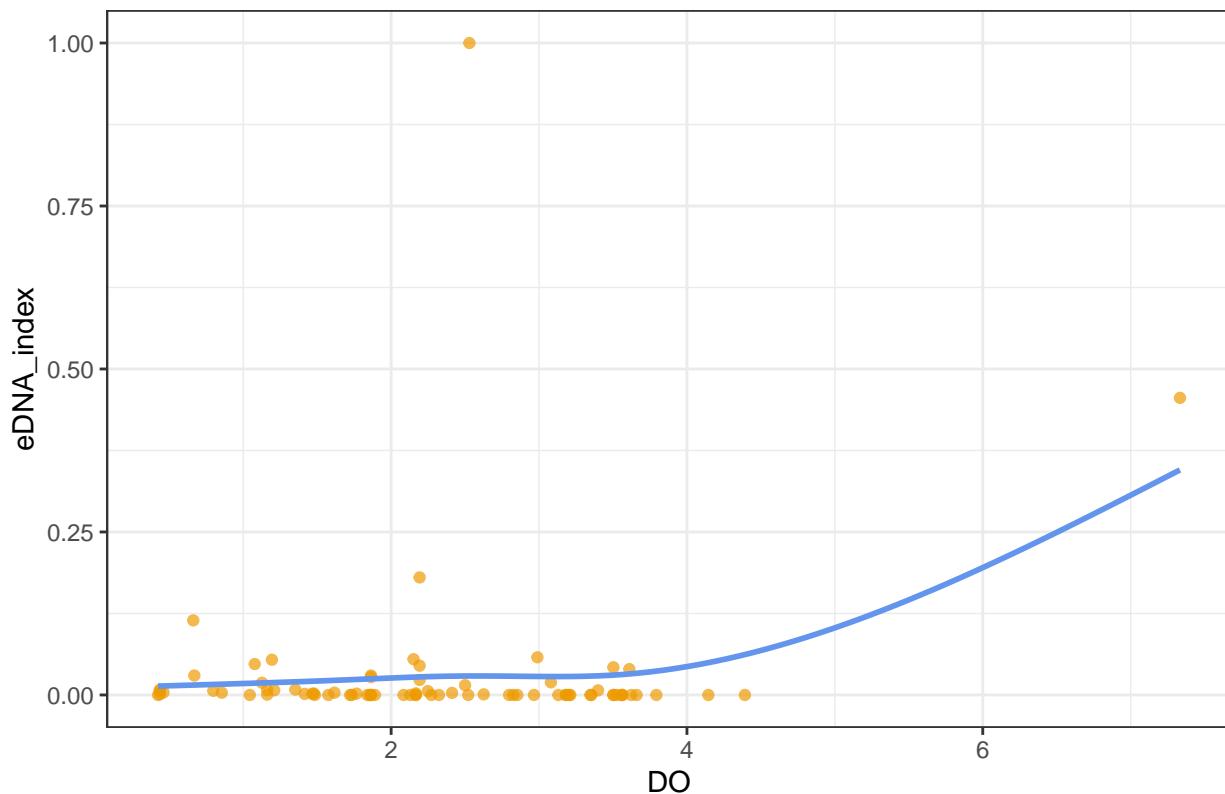
##
## Family: Beta regression(0.413)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.9383    0.1282 -15.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 2.041  2.558  6.205  0.064 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = -0.0944  Deviance explained = -8.52%
## -REML = -1132.9  Scale est. = 1      n = 78
## [1] "GAM quality"
## [1] "AIC: -2263.48718738886"
## GCV.Cp
## -1132.941

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

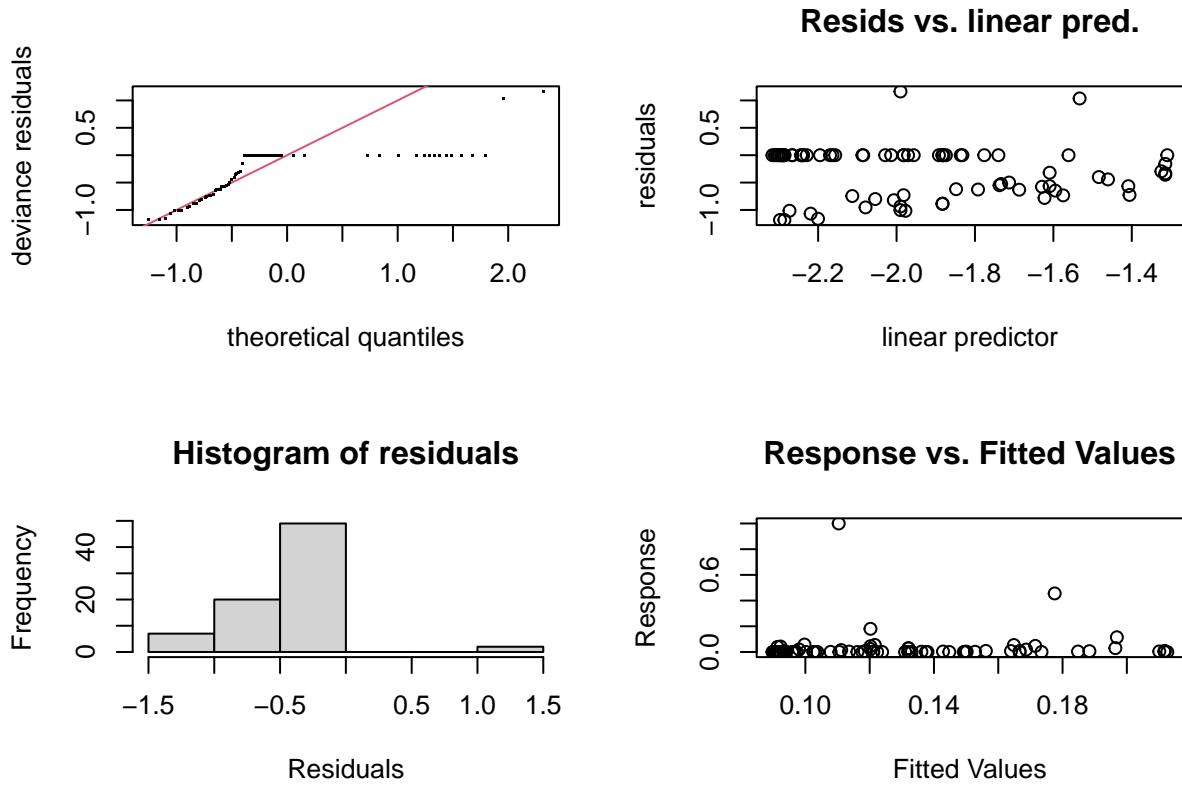
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Paracalanus sp. C AC–2013 eDNA Index vs Dissolved Oxygen



```
##  
## Method: REML    Optimizer: outer newton  
## full convergence after 3 iterations.  
## Gradient range [-8.4033e-06,9.109311e-05]  
## (score -1132.941 & scale 1).  
## Hessian positive definite, eigenvalue range [0.3018038,22.70014].  
## Model rank = 10 / 10  
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):  
## saturated likelihood may be inaccurate
```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##          k'   edf k-index p-value
## s(DO) 9.00 2.04     0.59    0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## $mfrow
## [1] 2 2
##
## [1] "Pseudocalanus newmani eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.354)
## Link function: logit
##

```

```

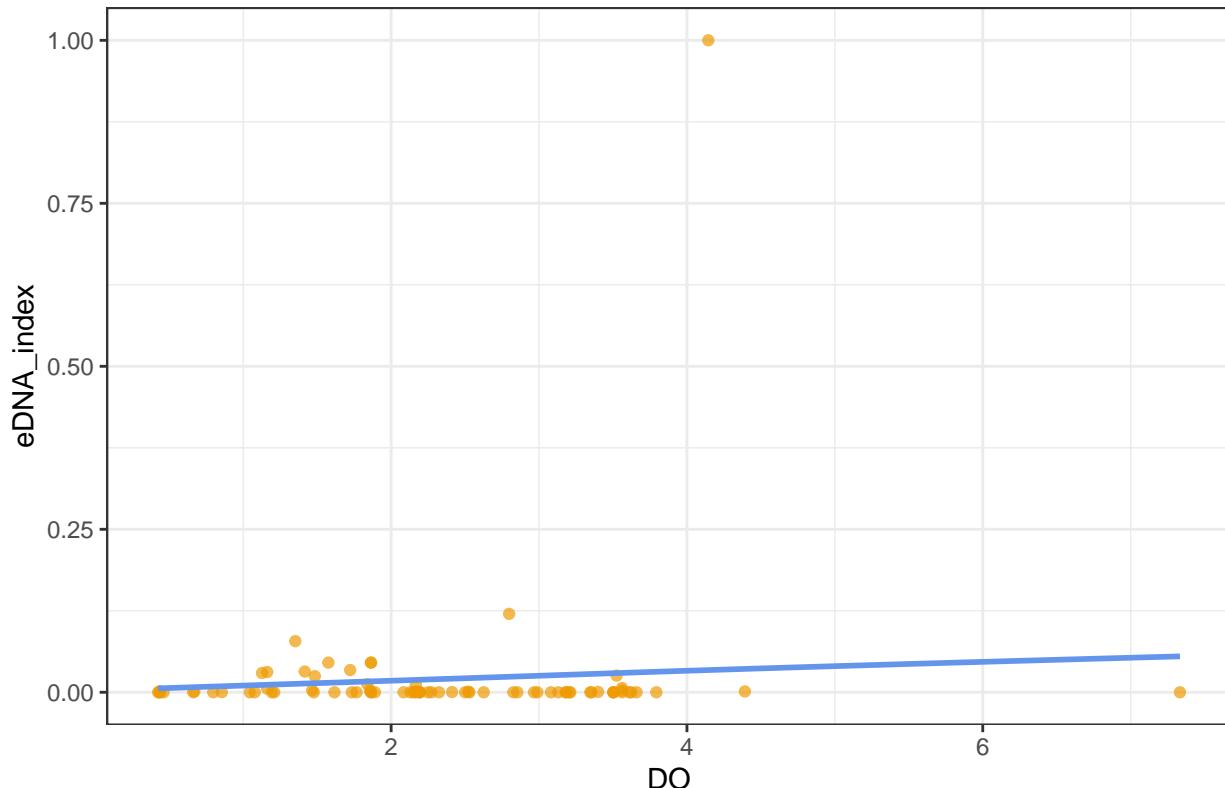
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.0286    0.1268    -16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df Chi.sq p-value
## s(DO) 1.001 1.003 0.122  0.728
##
## R-sq.(adj) = -0.0275  Deviance explained = -10.5%
## -REML = -1401.7  Scale est. = 1          n = 78
## [1] "GAM quality"
## [1] "AIC: -2802.03221853363"
## GCV.Cp
## -1401.72

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Pseudocalanus newmani eDNA Index vs Dissolved Oxygen



```

##
## Method: REML  Optimizer: outer newton
## full convergence after 7 iterations.

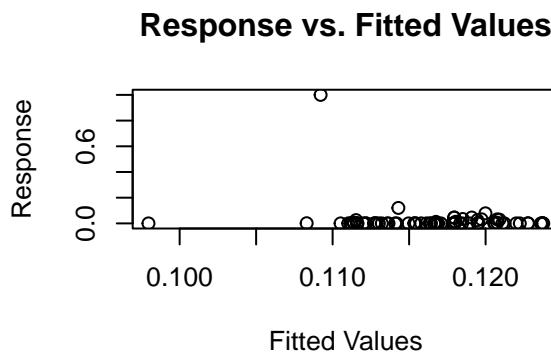
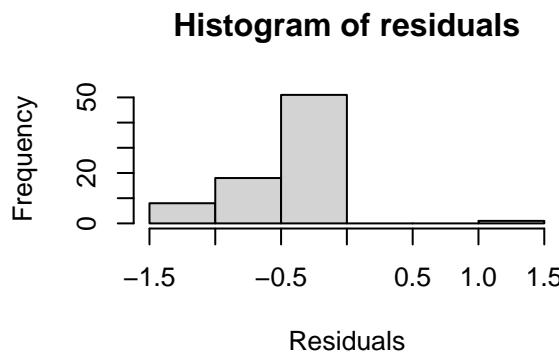
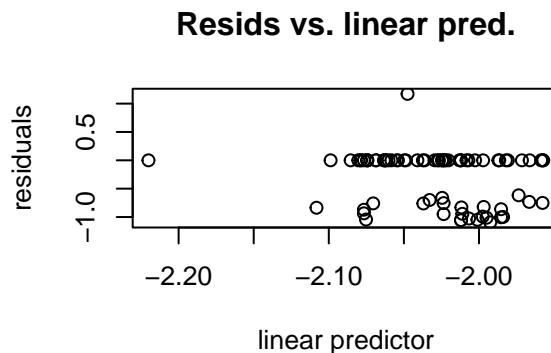
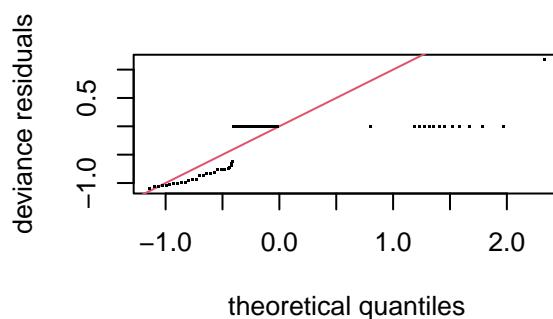
```

```

## Gradient range [-0.0006688737,0.0001181882]
## (score -1401.72 & scale 1).
## Hessian positive definite, eigenvalue range [0.0006681321,19.94783].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    0.66    0.18
## $mfrow
## [1] 2 2
##
## [1] "Thermocyclops inversus eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood

```

```

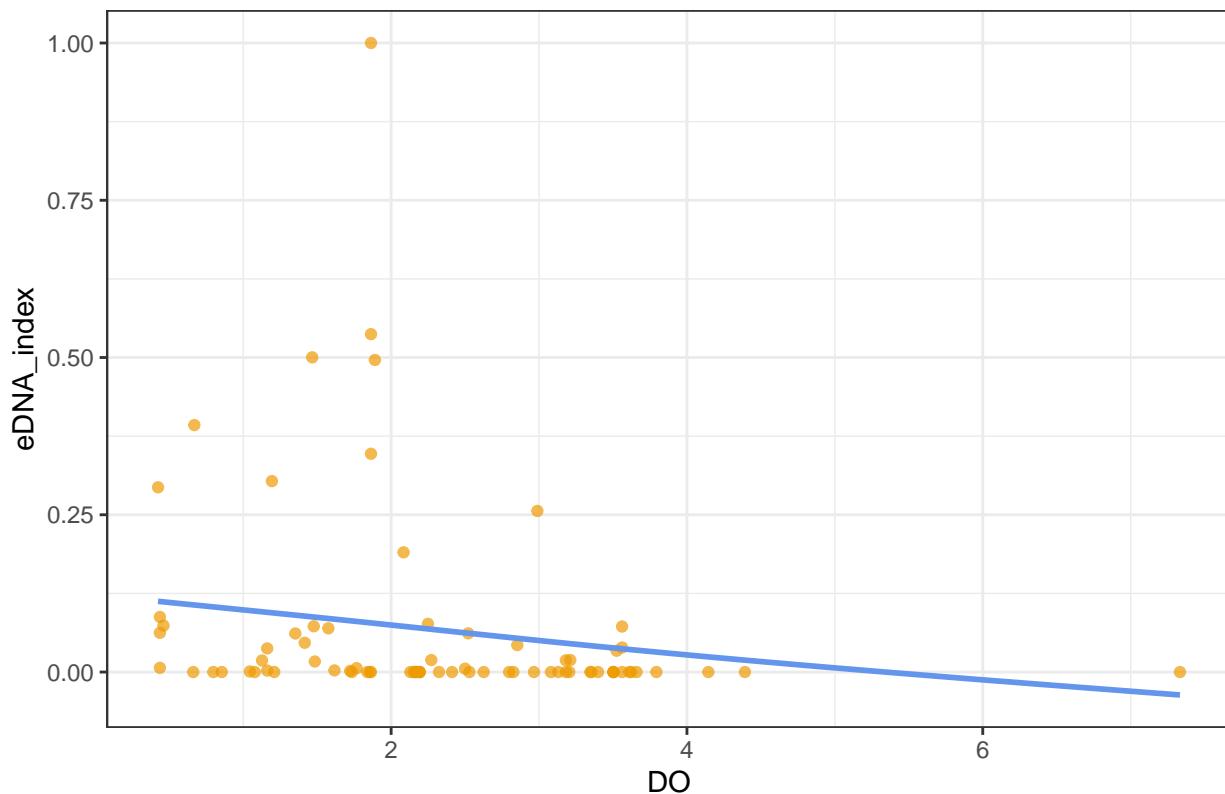
## may be inaccurate
##
## Family: Beta regression(0.386)
## Link function: logit
##
## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.8810    0.1291 -14.56  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.002 1.005  7.37 0.00667 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0385  Deviance explained = -2.28%
## -REML = -1141.7  Scale est. = 1           n = 78
## [1] "GAM quality"
## [1] "AIC: -2281.74143015514"
##      GCV.Cp
## -1141.689

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

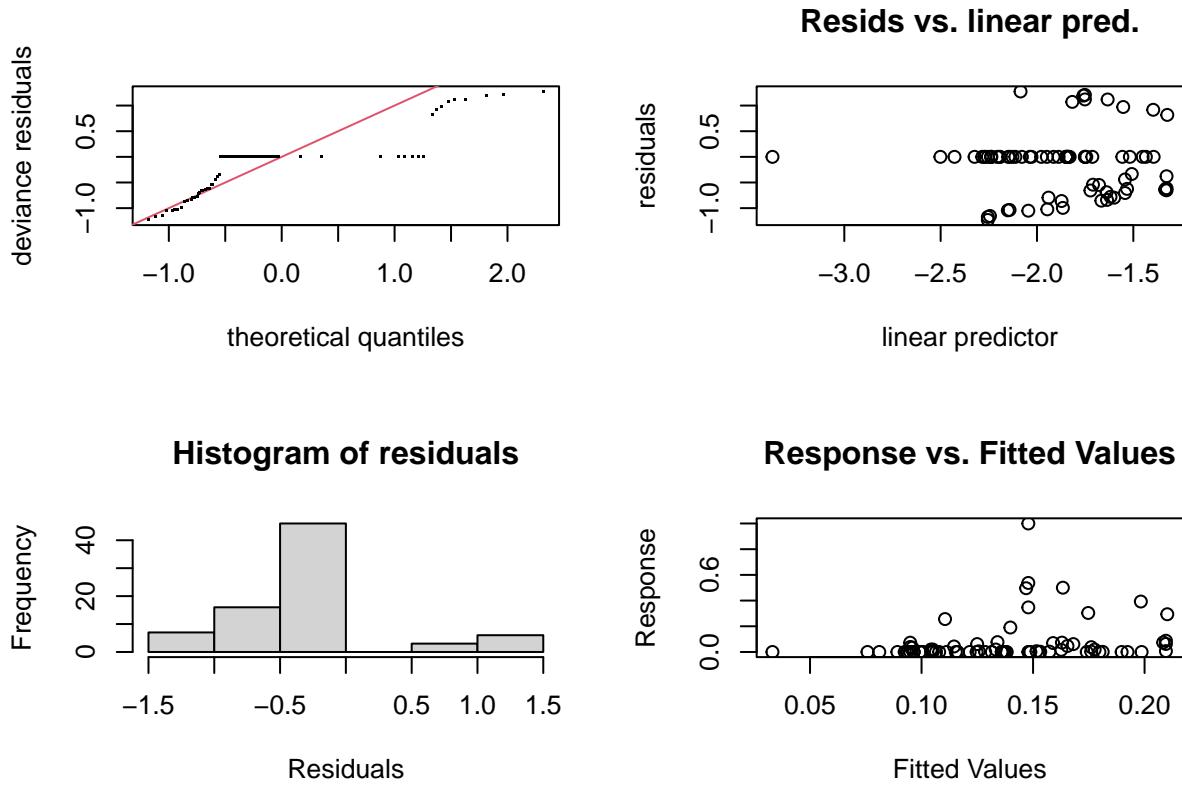
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```

### Thermocyclops inversus eDNA Index vs Dissolved Oxygen



```
##  
## Method: REML    Optimizer: outer newton  
## full convergence after 8 iterations.  
## Gradient range [-0.0002803939,-0.000258935]  
## (score -1141.689 & scale 1).  
## Hessian positive definite, eigenvalue range [0.00028175,24.03699].  
## Model rank = 10 / 10  
  
## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):  
## saturated likelihood may be inaccurate
```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO)  9    1    0.74   0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## $mfrow
## [1] 2 2
##
## [1] "Triconia minuta eDNA Index vs Dissolved Oxygen"
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Warning in family$saturated.ll(y, prior.weights, theta): saturated likelihood
## may be inaccurate
##
## Family: Beta regression(0.295)
## Link function: logit
##

```

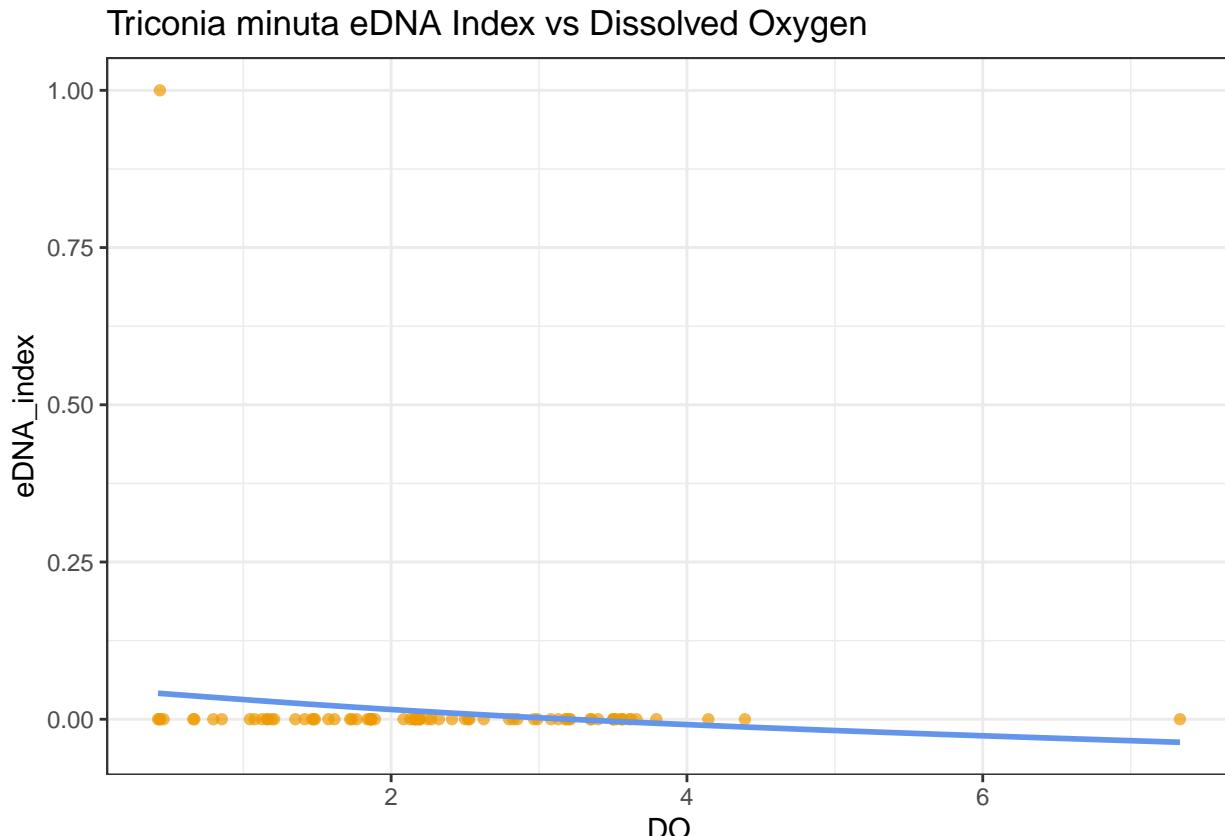
```

## Formula:
## eDNA_index ~ s(DO)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.2012     0.1249 -17.63 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(DO) 1.003 1.006 0.123 0.729
##
## R-sq.(adj) = -0.00119  Deviance explained = -4.22%
## -REML = -2088  Scale est. = 1           n = 78
## [1] "GAM quality"
## [1] "AIC: -4174.57420634434"
##      GCV.Cp
## -2087.999

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)) :
## saturated likelihood may be inaccurate

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)) :
## saturated likelihood may be inaccurate

```



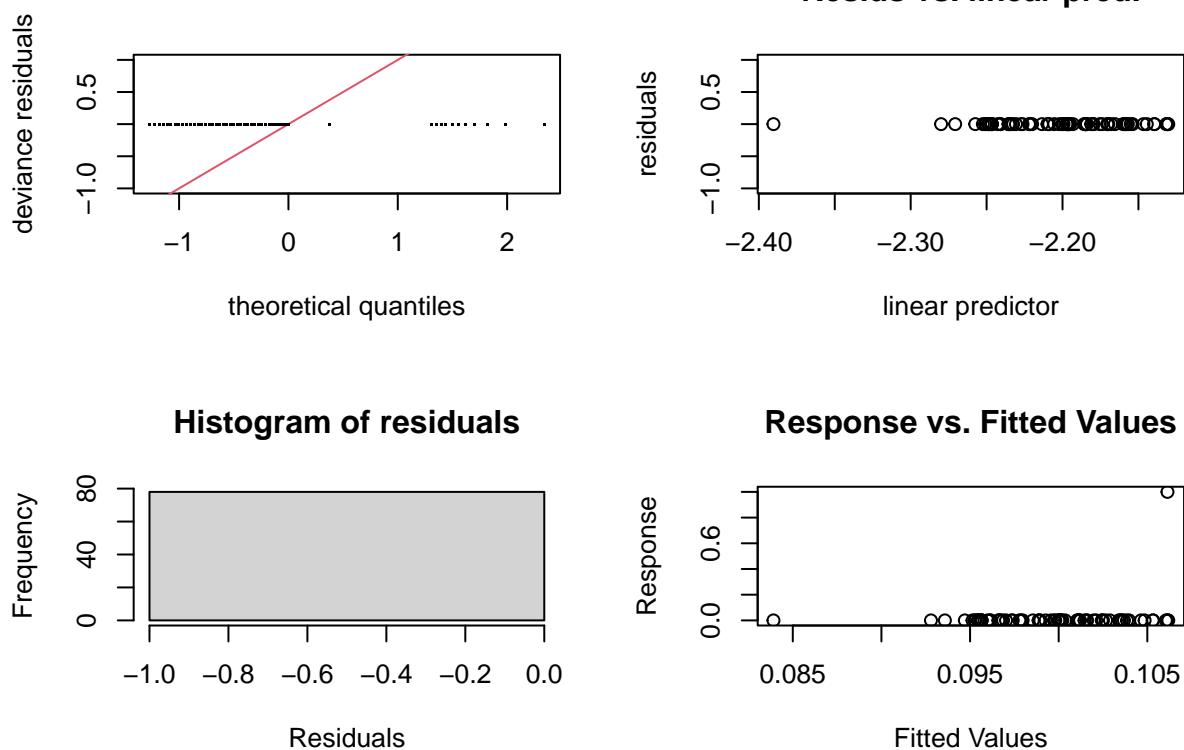
```
##  
## Method: REML    Optimizer: outer newton  
## full convergence after 8 iterations.
```

```

## Gradient range [-0.001166239,0.0001327572]
## (score -2087.999 & scale 1).
## Hessian positive definite, eigenvalue range [0.001164851,16.85319].
## Model rank = 10 / 10

## Warning in object$family$saturated.ll(y, wts, object$family$getTheta(TRUE)):
## saturated likelihood may be inaccurate

```



```

## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##      k'  edf k-index p-value
## s(DO) 9     1    NaN  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## $mfrow
## [1] 2 2

# I think I'm still having some issues w.r.t. the 2023 elimination issue

system("say G A M complete")

```

## Heatmap of DO x Temp x eDNA

```

for (i in 1:length(dfsplit)) { # For each species:
  species <- dfsplit[[i]]$Species[1] # Species name
  title <- paste(species, sep = " ", "eDNA Index vs Dissolved Oxygen & Temperature") #
  ↪ Plot title
  print(title)

```

```

# Plot + save

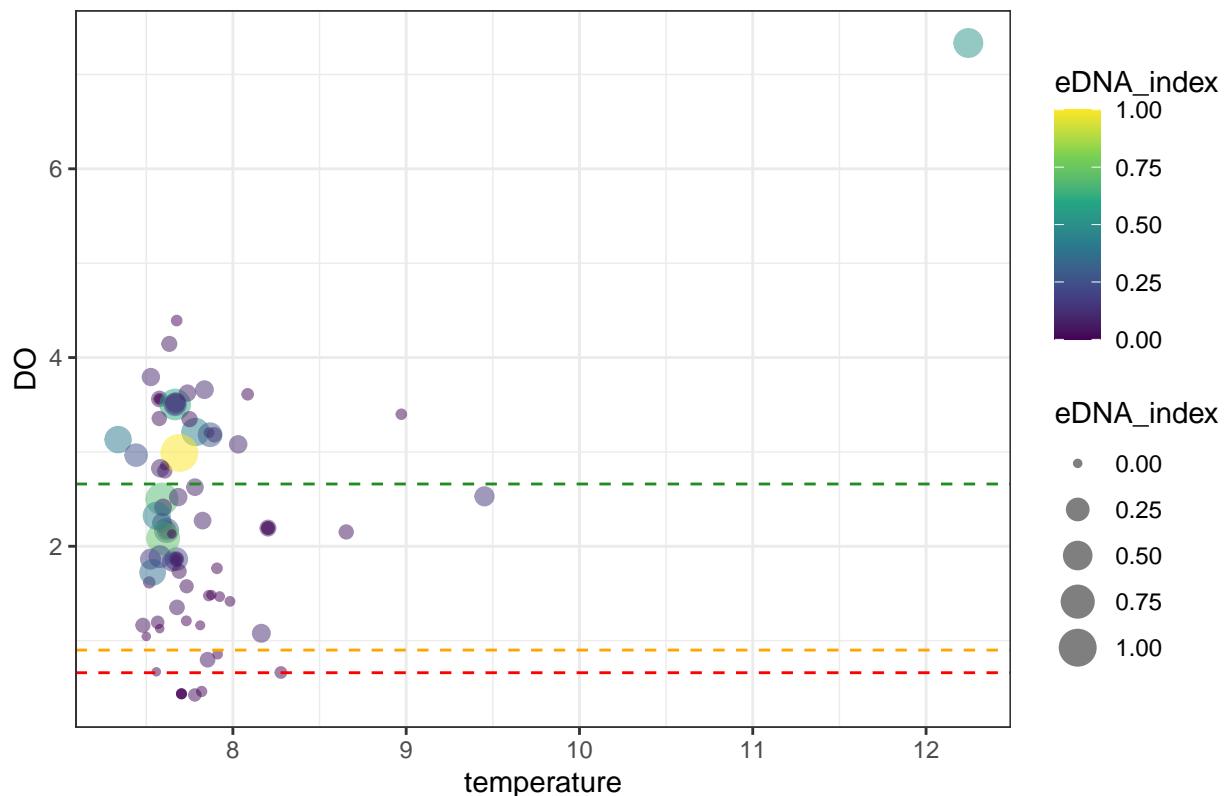
print( # plot this species
  ggplot(dfsplit[[i]], aes(x = temperature,
                            y = DO,
                            size = eDNA_index, color = eDNA_index)) +
  scale_color_continuous(type = "viridis") +
  geom_point(alpha = 0.5) +
  ggtitle(title) +
  geom_hline(yintercept = 0.66, linetype = 2, color = "red") + # most copepods die
  geom_hline(yintercept = 0.9, linetype = 2, color = "orange") + # ~50% of copepods
  → die
  geom_hline(yintercept = 2.66, linetype = 2, color = "forestgreen") + # many
  → copepods experience sublethal effects (e.g. less egg production)
  theme_bw()
)

ggsave(filename = paste(species, sep = "_", "eDNA_DO_Temp_ScatterSize.png"),
       path = here("eDNA_Index_Hypoxia", "Plots", "DO_Temp_Scatter"),
       width = 2500, height = 2000, units = "px")
}

## [1] "Acartia longiremis eDNA Index vs Dissolved Oxygen & Temperature"
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

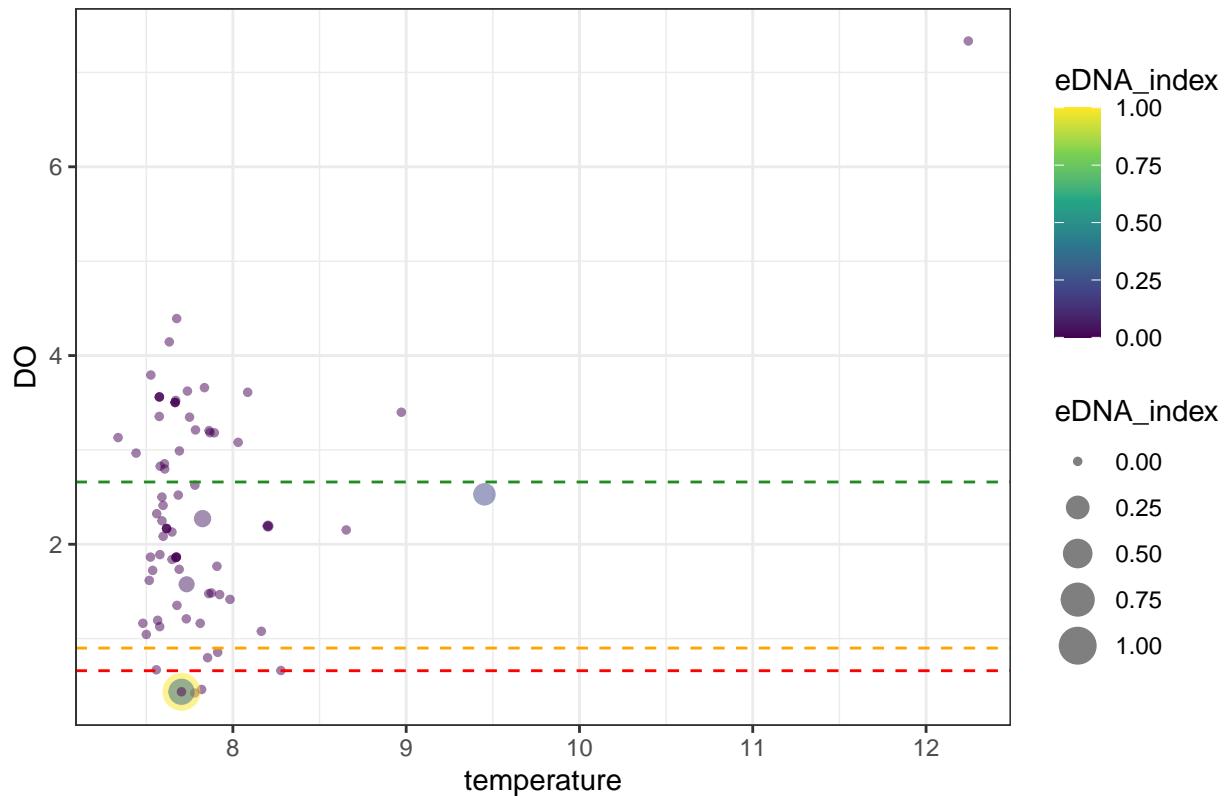
```

### Acartia longiremis eDNA Index vs Dissolved Oxygen & Temperature



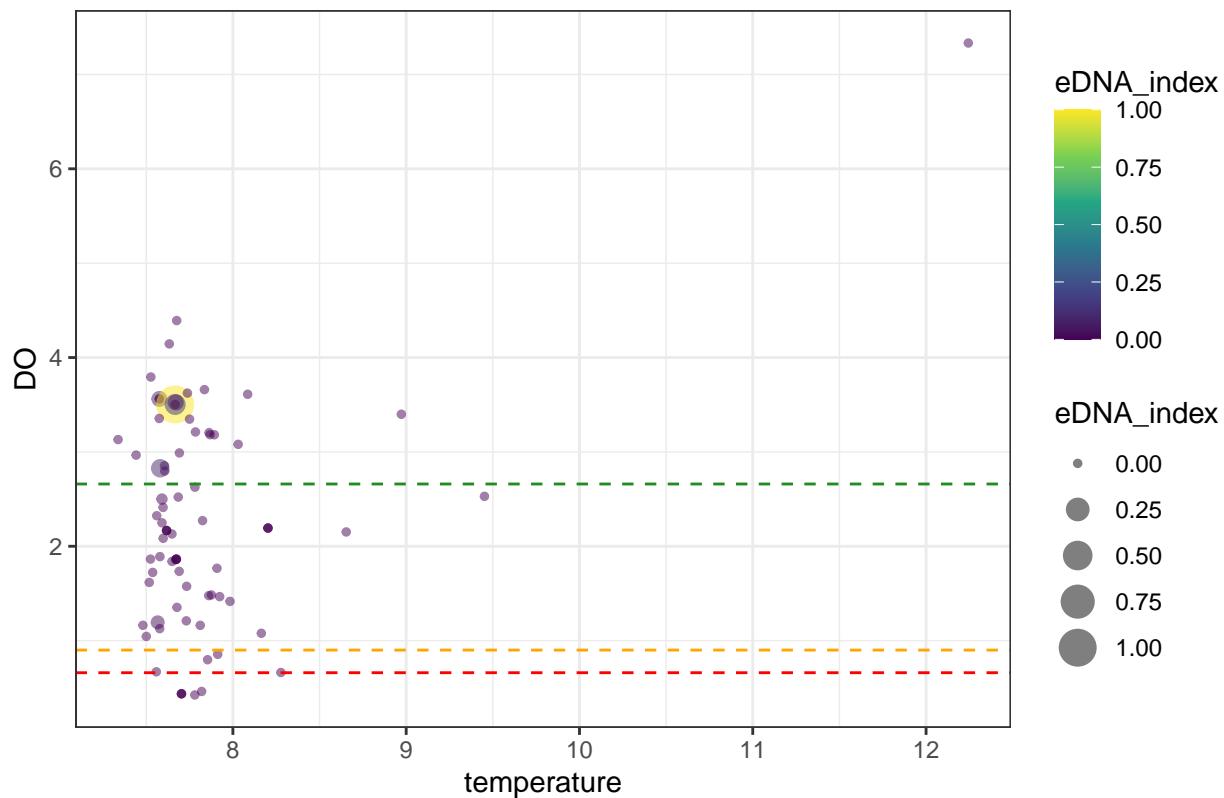
```
## [1] "Calanus pacificus eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Calanus pacificus eDNA Index vs Dissolved Oxygen & Temperature



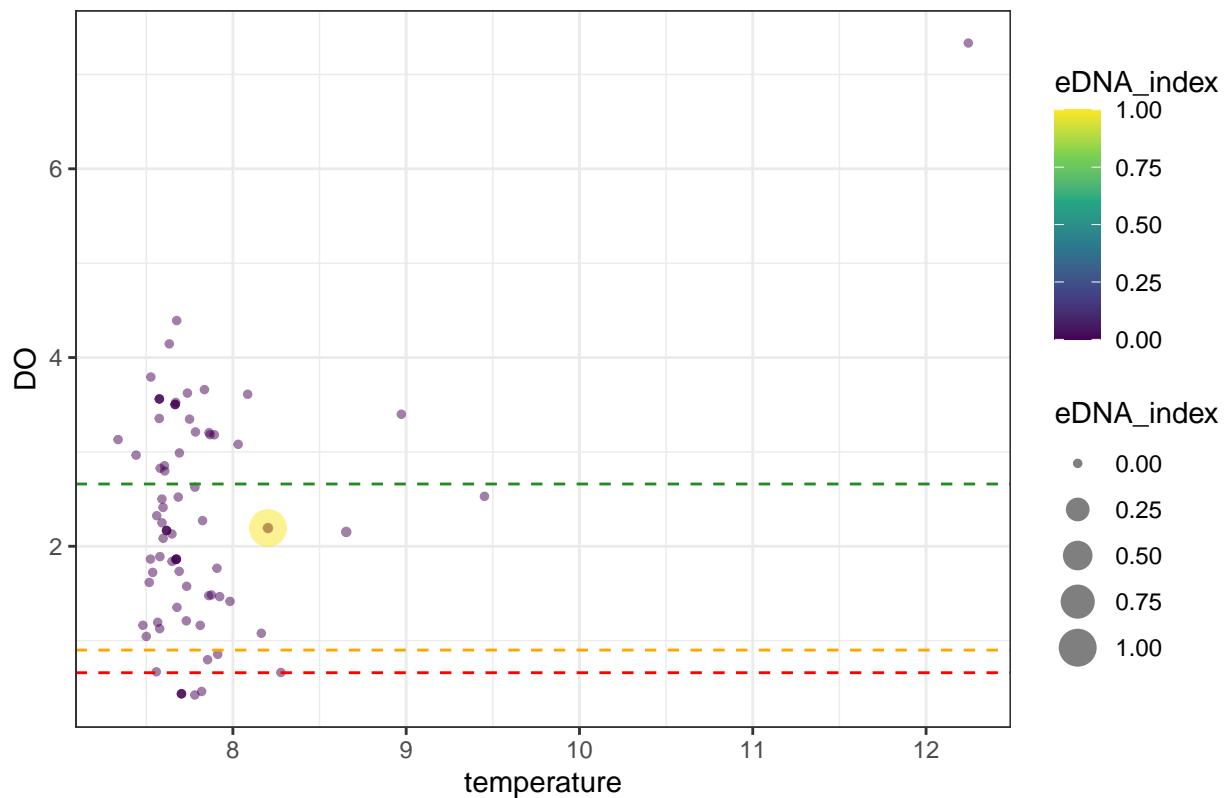
```
## [1] "Centropages abdominalis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Centropages abdominalis eDNA Index vs Dissolved Oxygen & Temperature



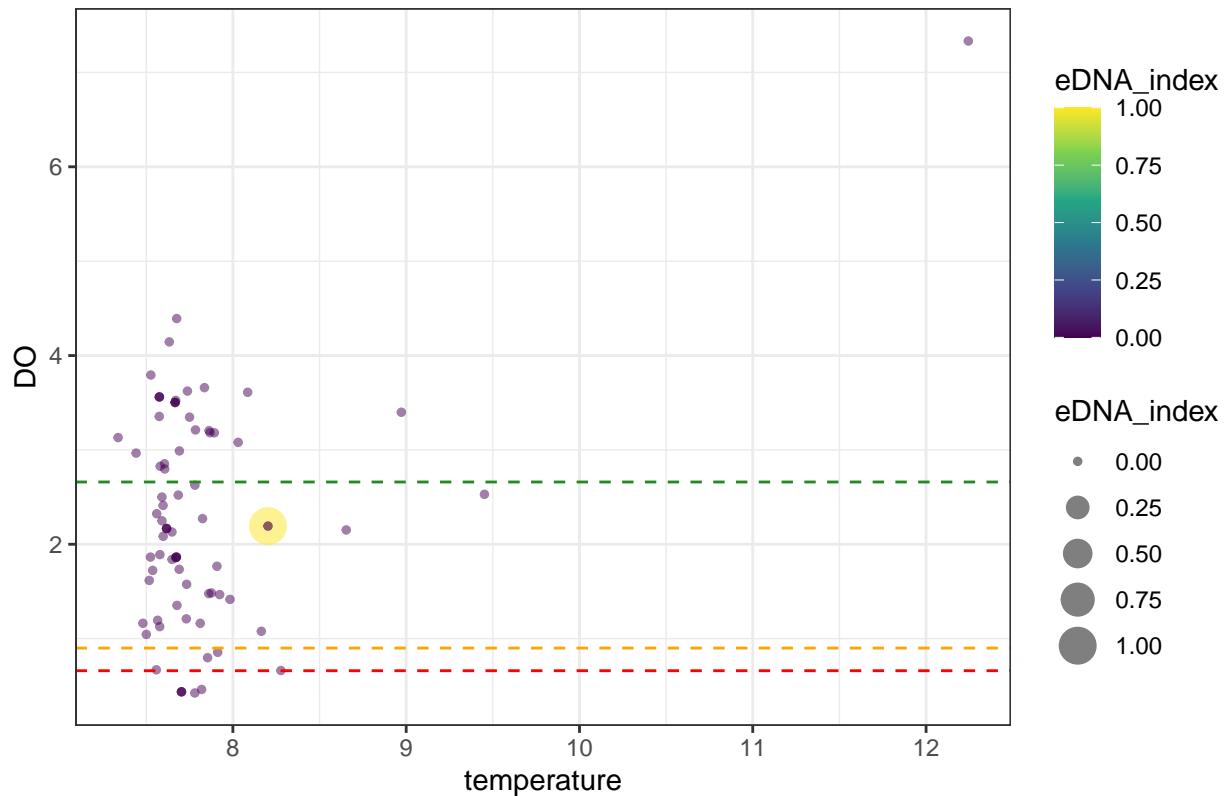
```
## [1] "Clausocalanus parapergens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Clausocalanus parapergens eDNA Index vs Dissolved Oxygen & Temperature



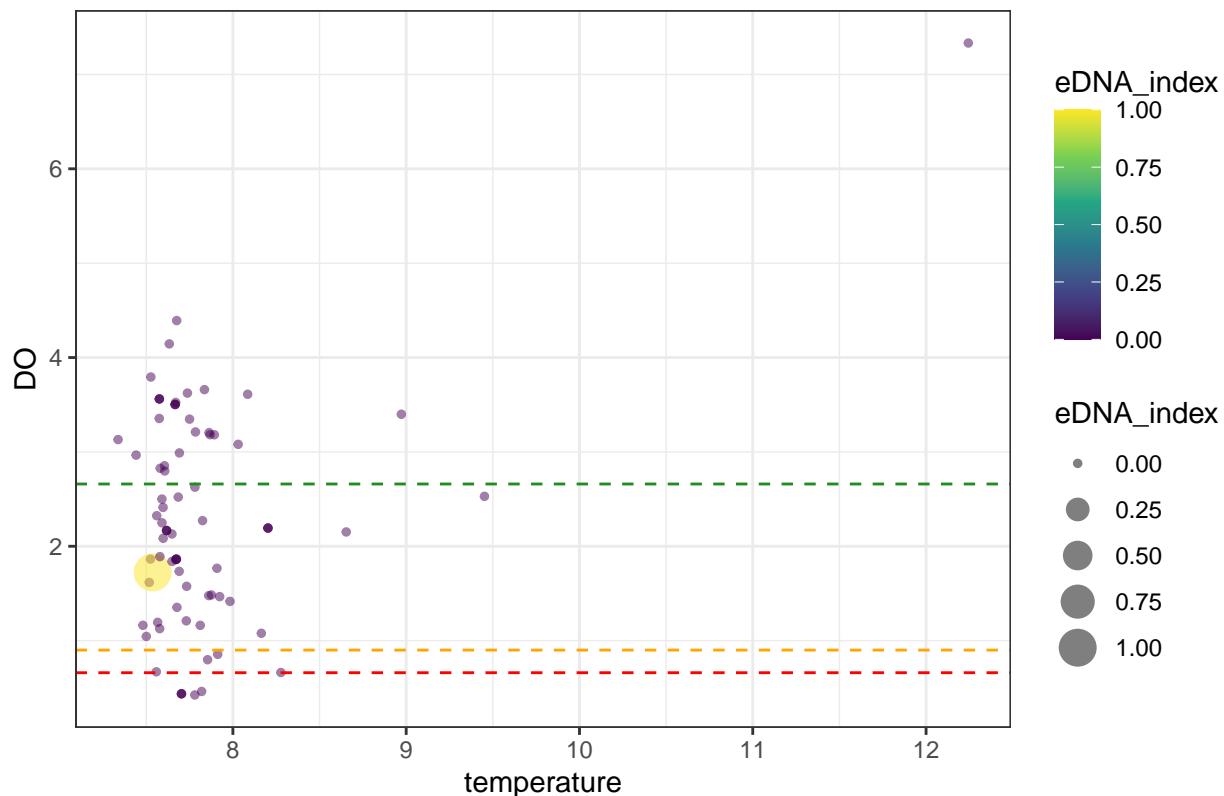
```
## [1] "Clausocalanus pergens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Clausocalanus pergens eDNA Index vs Dissolved Oxygen & Temperature



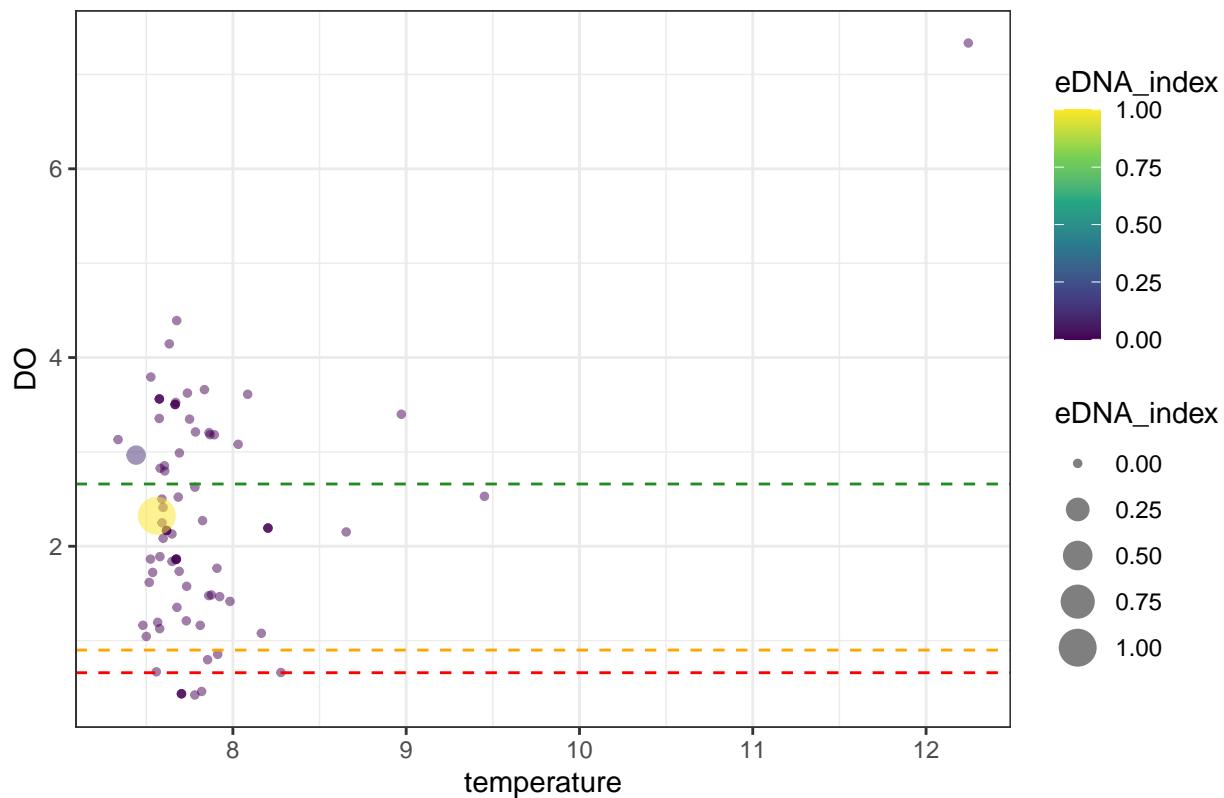
```
## [1] "Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen & Temperature



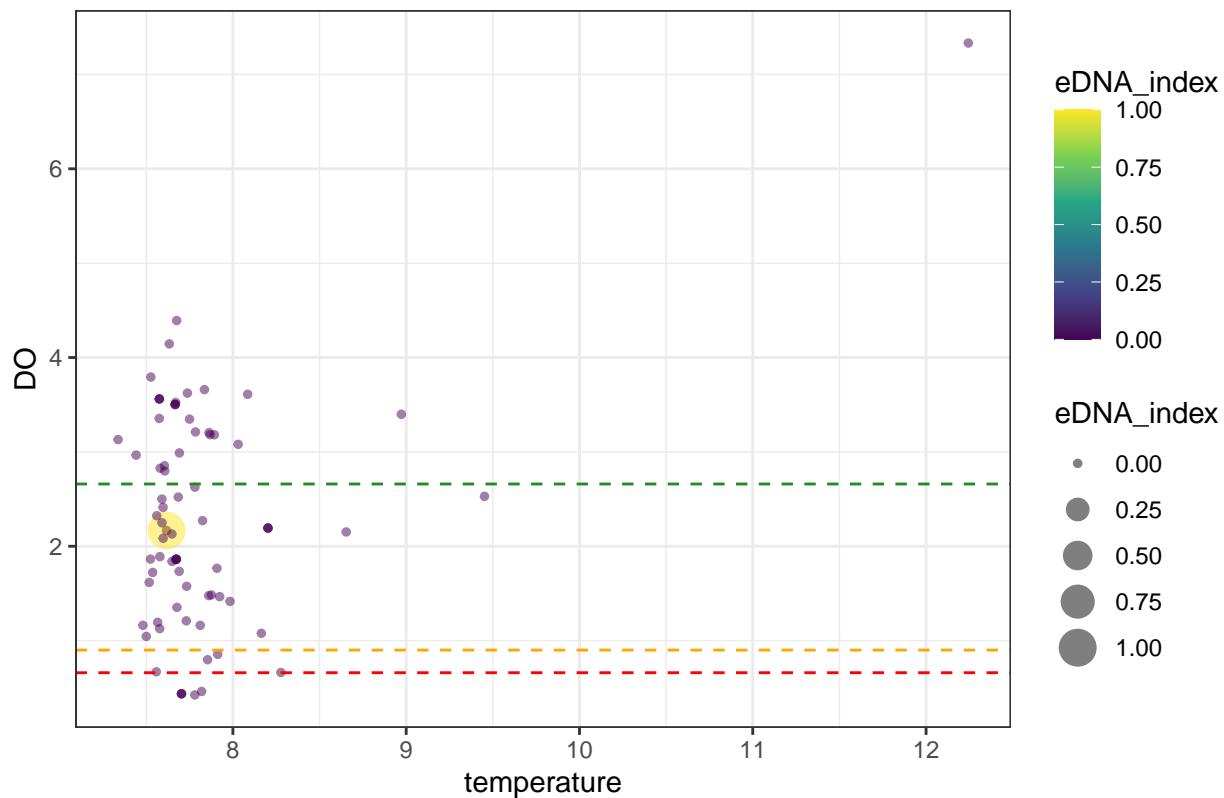
```
## [1] "Lucicutia flavigornis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Lucicutia flavigornis eDNA Index vs Dissolved Oxygen & Temperature



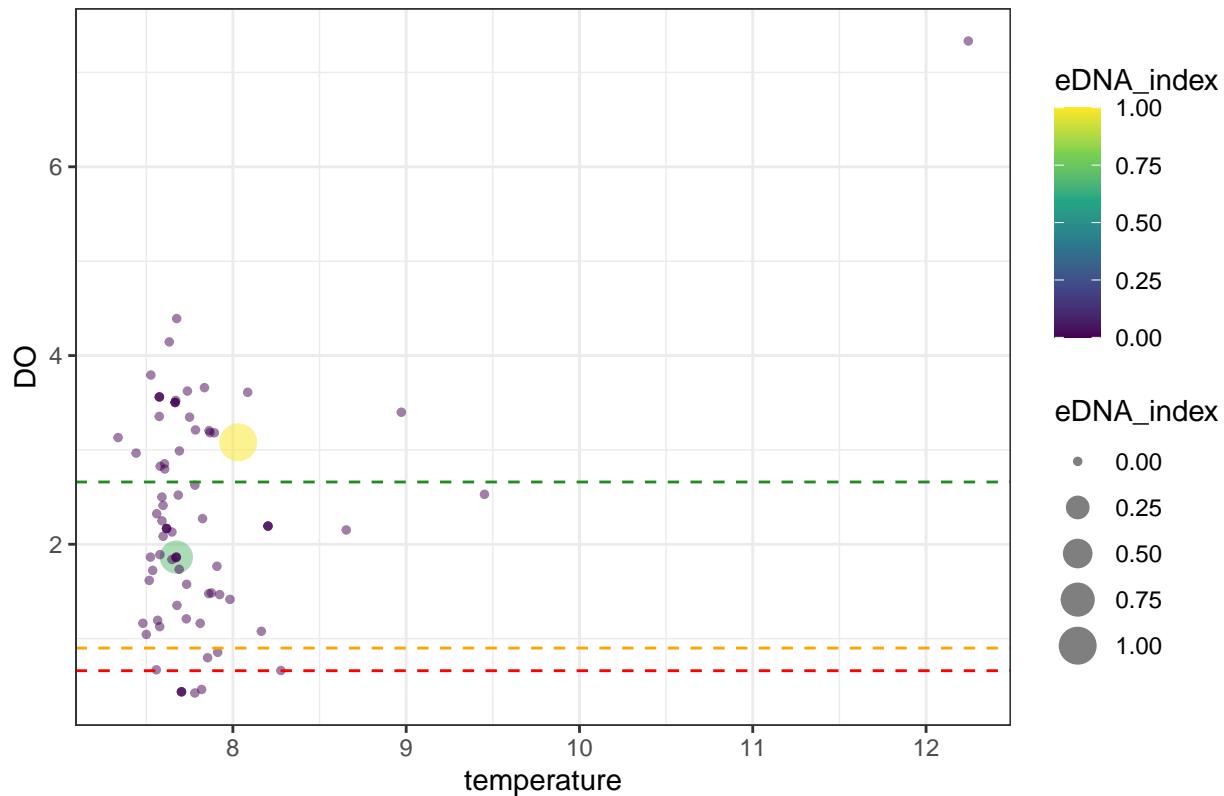
```
## [1] "Metridia lucens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Metridia lucens eDNA Index vs Dissolved Oxygen & Temperature



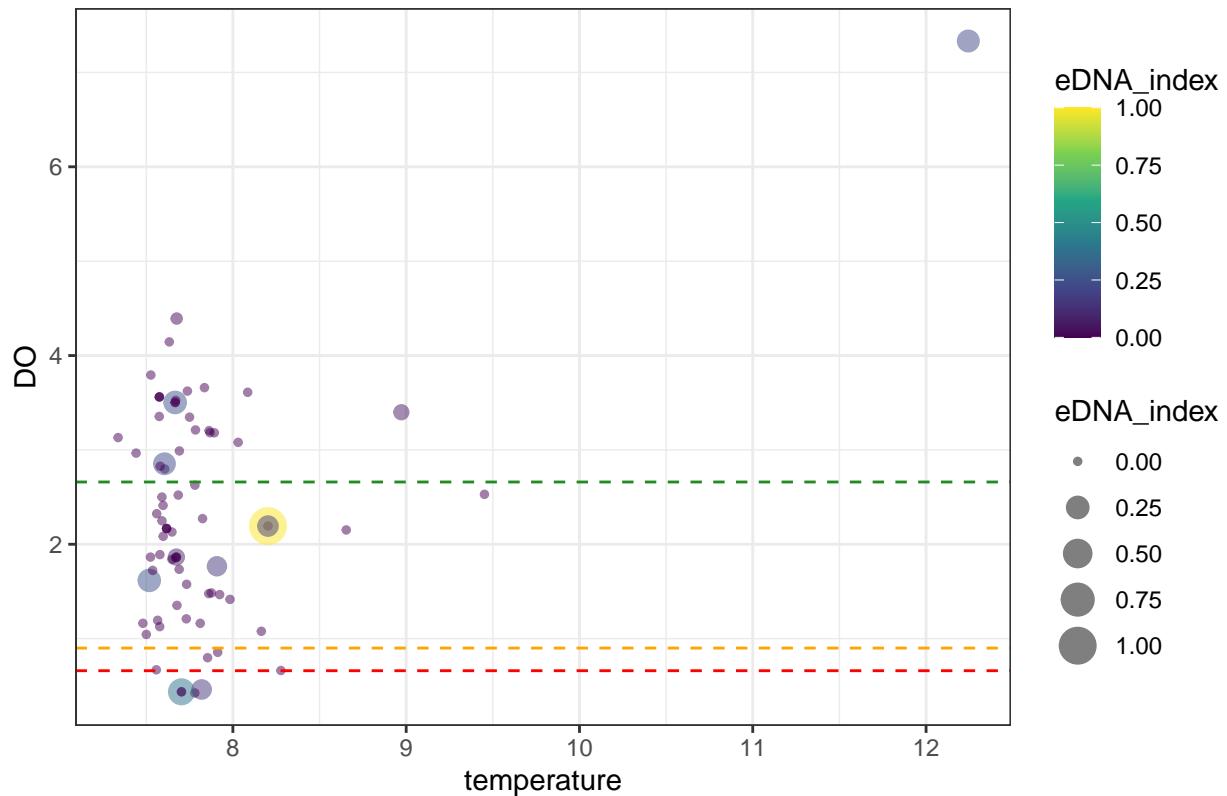
```
## [1] "Metridia pacifica eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Metridia pacifica eDNA Index vs Dissolved Oxygen & Temperature



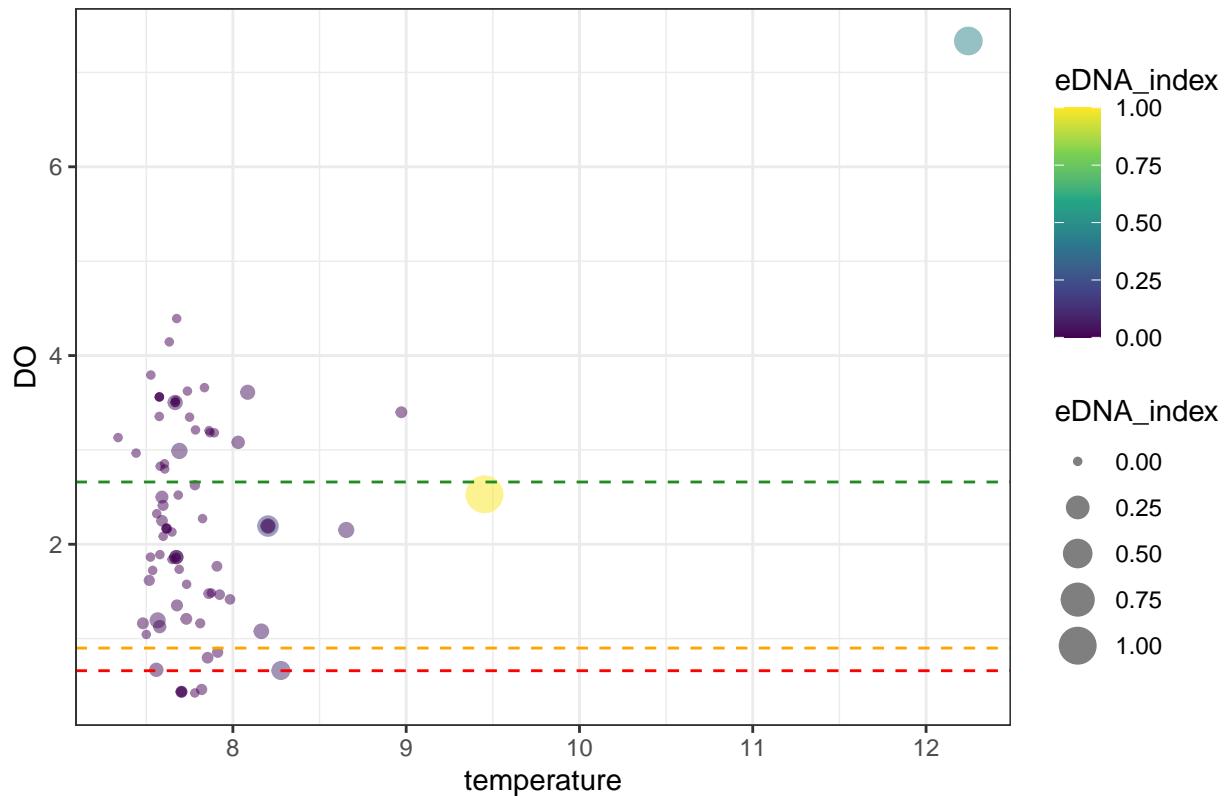
```
## [1] "Oithona similis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Oithona similis eDNA Index vs Dissolved Oxygen & Temperature



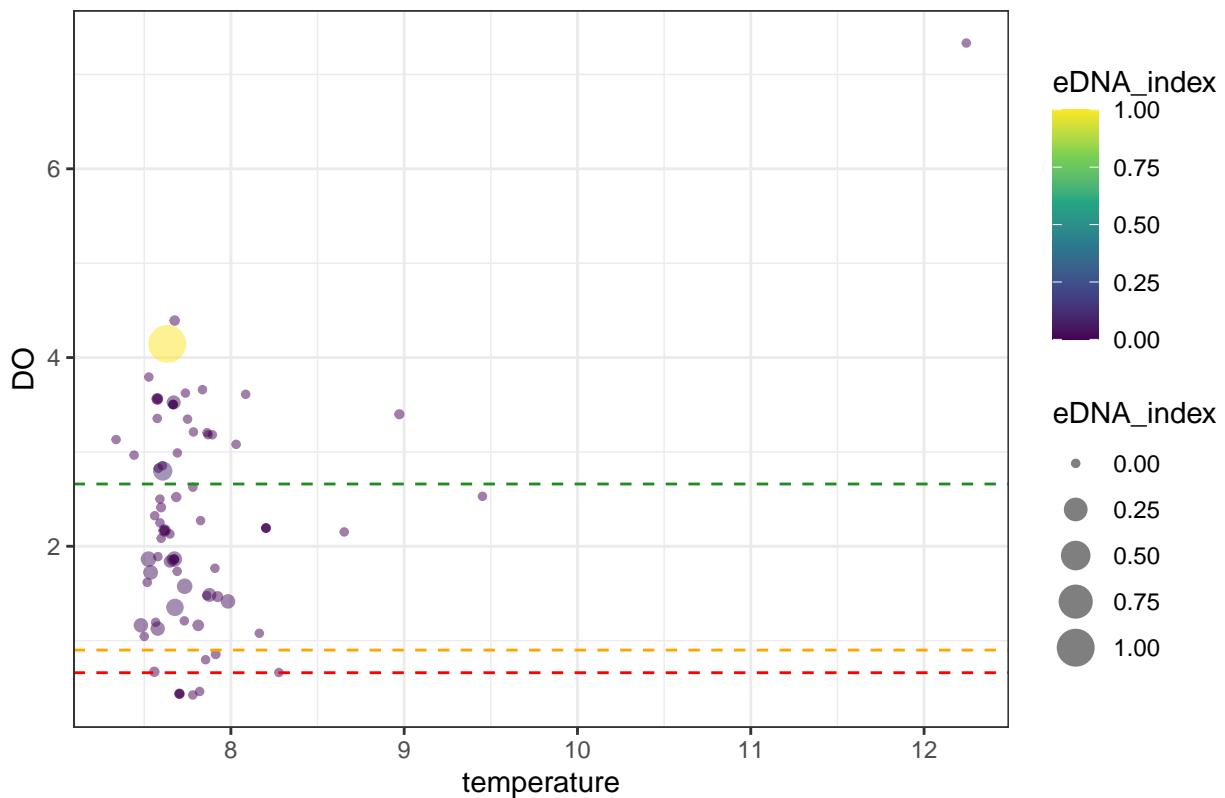
```
## [1] "Paracalanus sp. C AC-2013 eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Paracalanus sp. C AC–2013 eDNA Index vs Dissolved Oxygen & Temperature



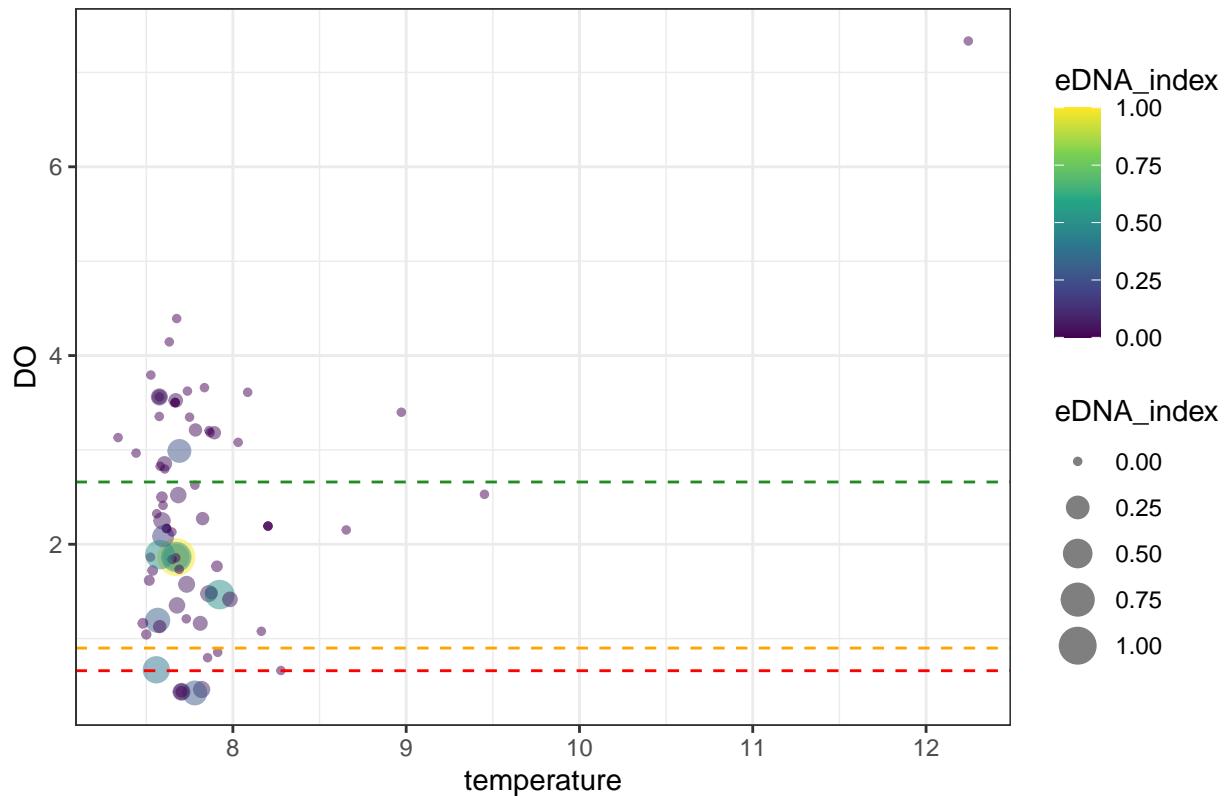
```
## [1] "Pseudocalanus newmani eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Pseudocalanus newmani eDNA Index vs Dissolved Oxygen & Temperature



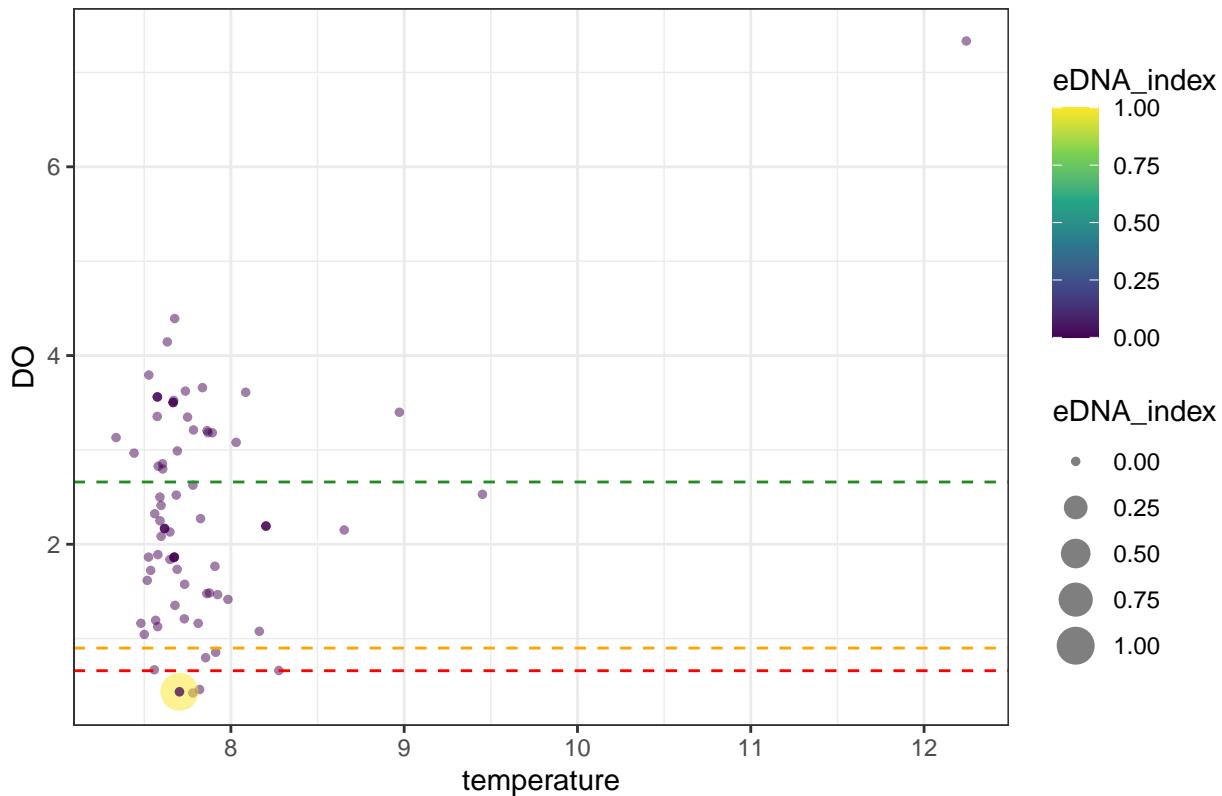
```
## [1] "Thermocyclops inversus eDNA Index vs Dissolved Oxygen & Temperature"
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```

## Thermocyclops inversus eDNA Index vs Dissolved Oxygen & Temperature



```
## [1] "Triconia minuta eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Triconia minuta eDNA Index vs Dissolved Oxygen & Temperature



```

for (i in 1:length(dfsplit)) { # For each species:
  species <- dfsplit[[i]]$Species[1] # Species name
  title <- paste(species, sep = " ", "eDNA Index vs Dissolved Oxygen & Temperature") #
  ↪ Plot title
  print(title)

# Plot + save

  print( # plot this species
    ggplot(dfsplit[[i]], aes(x = temperature,
                               y = DO,
                               color = eDNA_index)) +
    scale_color_continuous(type = "viridis") +
    geom_point(alpha = 0.7) +
    ggttitle(title) +
    geom_hline(yintercept = 0.66, linetype = 2, color = "red") + # most copepods die
    geom_hline(yintercept = 0.9, linetype = 2, color = "orange") + # ~50% of copepods
    ↪ die
    geom_hline(yintercept = 2.66, linetype = 2, color = "forestgreen") + # many
    ↪ copepods experience sublethal effects (e.g. less egg production)
    theme_bw()
  )

  ggsave(filename = paste(species, sep = "_", "eDNA_DO_Temp_Scatter.png"),
         path = here("eDNA_Index_Hypoxia", "Plots", "DO_Temp_Scatter"),
         width = 2500, height = 2000, units = "px")
}

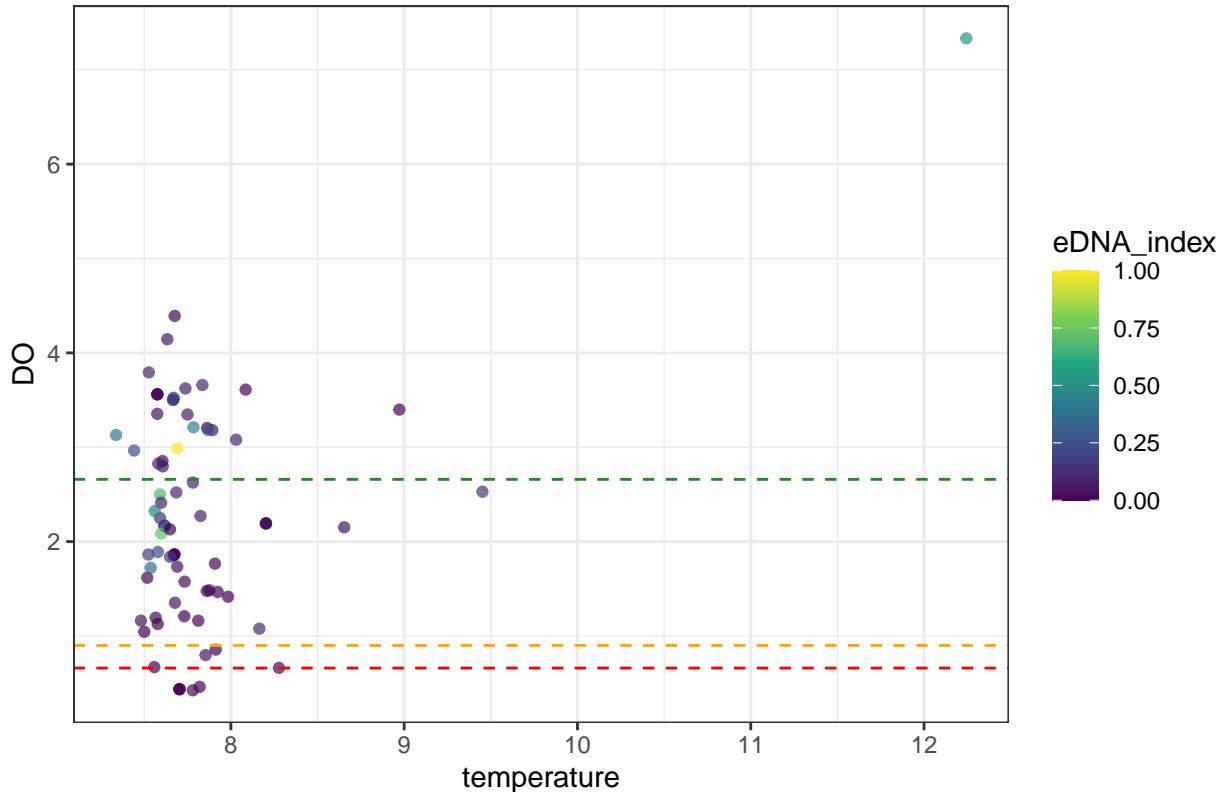
```

```

## [1] "Acartia longiremis eDNA Index vs Dissolved Oxygen & Temperature"
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

```

### Acartia longiremis eDNA Index vs Dissolved Oxygen & Temperature

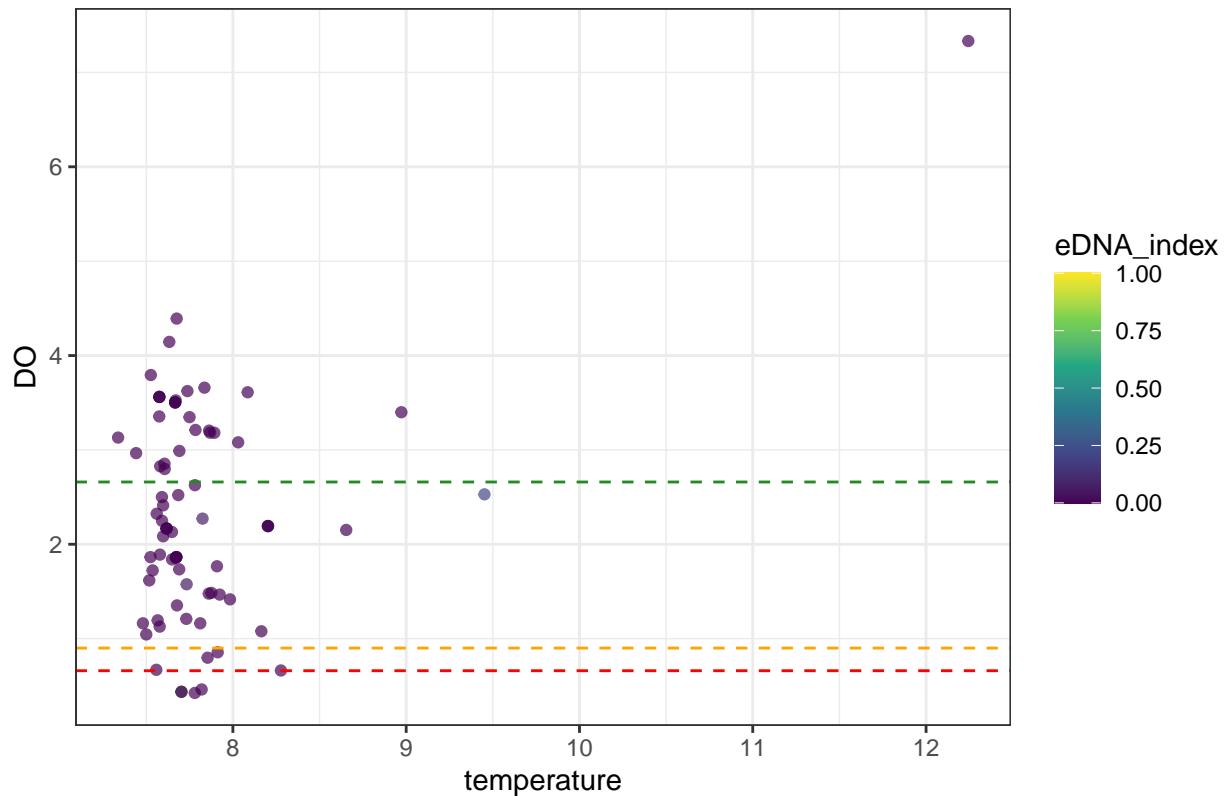


```

## [1] "Calanus pacificus eDNA Index vs Dissolved Oxygen & Temperature"
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

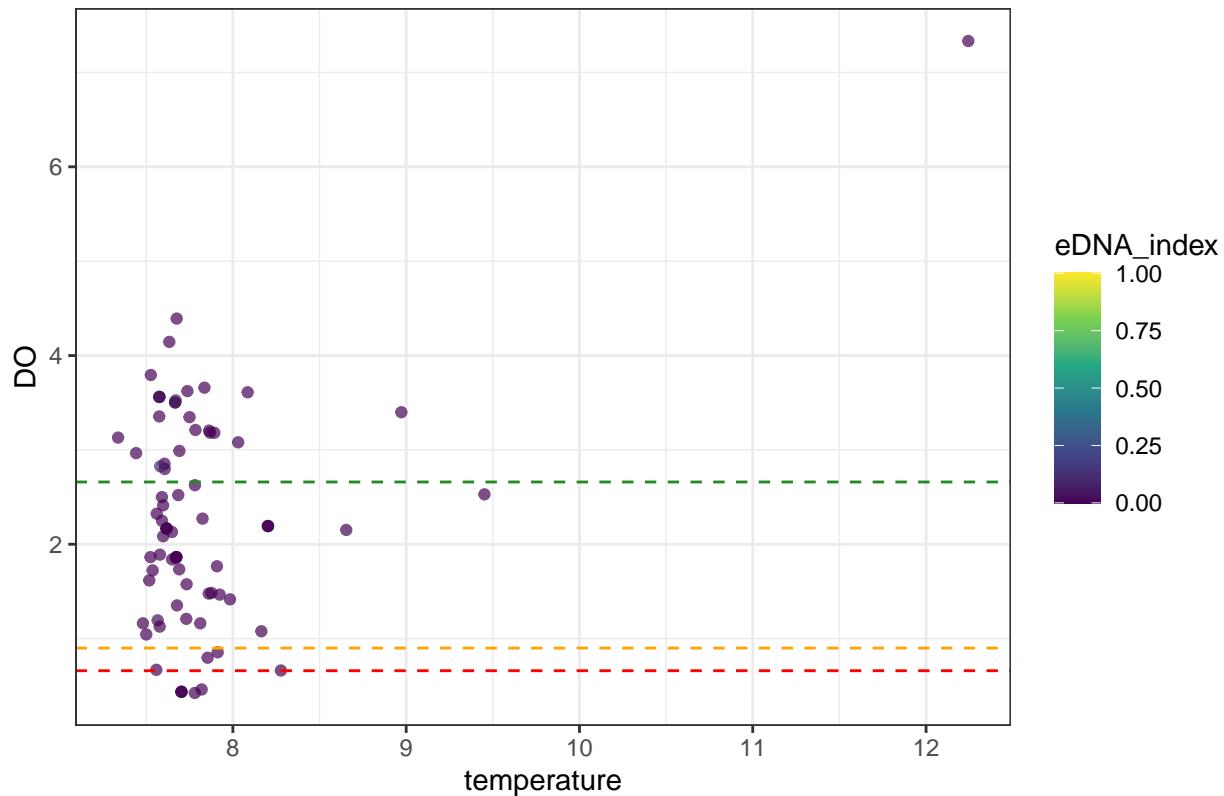
```

## Calanus pacificus eDNA Index vs Dissolved Oxygen & Temperature



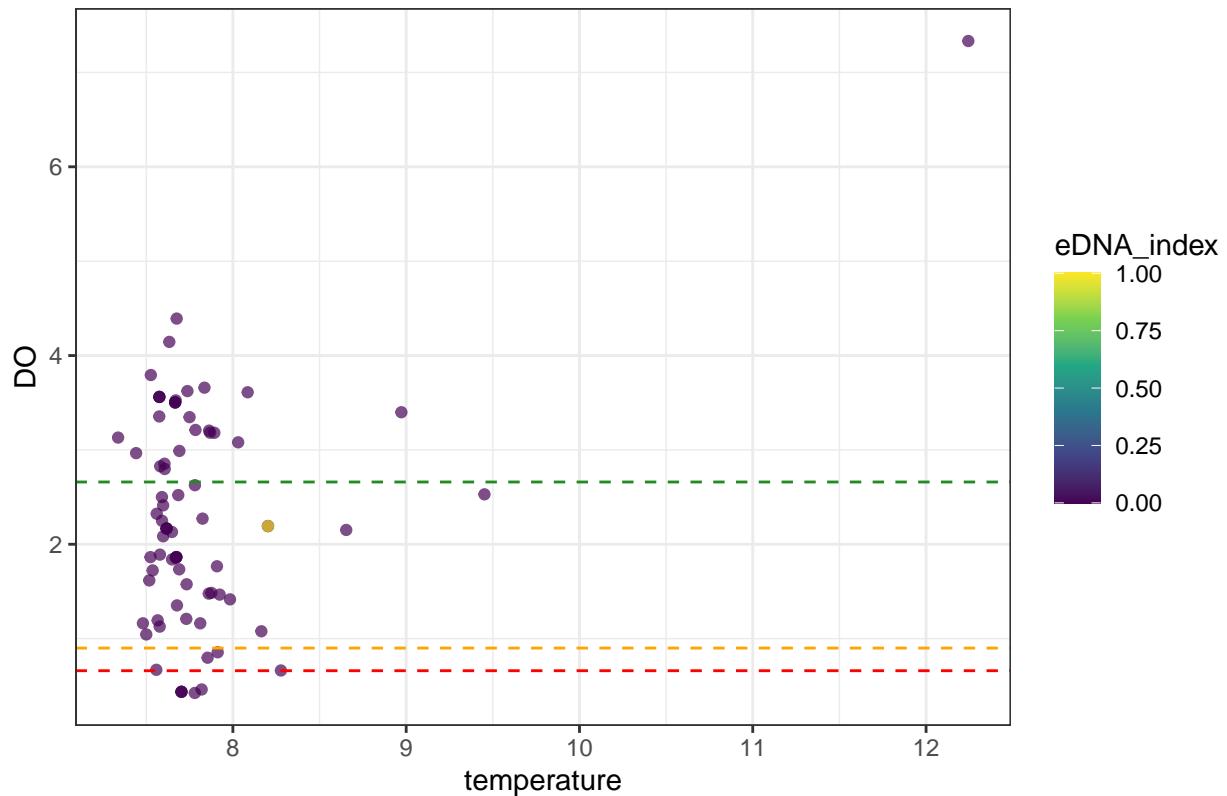
```
## [1] "Centropages abdominalis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Centropages abdominalis eDNA Index vs Dissolved Oxygen & Temperature



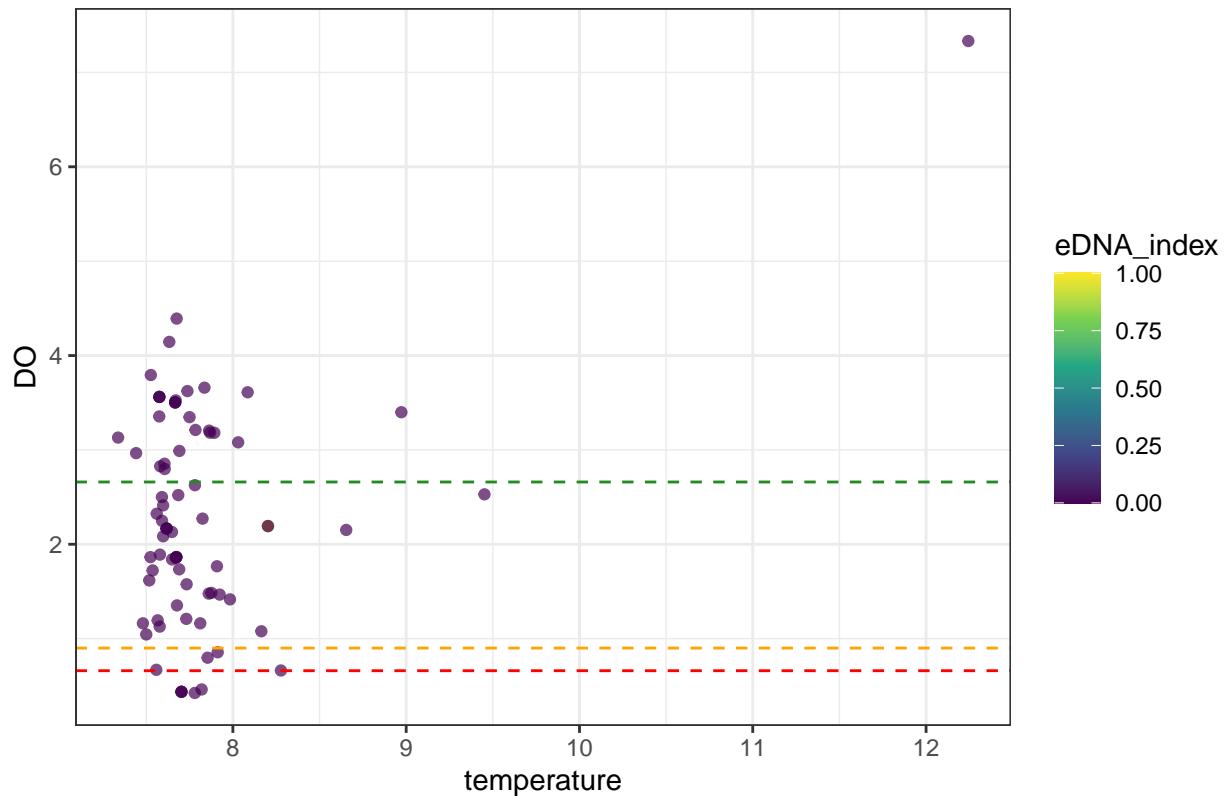
```
## [1] "Clausocalanus parapergens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Clausocalanus parapergens eDNA Index vs Dissolved Oxygen & Temperature



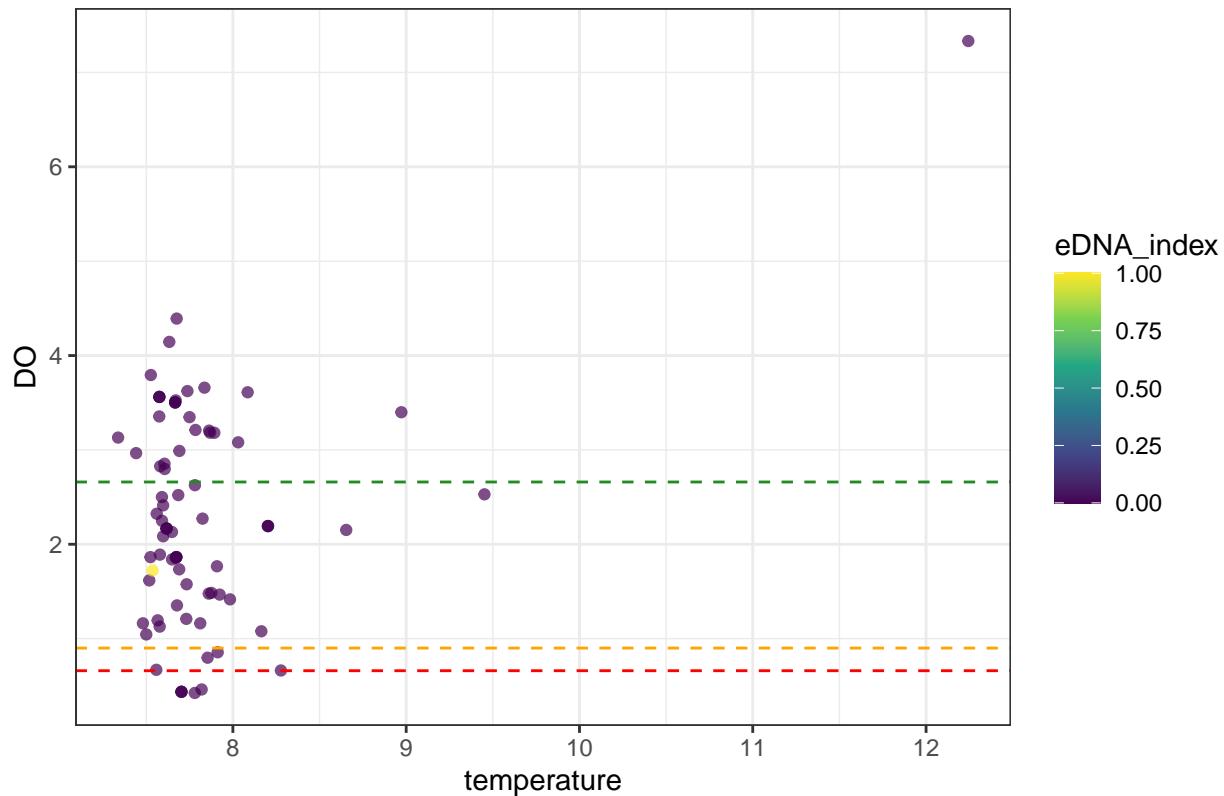
```
## [1] "Clausocalanus pergens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### Clausocalanus pergens eDNA Index vs Dissolved Oxygen & Temperature



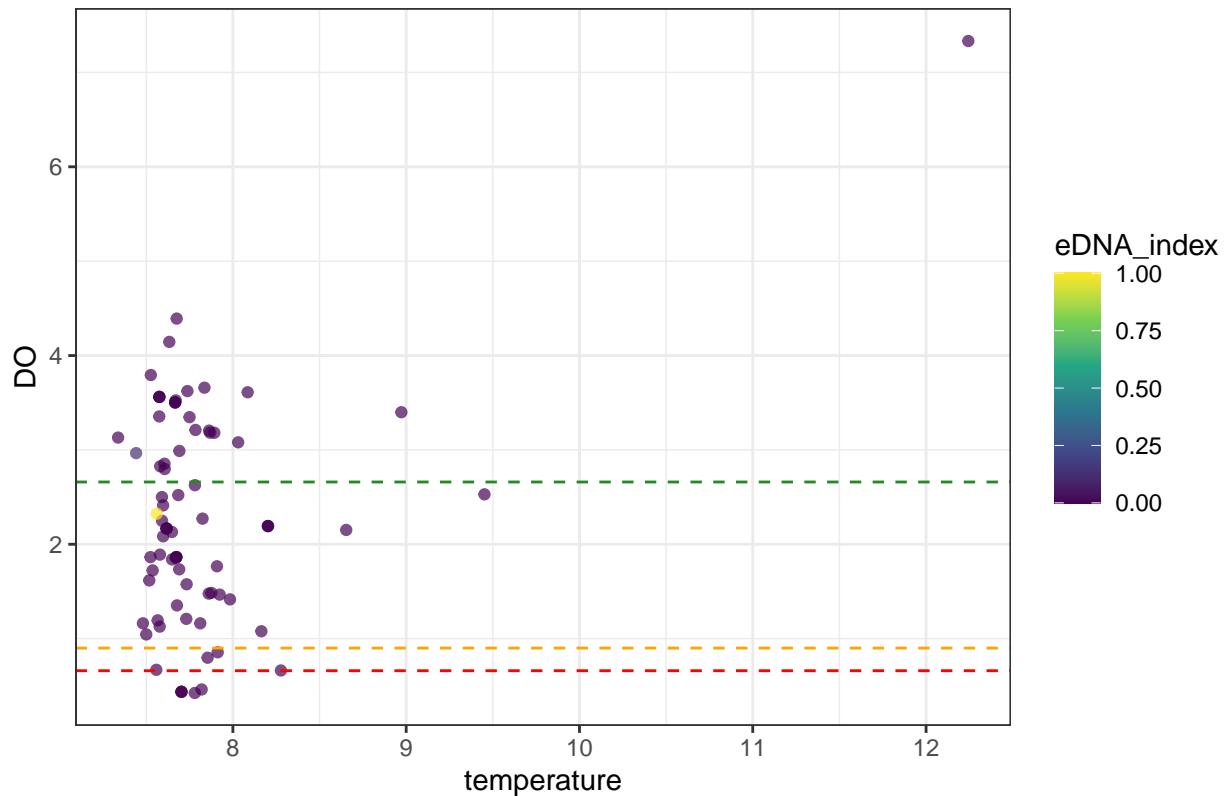
```
## [1] "Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### *Diacyclops incolotaenia* eDNA Index vs Dissolved Oxygen & Temperature



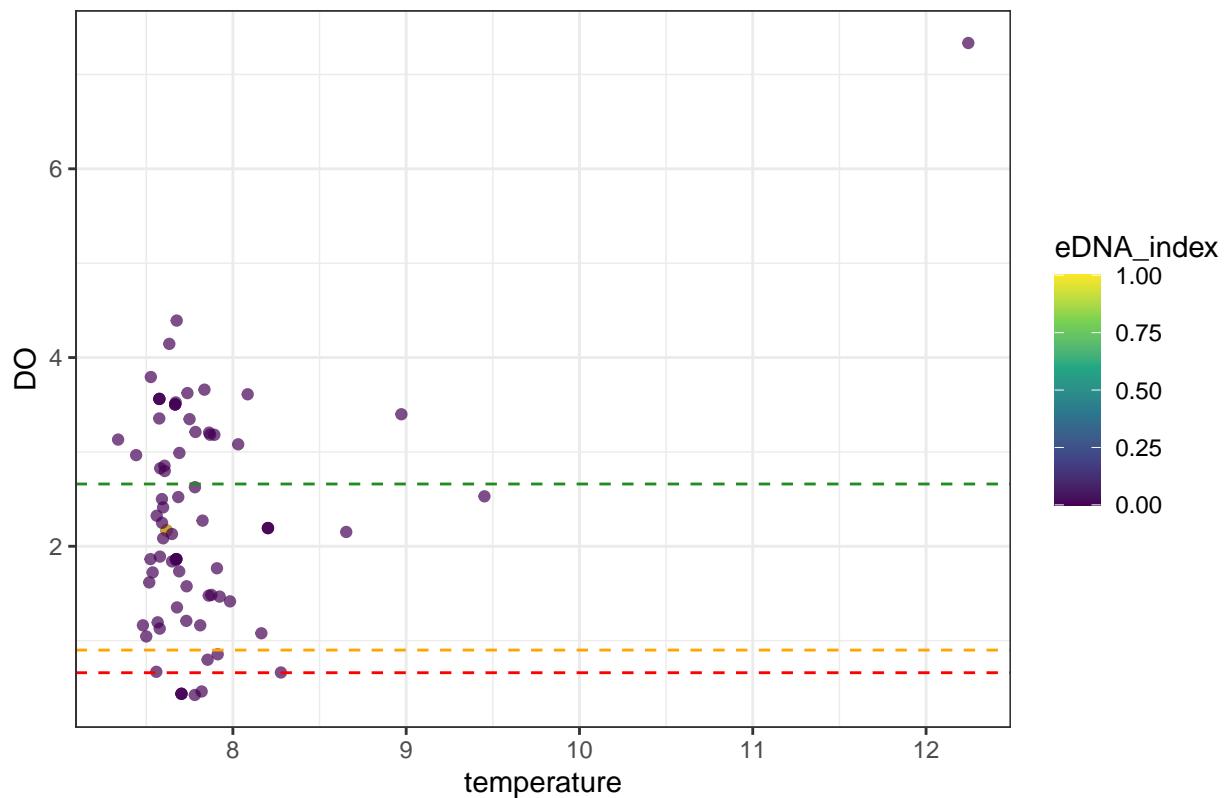
```
## [1] "Lucicutia flavicornis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Lucicutia flavigornis eDNA Index vs Dissolved Oxygen & Temperature



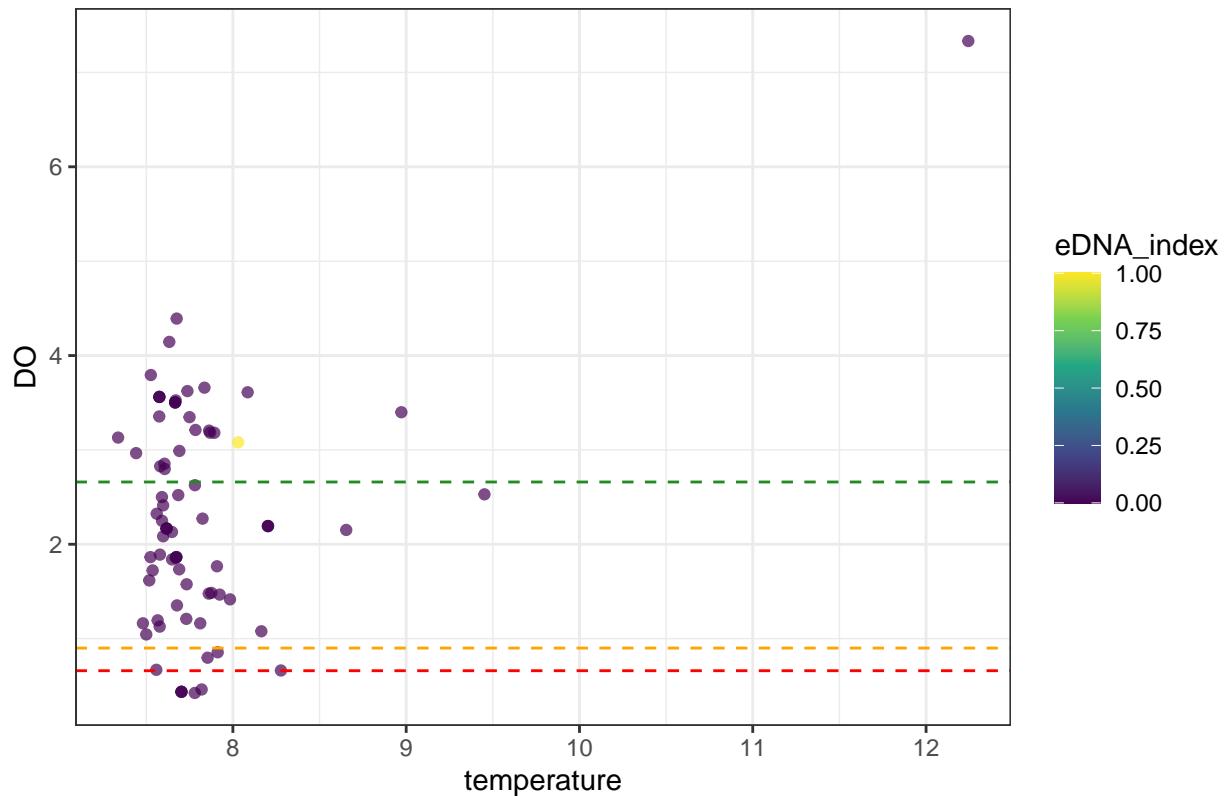
```
## [1] "Metridia lucens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Metridia lucens eDNA Index vs Dissolved Oxygen & Temperature



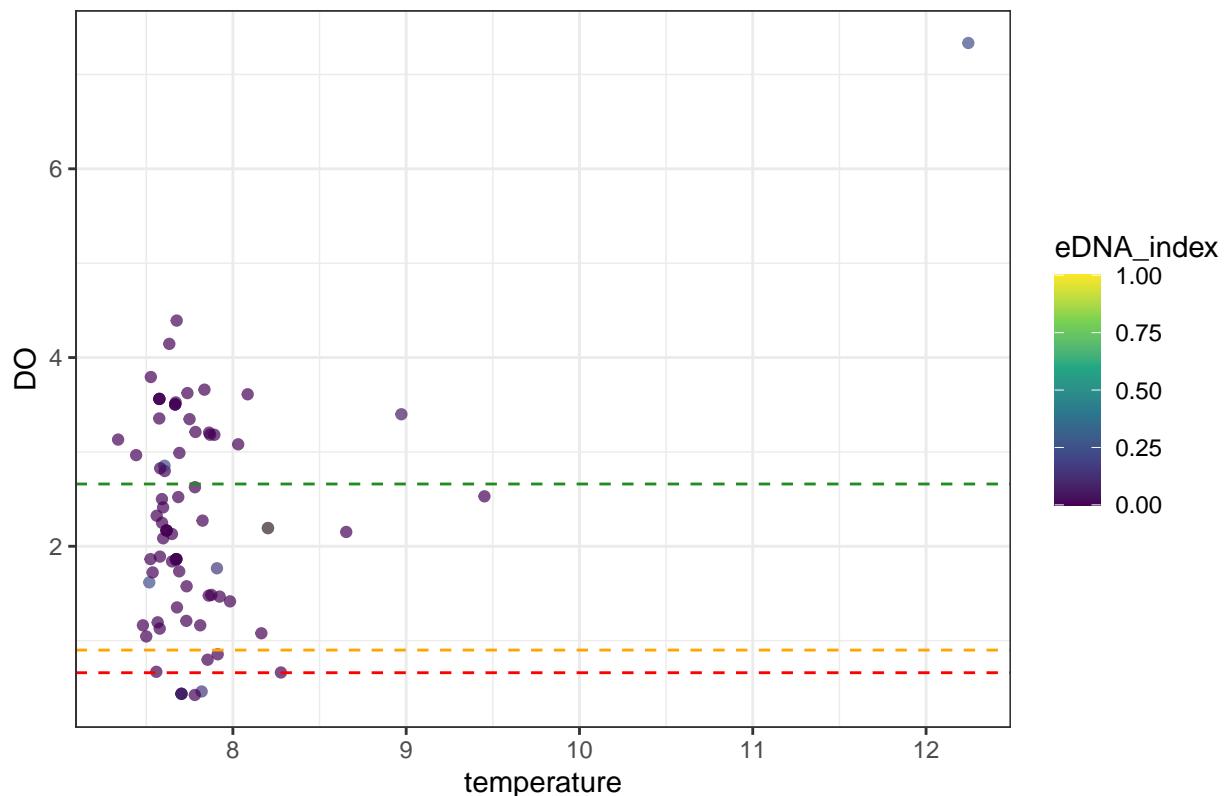
```
## [1] "Metridia pacifica eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Metridia pacifica eDNA Index vs Dissolved Oxygen & Temperature



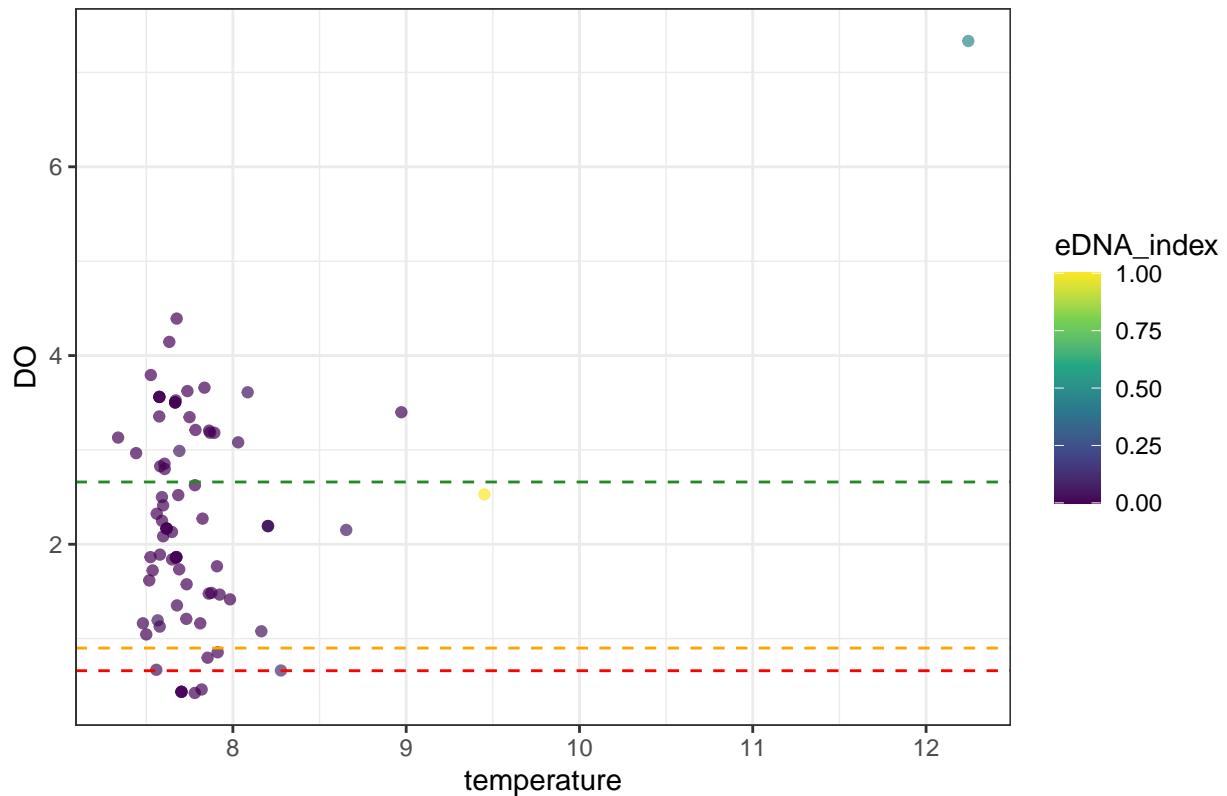
```
## [1] "Oithona similis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### Oithona similis eDNA Index vs Dissolved Oxygen & Temperature



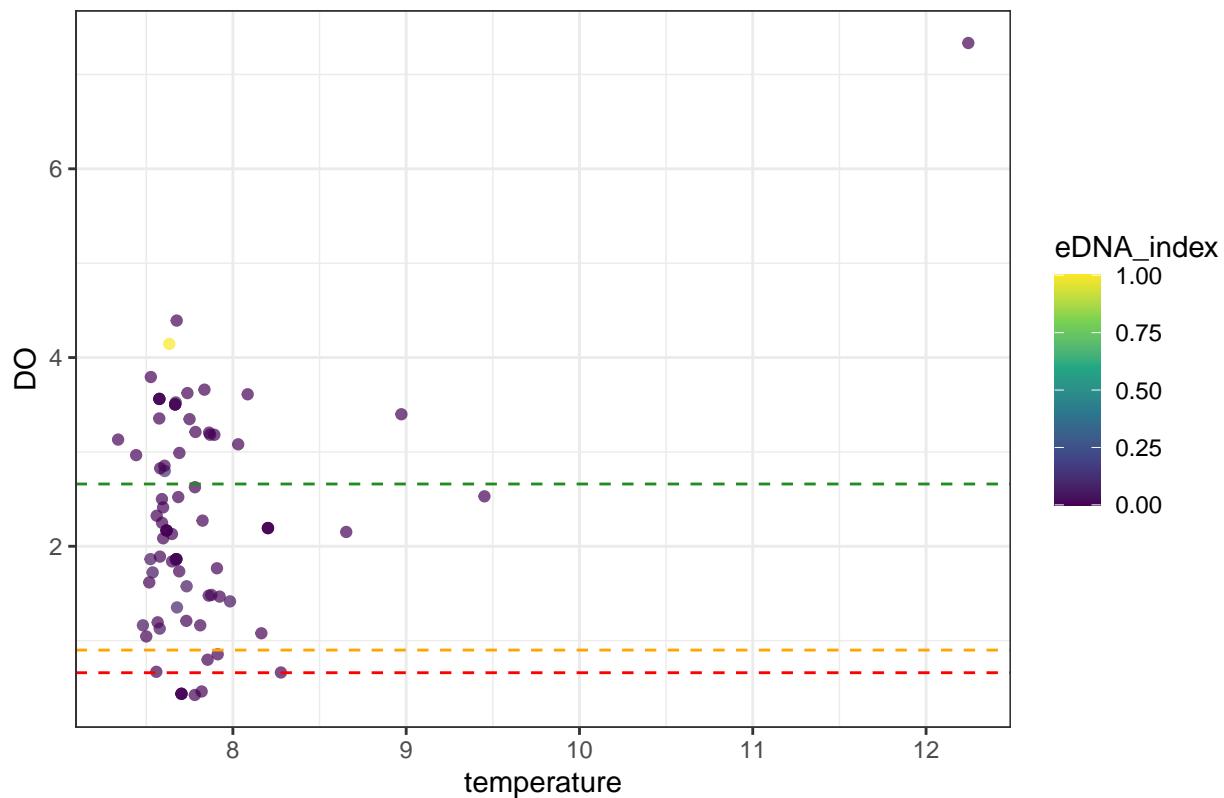
```
## [1] "Paracalanus sp. C AC-2013 eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Paracalanus sp. C AC–2013 eDNA Index vs Dissolved Oxygen & Temperature



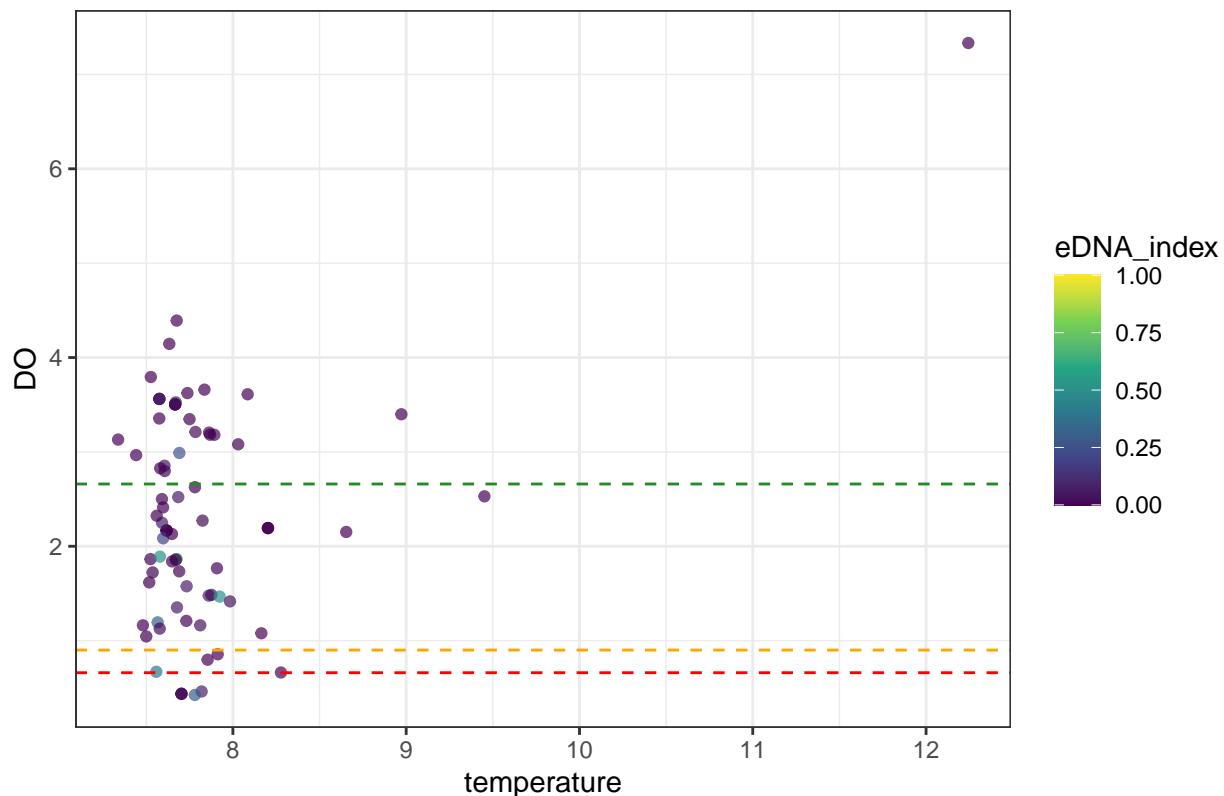
```
## [1] "Pseudocalanus newmani eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Pseudocalanus newmani eDNA Index vs Dissolved Oxygen & Temperature



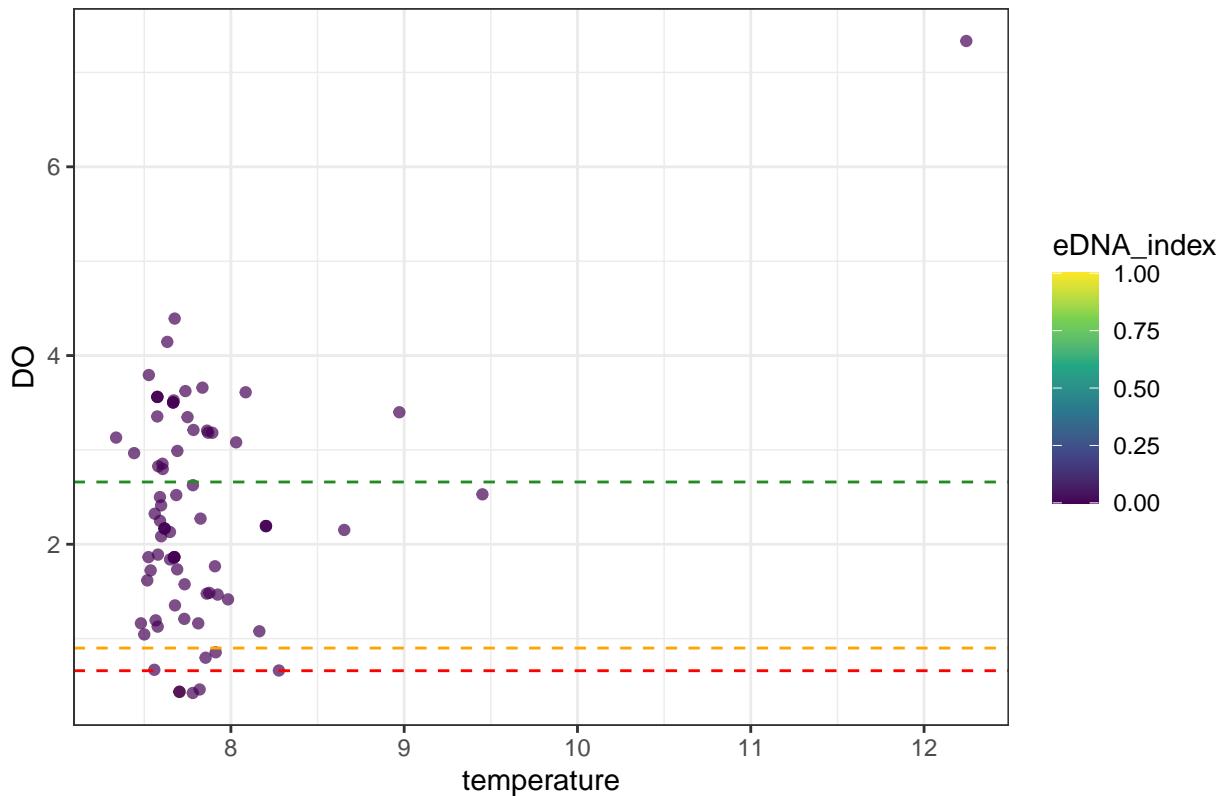
```
## [1] "Thermocyclops inversus eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### Thermocyclops inversus eDNA Index vs Dissolved Oxygen & Temperature



```
## [1] "Triconia minuta eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Triconia minuta eDNA Index vs Dissolved Oxygen & Temperature



```

for (i in 1:length(dfsplit)) { # For each species:
  species <- dfsplit[[i]]$Species[1] # Species name
  title <- paste(species, sep = " ", "eDNA Index vs Dissolved Oxygen & Temperature") #
  ↪ Plot title
  print(title)

# Plot + save

  print( # plot this species
    ggplot(dfsplit[[i]], aes(x = temperature,
                               y = DO,
                               color = eDNA_index)) +
    scale_color_continuous(type = "viridis") +
    geom_point(alpha = 0.5, shape = 15, size = 3) +
    ggttitle(title) +
    geom_hline(yintercept = 0.66, linetype = 2, color = "red") + # most copepods die
    geom_hline(yintercept = 0.9, linetype = 2, color = "orange") + # ~50% of copepods
    ↪ die
    geom_hline(yintercept = 2.66, linetype = 2, color = "forestgreen") + # many
    ↪ copepods experience sublethal effects (e.g. less egg production)
    theme_bw()
  )

  ggsave(filename = paste(species, sep = "_", "eDNA_DO_Temp_Heat.png"),
         path = here("eDNA_Index_Hypoxia", "Plots", "DO_Temp_Heat"),
         width = 2500, height = 2000, units = "px")
}

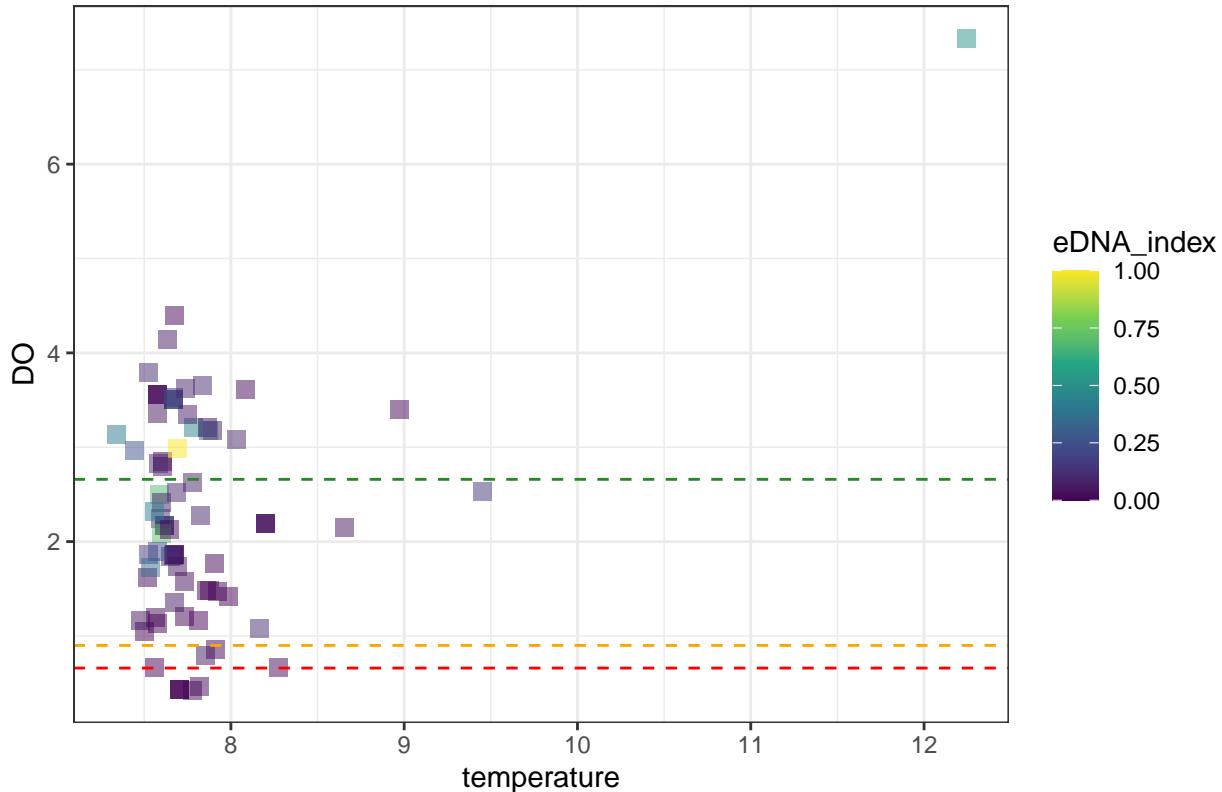
```

```

## [1] "Acartia longiremis eDNA Index vs Dissolved Oxygen & Temperature"
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

```

### Acartia longiremis eDNA Index vs Dissolved Oxygen & Temperature

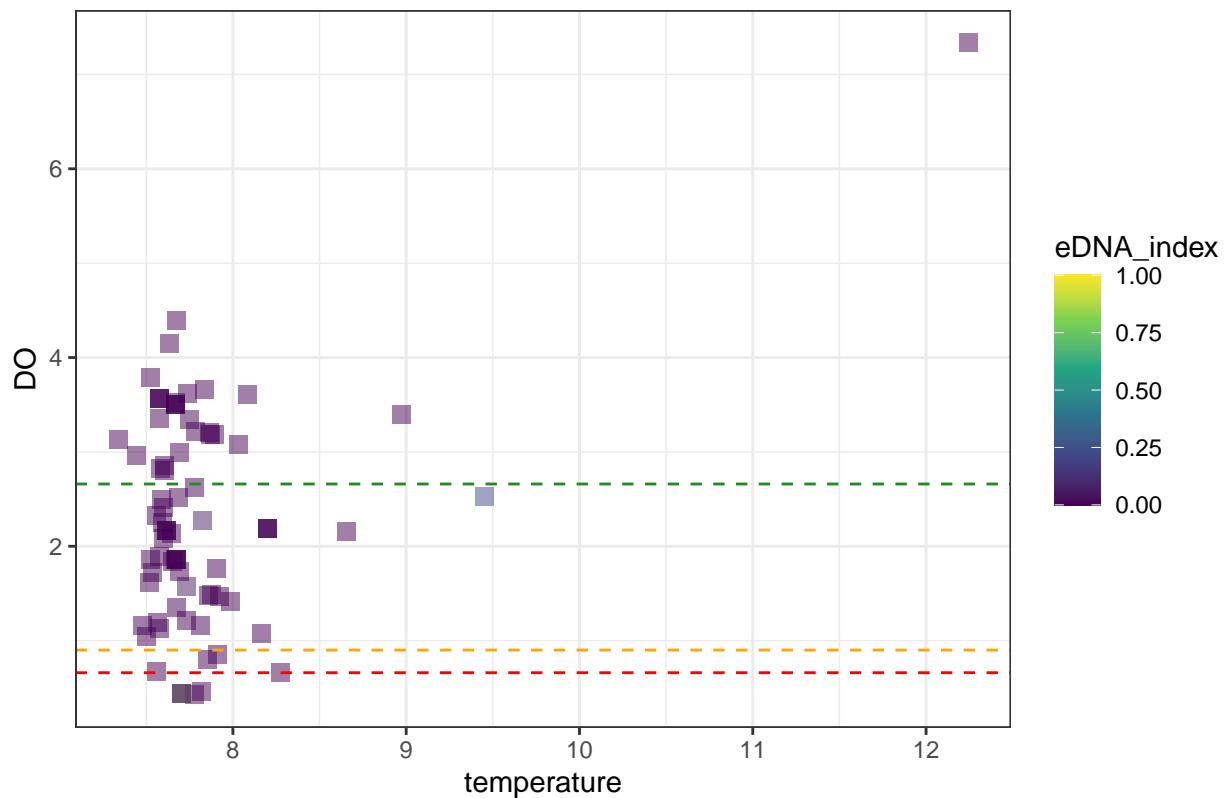


```

## [1] "Calanus pacificus eDNA Index vs Dissolved Oxygen & Temperature"
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

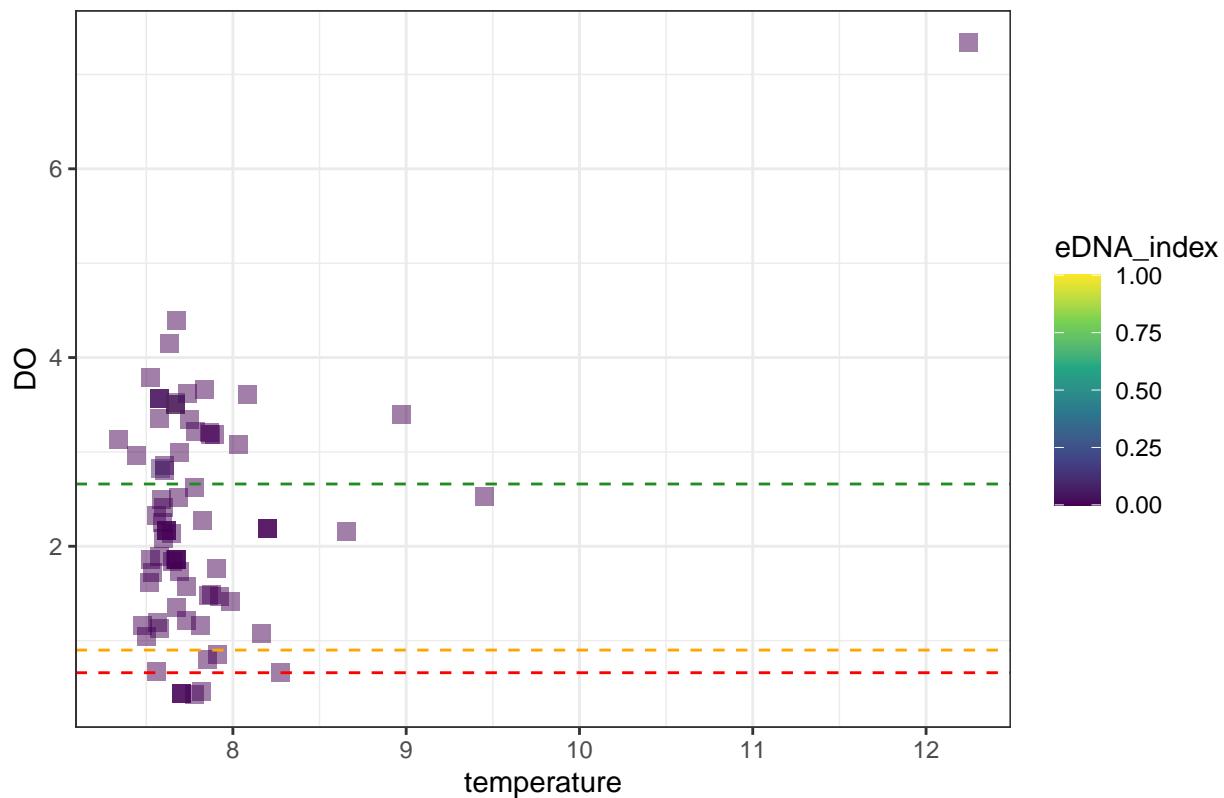
```

### Calanus pacificus eDNA Index vs Dissolved Oxygen & Temperature



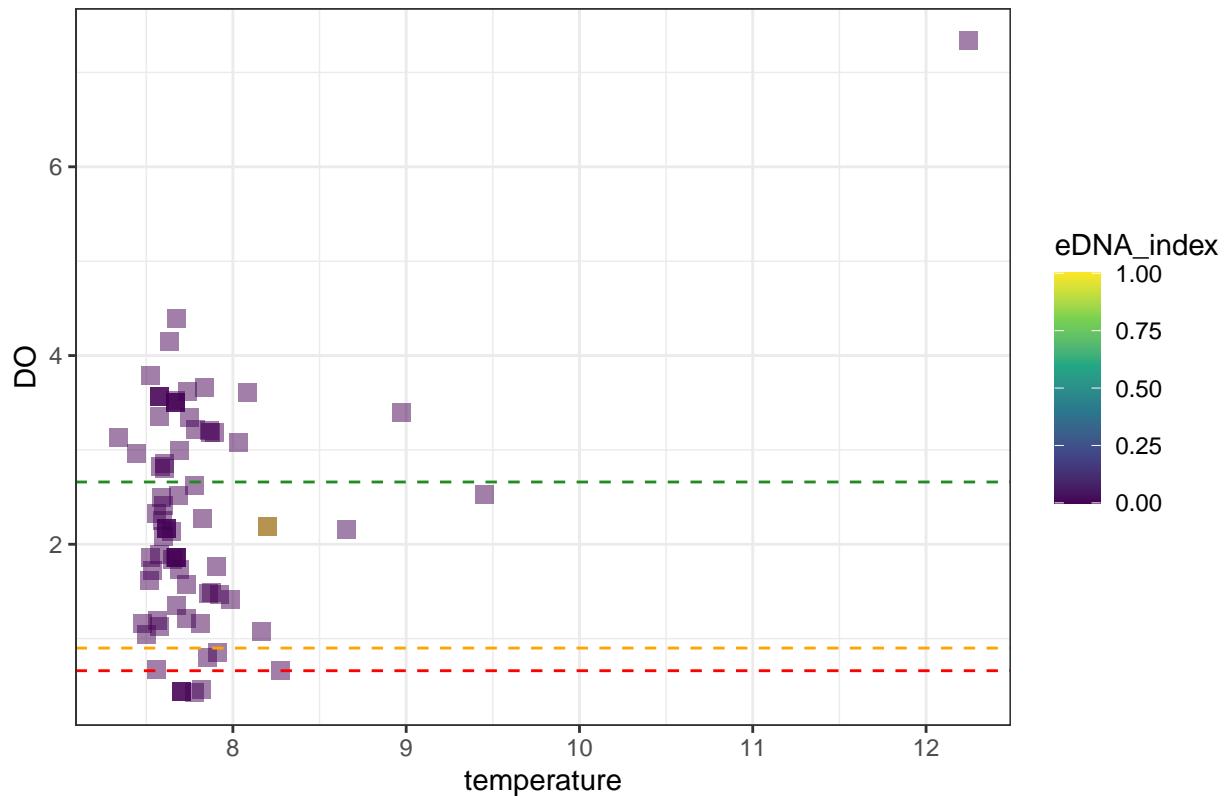
```
## [1] "Centropages abdominalis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Centropages abdominalis eDNA Index vs Dissolved Oxygen & Temperature



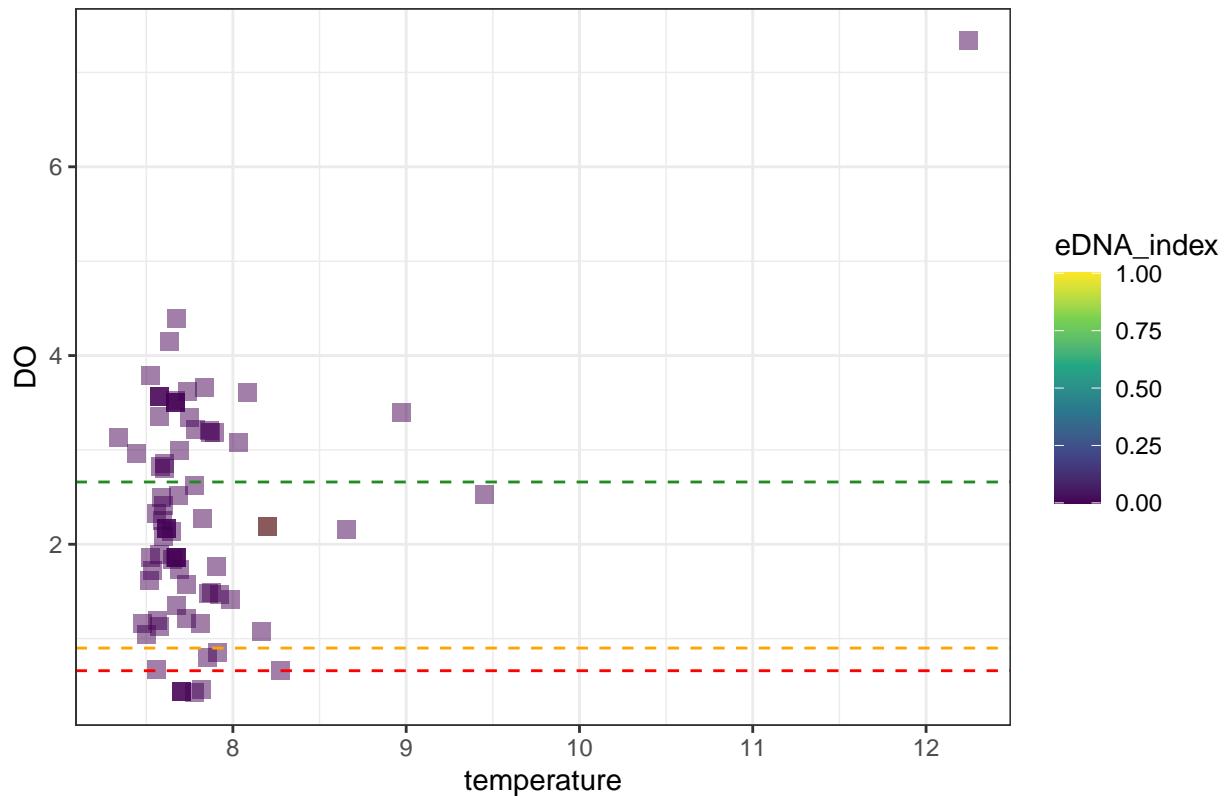
```
## [1] "Clausocalanus parapergens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Clausocalanus parapergens eDNA Index vs Dissolved Oxygen & Temperature



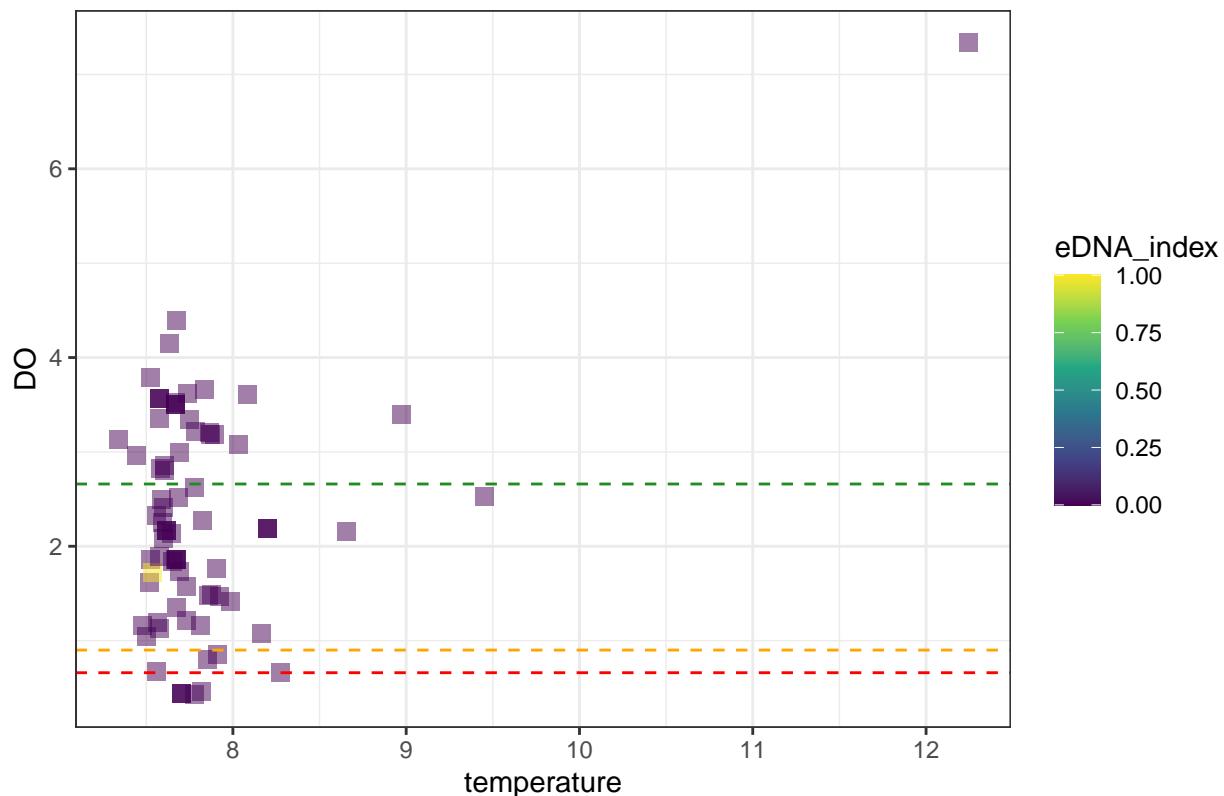
```
## [1] "Clausocalanus pergens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### Clausocalanus pergens eDNA Index vs Dissolved Oxygen & Temperature



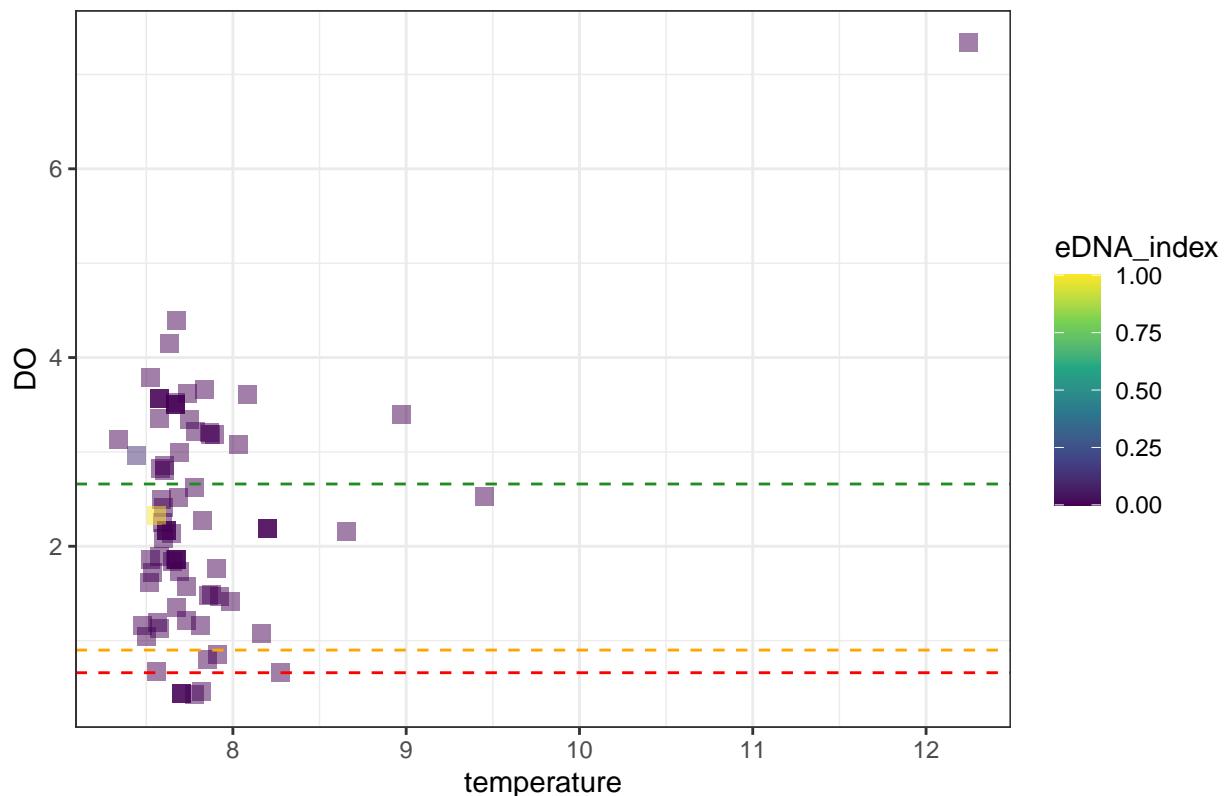
```
## [1] "Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Diacyclops incolotaenia eDNA Index vs Dissolved Oxygen & Temperature



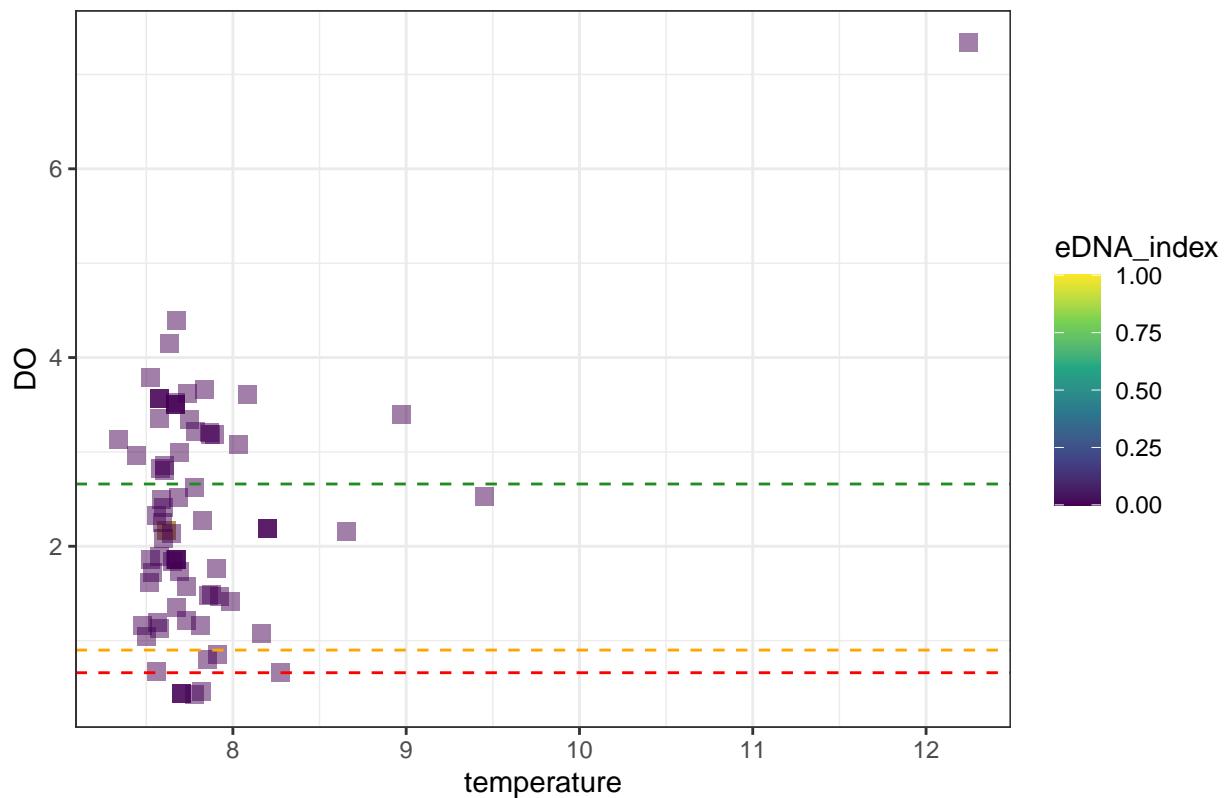
```
## [1] "Lucicutia flavicornis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### Lucicutia flavigornis eDNA Index vs Dissolved Oxygen & Temperature



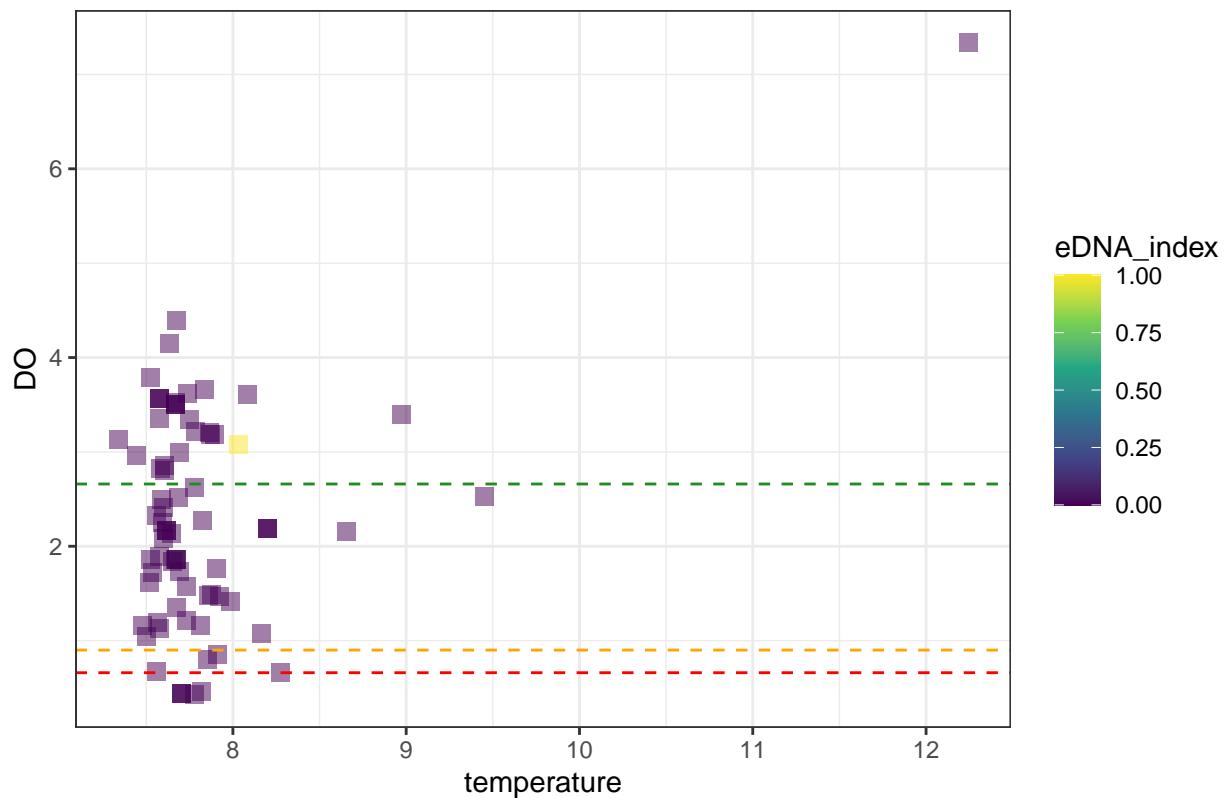
```
## [1] "Metridia lucens eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Metridia lucens eDNA Index vs Dissolved Oxygen & Temperature



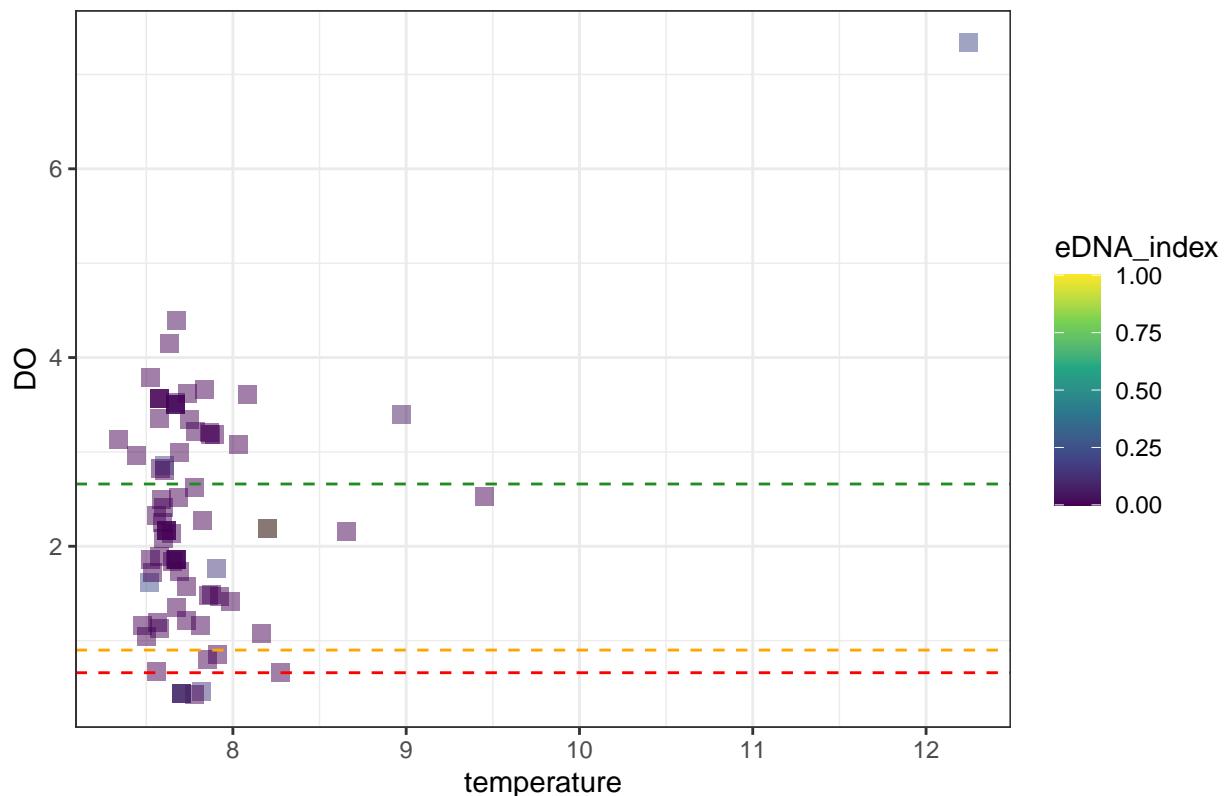
```
## [1] "Metridia pacifica eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Metridia pacifica eDNA Index vs Dissolved Oxygen & Temperature



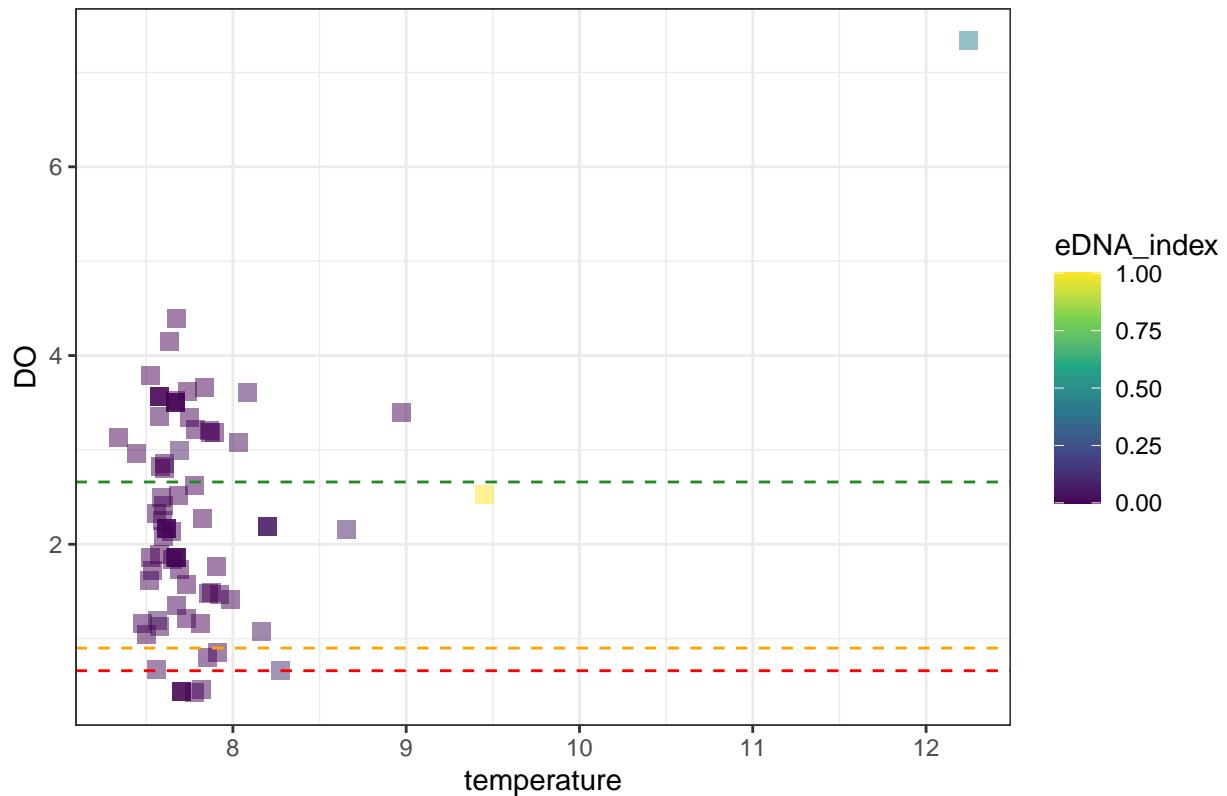
```
## [1] "Oithona similis eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Oithona similis eDNA Index vs Dissolved Oxygen & Temperature



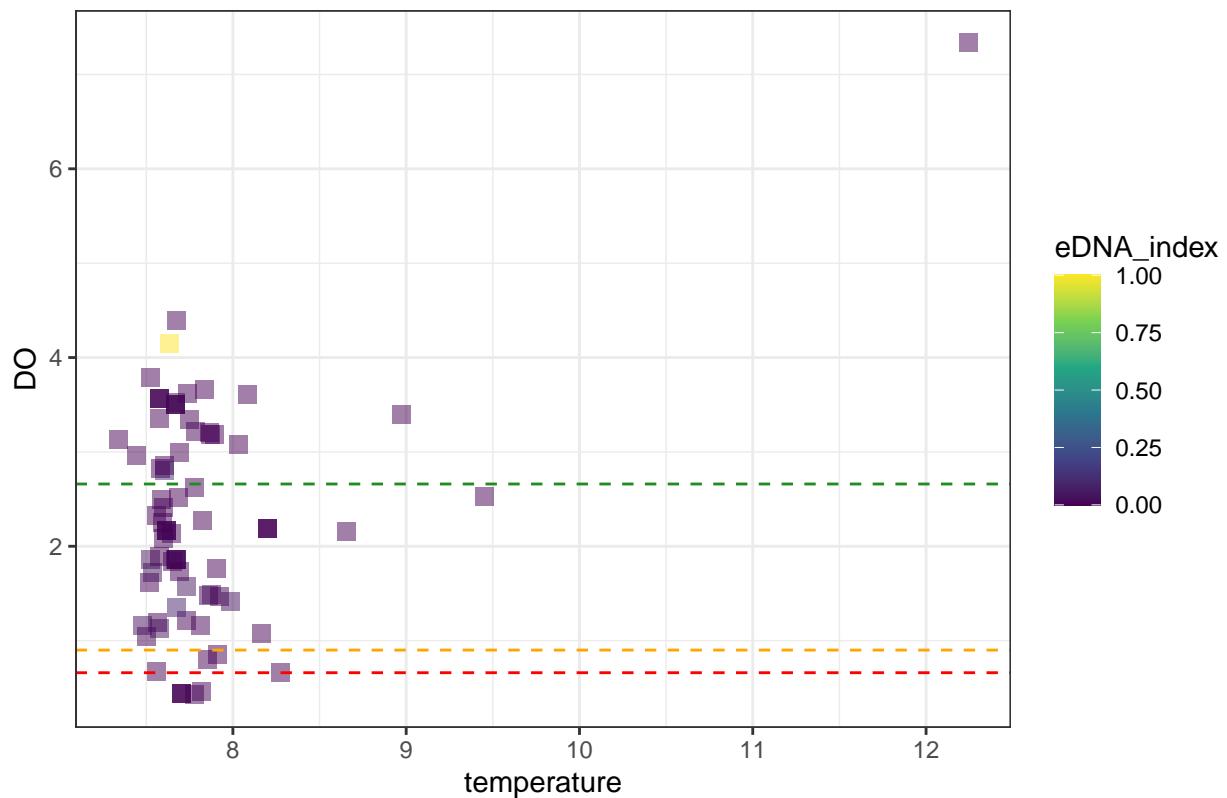
```
## [1] "Paracalanus sp. C AC-2013 eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Paracalanus sp. C AC-2013 eDNA Index vs Dissolved Oxygen & Temperature



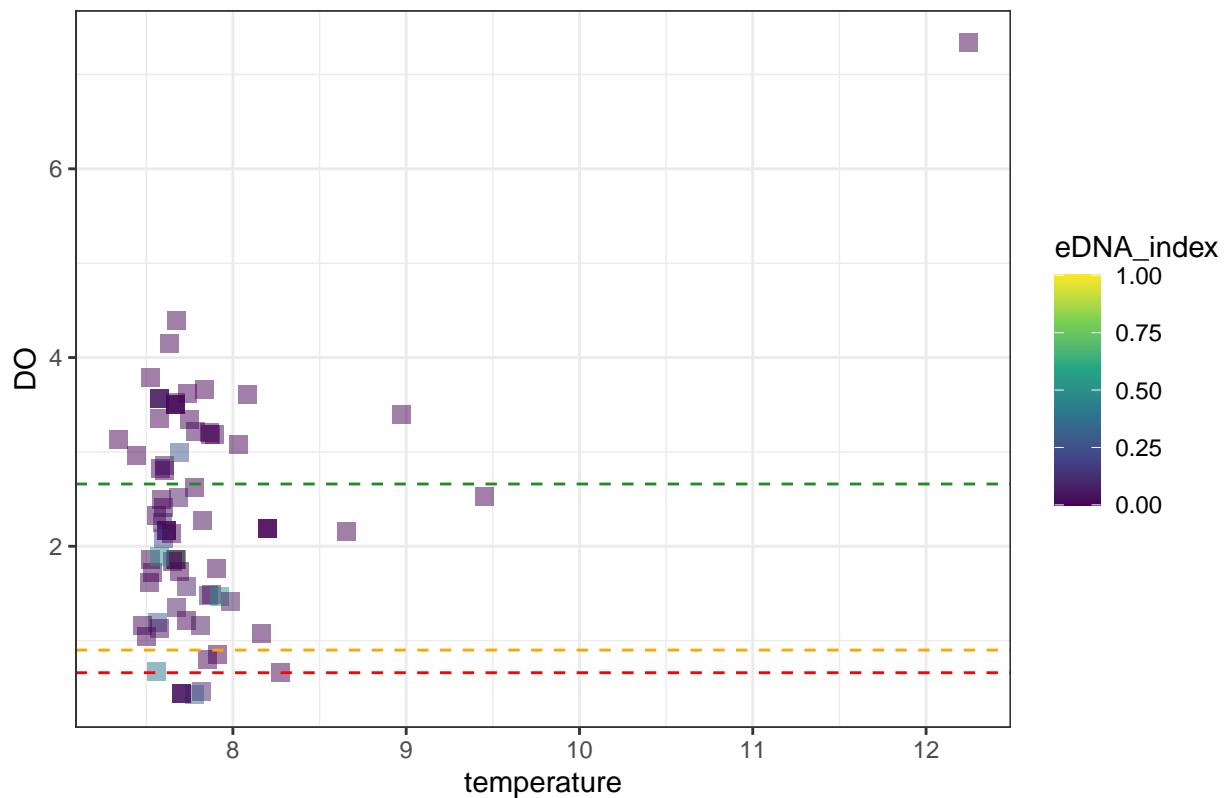
```
## [1] "Pseudocalanus newmani eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Pseudocalanus newmani eDNA Index vs Dissolved Oxygen & Temperature



```
## [1] "Thermocyclops inversus eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

## Thermocyclops inversus eDNA Index vs Dissolved Oxygen & Temperature



```
## [1] "Triconia minuta eDNA Index vs Dissolved Oxygen & Temperature"  
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom_point()`).
```

### Triconia minuta eDNA Index vs Dissolved Oxygen & Temperature

