"OCNMS CTD Data Exploration"

Eleanor (Ella) Crotty

Contents

Setup

```
# Warnings and startup messages suppressed
library(tidyverse)
library(patchwork) # Put plots together
library(scales) # Rescale datetime axes
library(ggrepel)
library(readxl)
library(here) # Project/filepath management
library(maps)
library(RColorBrewer) # Color palettes
library(colorRamps) # Color palettes
```

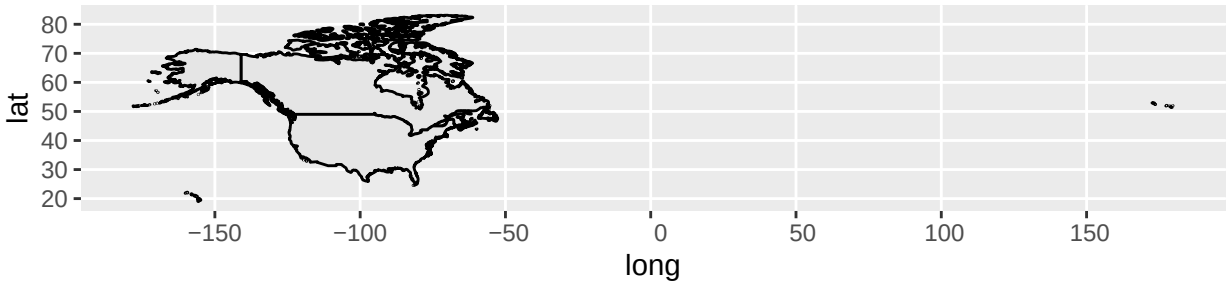
```
wd <- "OCNMS_Hypoxia/Data/CTD_Data"
exp <- "OCNMS_Hypoxia/Outputs"
pltpath <- "OCNMS_Hypoxia/Plots/CTD_Data_Exploration_Plots"

OME_CTD <- read.csv(here(wd, "OCNMS_OME_ctd_output_copy.csv"))
OCNMS_OME_CTD <- read.csv(here(wd, "OCNMS_OMEsites_ctd_output_copy.csv"))
OCNMS_All_CTD <- read.csv(here(wd, "OCNMS_Allsites_ctd_output_copy.csv"))

# Problem: all the longitudes are positive, they need to be negative
fixlong <- function(df) {
  df$longitude <- df$longitude*-1
  df
}

head(fixlong(OME_CTD))
```

```
##             startTime             filename         cast_info station_id
```

```
## 1 2021-10-18T18:21:17Z CAST_CE042_211018.cnv CAST_CE042_211018      CE042
## 2 2021-10-18T18:21:17Z CAST_CE042_211018.cnv CAST_CE042_211018      CE042
## 3 2021-10-18T18:21:17Z CAST_CE042_211018.cnv CAST_CE042_211018      CE042
## 4 2021-10-18T18:21:17Z CAST_CE042_211018.cnv CAST_CE042_211018      CE042
## 5 2021-10-18T18:21:17Z CAST_CE042_211018.cnv CAST_CE042_211018      CE042
## 6 2021-10-18T18:21:17Z CAST_CE042_211018.cnv CAST_CE042_211018      CE042
##         date    timeJ pressure  depth temperature conductivity salinity
## 1 2021-10-18 291.7648    -4.39 -4.353     12.2923     0.000788   0.0095
## 2 2021-10-18 291.7648    -4.39 -4.353     12.3163     0.000788   0.0095
## 3 2021-10-18 291.7648    -4.12 -4.086     12.3431     0.000788   0.0095
## 4 2021-10-18 291.7648    -4.12 -4.086     12.3713     0.000788   0.0095
## 5 2021-10-18 291.7648    -4.12 -4.086     12.3996     0.000788   0.0095
## 6 2021-10-18 291.7648    -4.12 -4.086     12.4278     0.000788   0.0095
##   soundSpeed   oxygen fluorescence beamAttenuation beamTransmission descentRate
## 1    1456.07 -2.85347            0          55.262          -1.2928  1.1369e-15
## 2    1456.16 -2.85191            0          55.262          -1.2928 -1.7760e-16
## 3    1456.26 -2.85026            0          55.262          -1.2928  1.0800e-01
## 4    1456.37 -2.84843            0          55.262          -1.2928  1.6200e-01
## 5    1456.48 -2.84659            0          55.262          -1.2928  1.6200e-01
## 6    1456.58 -2.84477            0          55.262          -1.2928  1.0800e-01
##   bpos scan flag   station_name latitude longitude bottom_depth_m
## 1    0    2    0 Cape Elizabeth  47.3531 -124.4887             42
## 2    0    3    0 Cape Elizabeth  47.3531 -124.4887             42
## 3    0    4    0 Cape Elizabeth  47.3531 -124.4887             42
## 4    0    5    0 Cape Elizabeth  47.3531 -124.4887             42
## 5    0    6    0 Cape Elizabeth  47.3531 -124.4887             42
## 6    0    7    0 Cape Elizabeth  47.3531 -124.4887             42
```

```
OME_CTD <- fixlong(OME_CTD)
OCNMS_OME_CTD <- fixlong(OCNMS_OME_CTD)
OCNMS_All_CTD <- fixlong(OCNMS_All_CTD)
# All better!

# Problem #2: OME used oxygen, OCNMS uses dissolved_oxygen
OCNMS_All_CTD <- OCNMS_All_CTD %>%
  rename(DO = dissolved_oxygen)
OCNMS_OME_CTD <- OCNMS_OME_CTD %>%
  rename(DO = dissolved_oxygen)
OME_CTD <- OME_CTD %>%
  rename(DO = oxygen)
```

```
mapUC <- map_data("world", region = c("usa", "canada"))

ggplot(mapUC, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "gray90", color = "black") +
  coord_sf() # coord_quickmap is an approximation to preserve straight lines, which works
  ↪  best for small areas close to the equator. projection can be defined (see
  ↪  mapproj::mapproject() for list) and R now recommends using coord_sf(). coord_sf()
  ↪  takes xlim, ylim, crs
```



Some excerpts from the R for Data Science tutorial

```
nz <- map_data("nz")

ggplot(nz, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "white", color = "black") +
  coord_quickmap() # This will fix the weird stretch usually
```

```
# Making a ggplot with label changes
histogram <- function(df, var, binwidth) {
  label <- rlang::englue("A histogram of {{var}} with binwidth {binwidth}")

  df |>
    ggplot(aes(x = {{ var }})) +
    geom_histogram(binwidth = binwidth) +
    labs(title = label)
}

diamonds |> histogram(carat, 0.1)
diamonds |> histogram(price, 1000)
```

```
df <- tribble(
  ~id, ~measurement, ~value,
  "A",        "bp1",    100,
  "B",        "bp1",    140,
  "B",        "bp2",    115,
  "A",        "bp2",    120,
  "A",        "bp3",    105
)
```

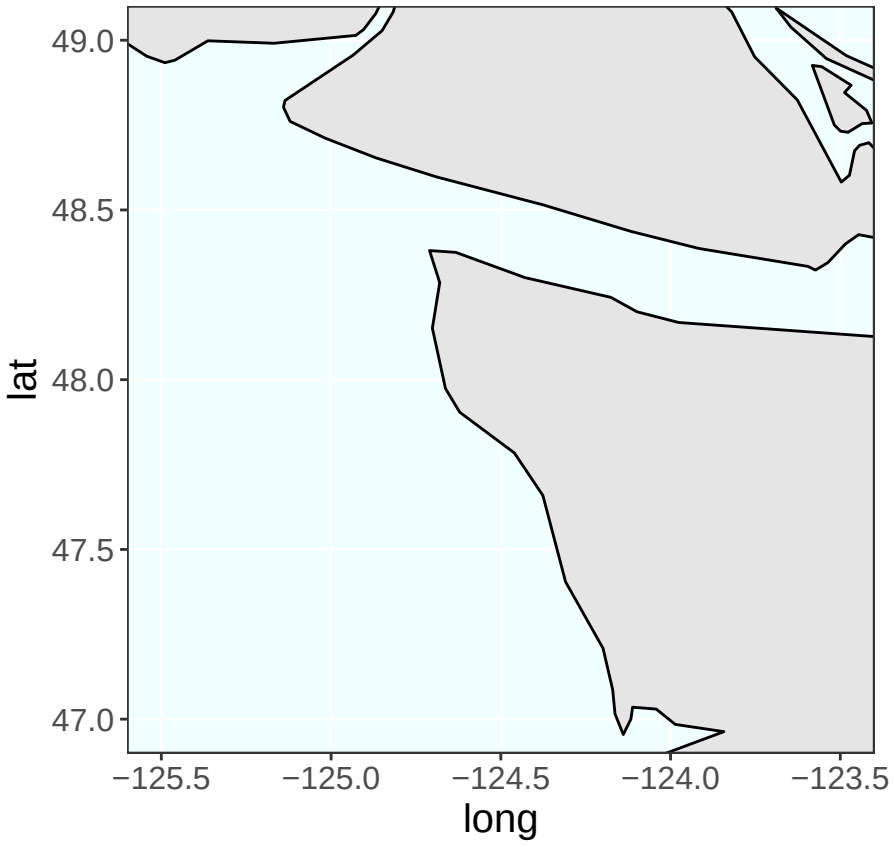
Functions

```
# Variable to make a base map of OCNMS, which I can then add data points to

rangeOC <- tribble(
  ~MinLong, ~MaxLong, ~MinLat, ~MaxLat,
  min(OME_CTD$longitude), max(OME_CTD$longitude), min(OME_CTD$latitude),
  ↪  max(OME_CTD$latitude)
)

OCNMS_x <- c(-123.5, -125.5)
OCNMS_y <- c(47,49)

mapOCNMS <- ggplot(mapUC, aes(x = long, y = lat, group = group)) +
```

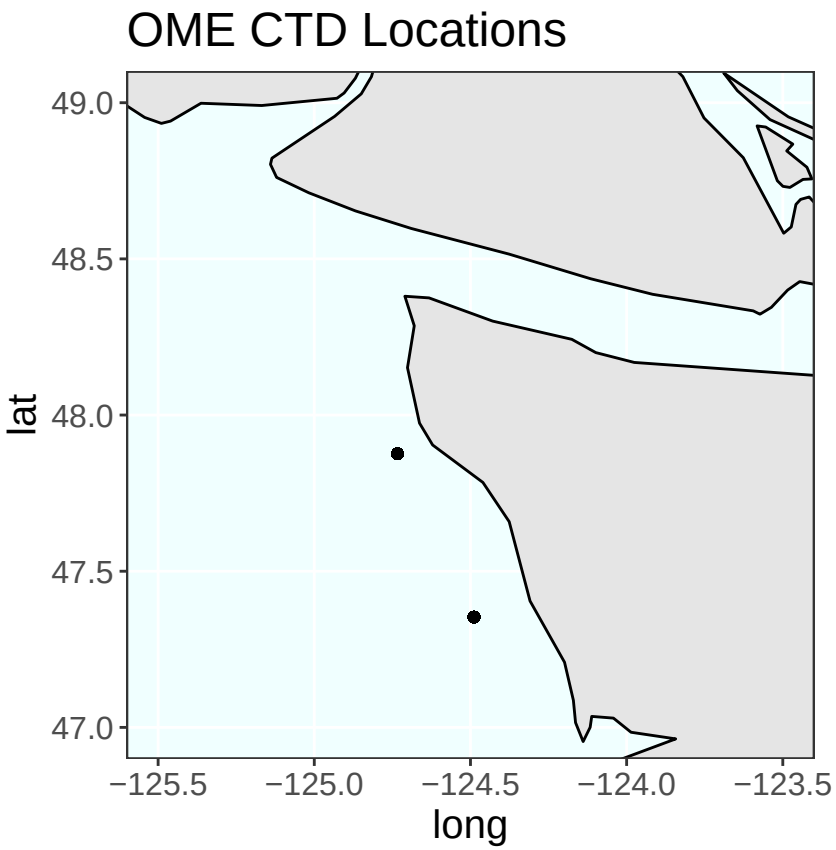
```
  geom_polygon(fill = "gray90", color = "black") +
  coord_sf(xlim = OCNMS_x,
           ylim = OCNMS_y) +
  theme_bw() +
  theme(text = element_text(size=15),
        panel.background = element_rect(fill = "azure1",
                                        colour = "azure1"),
        legend.key = element_rect(fill = "white",
                                  color = "white"),
        # It added in blue behind the dots in the key and I don't want that
        panel.grid.major = element_line(size = 0.5,
                                        linetype = 'solid',
                                        colour = "white"),
        panel.grid.minor = element_line(size = 0.25,
                                        linetype = 'solid',
                                        colour = "white")) # +
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

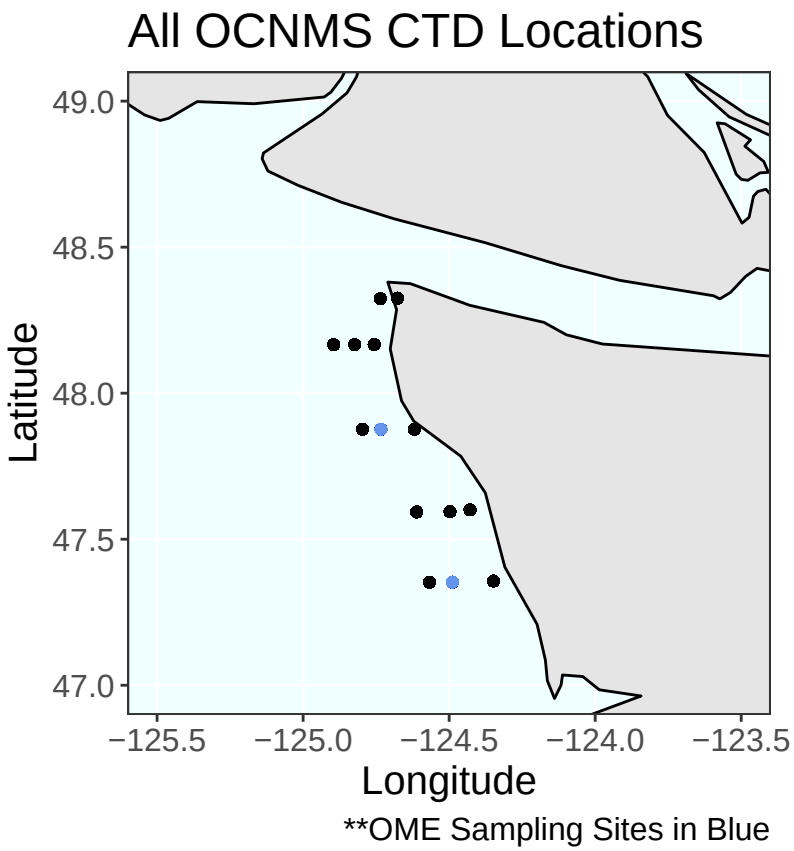
```
  # coord_quickmap() # This will fix the weird stretch usually, can also do
  ↪  coord_fixed(ratio = 1.3)

mapOCNMS
```



```
mapOCNMS +
  geom_point(data = OME_CTD, aes(x = longitude, y = latitude, group = NA)) + # aes(group
  ↪  = NA) or geom_point(inherit = FALSE) keeps it from looking for a group since the
  ↪  base layer has groups
  ggtitle("OME CTD Locations")
```




```
mapOCNMS +
  geom_point(data = OCNMS_All_CTD, aes(x = longitude, y = latitude, group = NA)) + #
  ↪  aes(group = NA) or geom_point(inherit = FALSE) keeps it from looking for a group
  ↪  since the base layer has groups
  geom_point(data = OME_CTD, aes(x = longitude, y = latitude, group = NA), color =
  ↪  "cornflowerblue") +
  labs(title = "All OCNMS CTD Locations", caption = "**OME Sampling Sites in Blue", x =
  ↪  "Longitude", y = "Latitude")
```




```
ggsave("CTD_Locations.png", path = here(pltpath), dpi = 500)
```

```
## Saving 6.5 x 4.5 in image
```

```
# Histogram function
histogram_fill <- function(df, var, binwidth, fill = "darkgray") {
  label <- rlang::englue("A histogram of {{var}} in {{df}} with binwidth {binwidth}")

  df |>
    ggplot(aes(x = {{ var }}, fill = {{fill}})) +
    geom_histogram(binwidth = binwidth, color = "black") +
    labs(title = label) +
    theme_bw()
}

histogram2 <- function(df, var, binwidth, fill = NA) { # Not currently functional
  label <- rlang::englue("A histogram of {{var}} in {{df}} with binwidth {binwidth}")

  if (is.na(fill)) {
    df |>
    ggplot(aes(x = {{ var }})) +
    geom_histogram(binwidth = binwidth, color = "black", fill = "darkgray") +
    labs(title = label) +
    theme_bw()
  } else {
  df |>
    ggplot(aes(x = {{ var }}, fill = {{fill}})) +
```

```
    geom_histogram(binwidth = binwidth, color = "black") +
    labs(title = label) +
    theme_bw()
  }

}

histogram <- function(df, var, binwidth) {
  label <- rlang::englue("A histogram of {{var}} in {{df}} with binwidth {binwidth}")

  df |>
    ggplot(aes(x = {{ var }})) +
    geom_histogram(binwidth = binwidth, color = "black", fill = "darkgray") +
    labs(title = label) +
    theme_bw()
}
```

Cleaning

Add dates & years

```
OME_CTD <- OME_CTD %>%
  mutate(date = as.Date(date)) %>%
  mutate(year = as.factor(year(date)))

OME_CTD %>%
  group_by(year) %>%
  summarize(Observations_OME = n())
```

```
## # A tibble: 4 x 2
##   year  Observations_OME
##   <fct>            <int>
## 1 2021              3301
## 2 2022              6130
## 3 2023              2289
## 4 2024              1226
```

```
OCNMS_OME_CTD <- OCNMS_OME_CTD %>%
  mutate(date = as.Date(date)) %>%
  mutate(year = as.factor(year(date)))

OCNMS_OME_CTD %>%
  group_by(year) %>%
  summarize(Observations_OCNMS_filtOME = n())
```

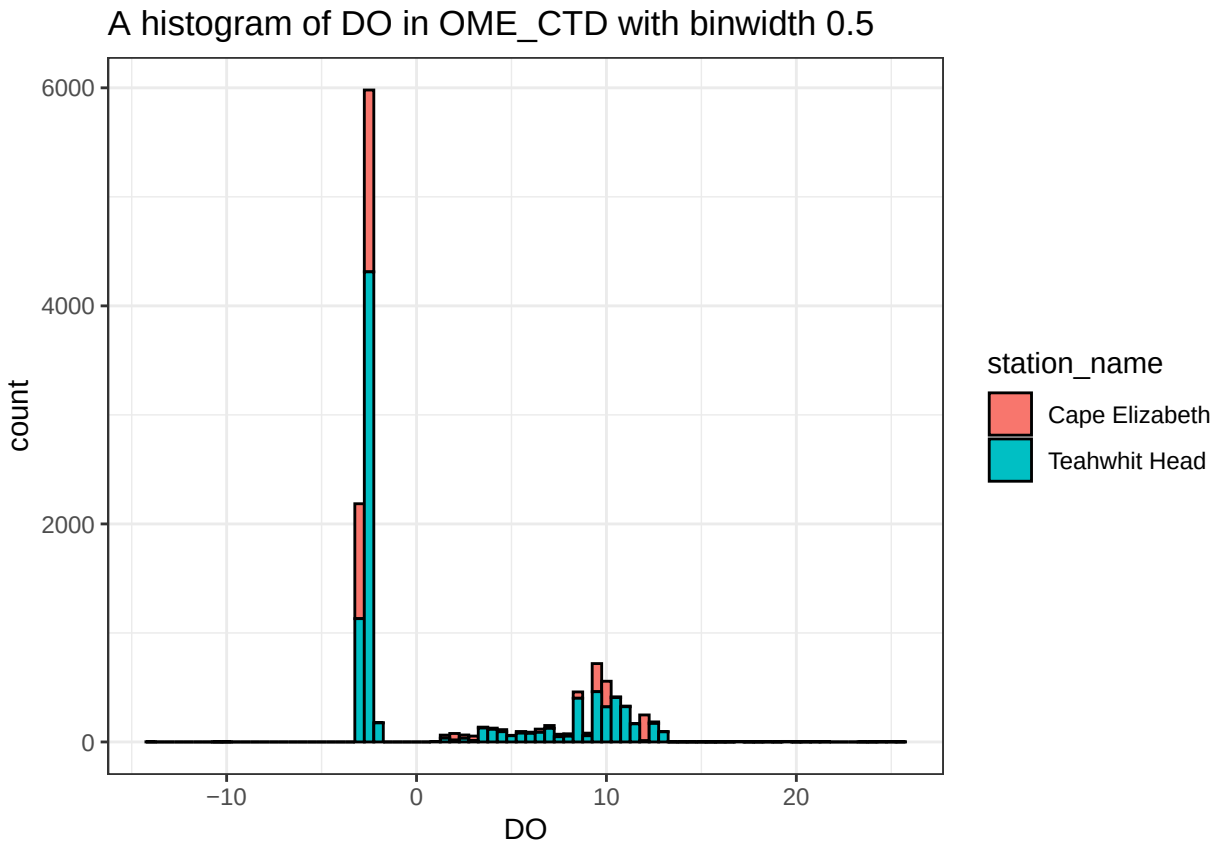
```
## # A tibble: 19 x 2
##    year  Observations_OCNMS_filtOME
##    <fct>                      <int>
##  1 2005                          83
##  2 2006                         395
##  3 2007                         369
##  4 2008                         329
##  5 2009                         251
```

```
##  6 2010                          272
##  7 2011                          218
##  8 2012                          503
##  9 2013                          336
## 10 2014                          421
## 11 2015                          343
## 12 2016                          383
## 13 2017                          423
## 14 2018                          344
## 15 2019                          454
## 16 2020                          208
## 17 2021                          337
## 18 2022                          409
## 19 2023                          410
```

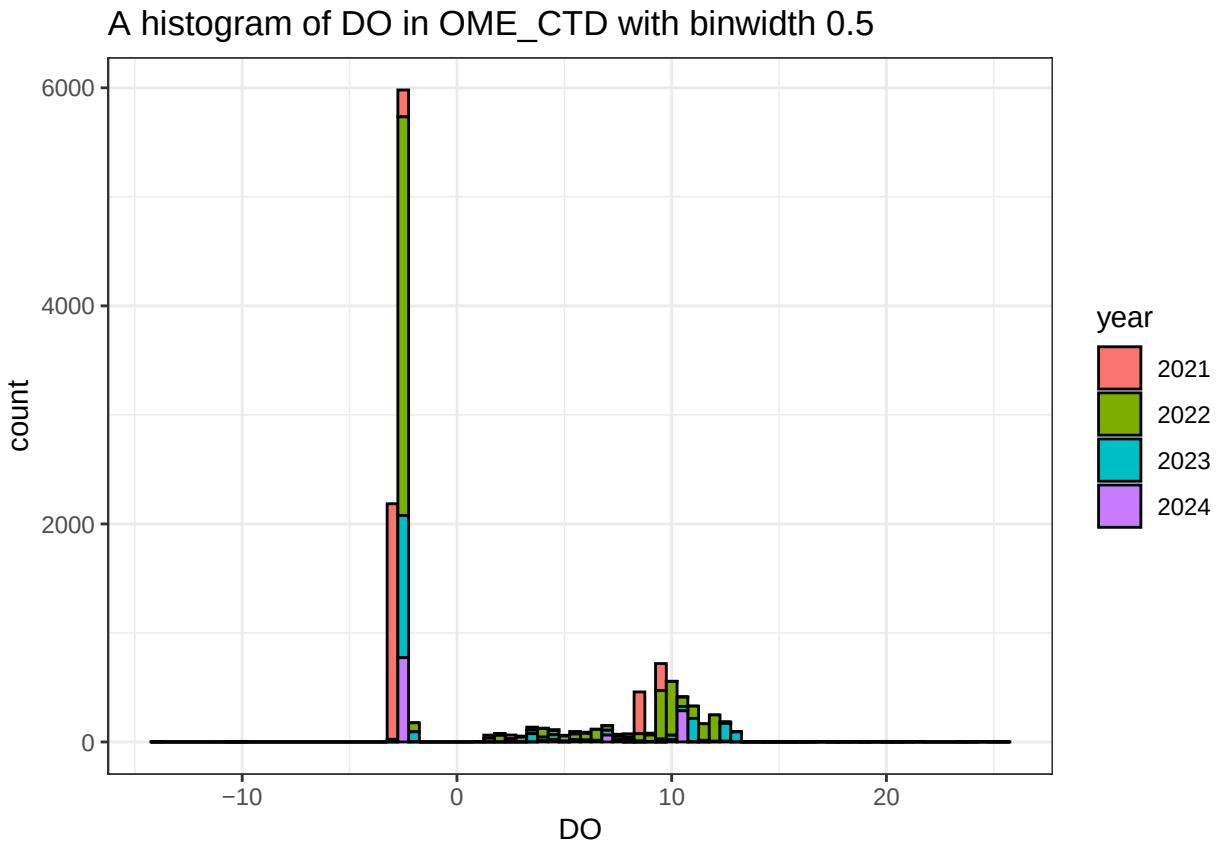
```
OCNMS_All_CTD <- OCNMS_All_CTD %>%
  mutate(date = as.Date(date)) %>%
  mutate(year = as.factor(year(date)))

OCNMS_All_CTD %>%
  group_by(year) %>%
  summarize(Observations_OCNMS = n())
```

```
## # A tibble: 19 x 2
##    year  Observations_OCNMS
##    <fct>              <int>
##  1 2005                 336
##  2 2006                1534
##  3 2007                2058
##  4 2008                1318
##  5 2009                1151
##  6 2010                1495
##  7 2011                1514
##  8 2012                1603
##  9 2013                1233
## 10 2014                1375
## 11 2015                1118
## 12 2016                1122
## 13 2017                1323
## 14 2018                1116
## 15 2019                1338
## 16 2020                 733
## 17 2021                 919
## 18 2022                1173
## 19 2023                1238
```

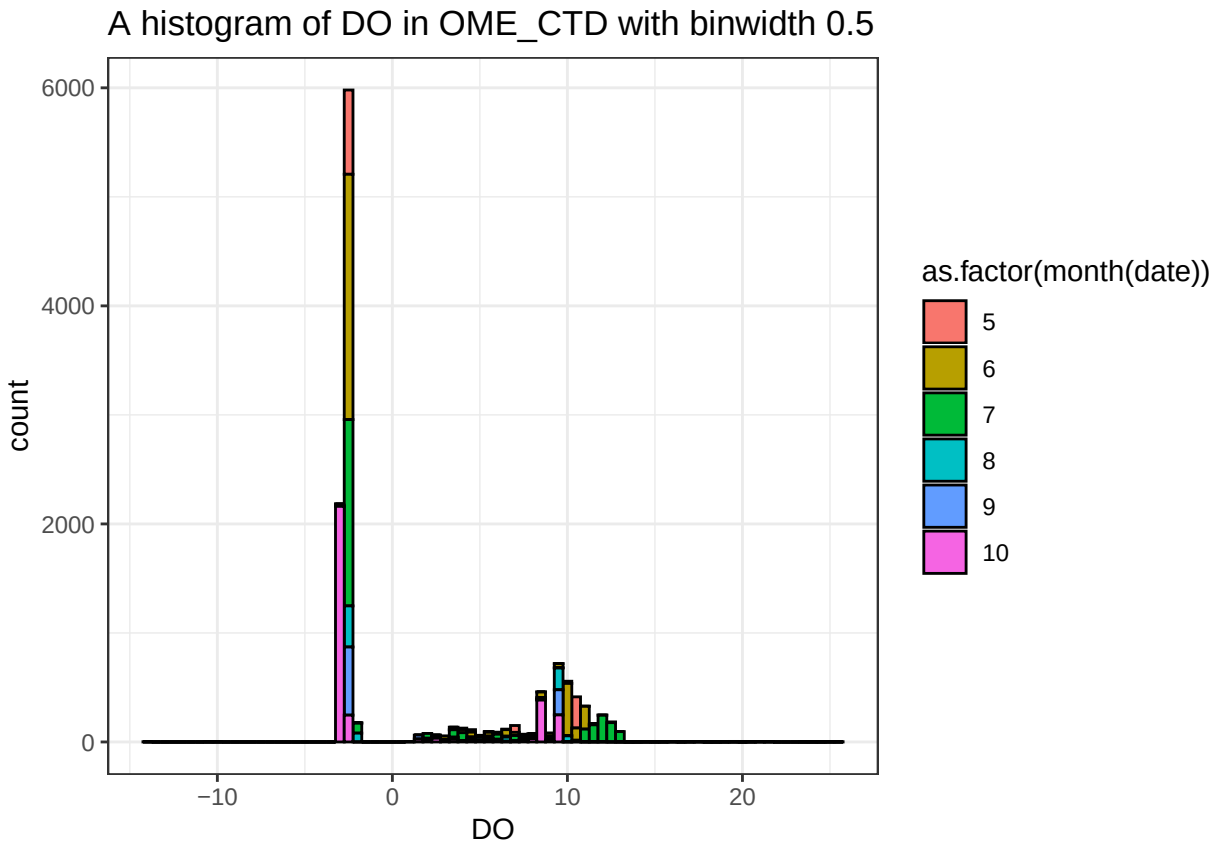
```
# Why is some of it negative???
histogram_fill(OME_CTD, DO, 0.5, fill = station_name)
```




```
histogram_fill(OME_CTD, DO, 0.5, fill = year)
```

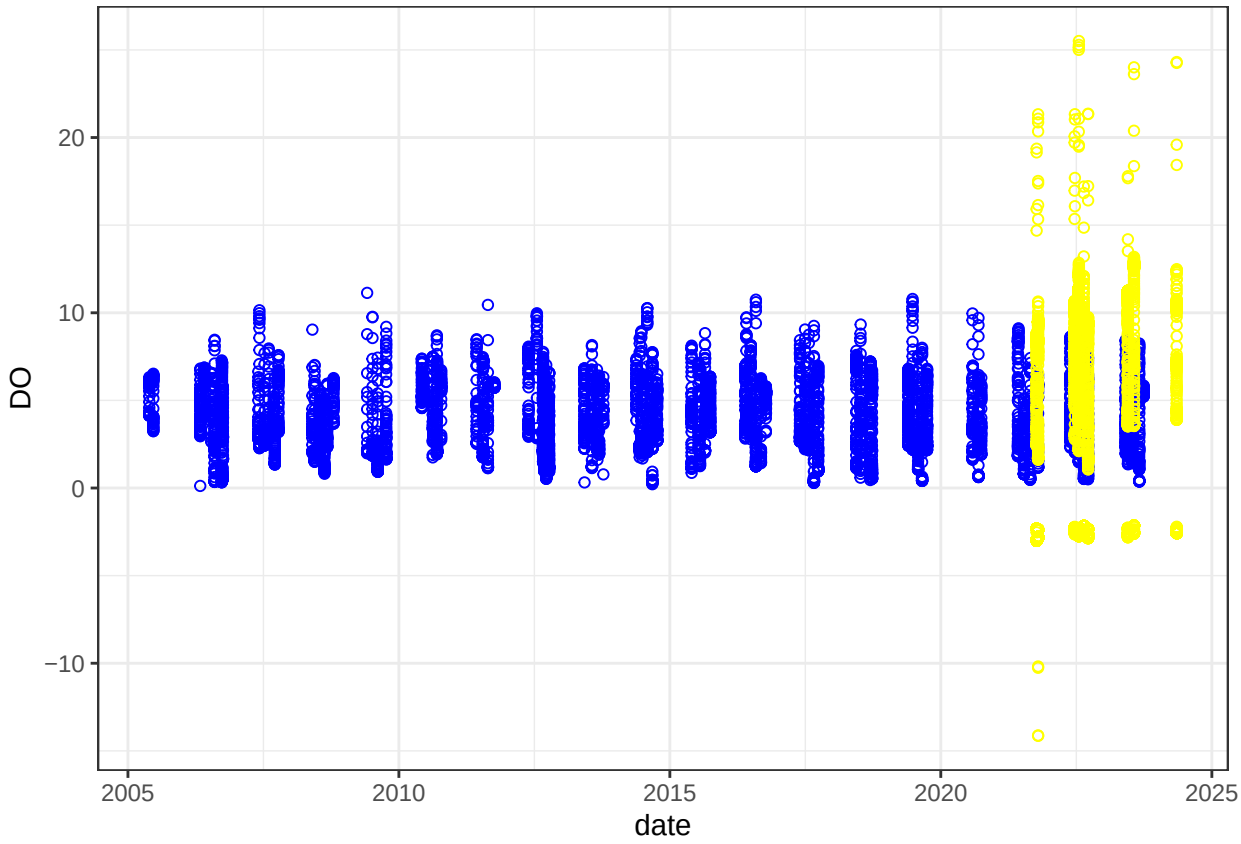



```
histogram_fill(OME_CTD, DO, 0.5, fill = as.factor(month(date)))
```




```
# Shannon says the OME files are redundant and the OCNMS published data probably just
↪  eliminated that because the quality score was low or something. Shannon will check
↪  with Sam. Use OCNMS_OME_CTD for now.

# Plot OME vs. OCNMS to see if redundant
ggplot(OCNMS_OME_CTD, aes(x = date, y = DO)) +
  geom_point(shape = 1, color = "blue") +
  theme_bw() +
  geom_point(data = OME_CTD, aes(x = date, y = DO), color = "yellow", shape = 1)
```

```
## Warning: Removed 7 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
# Probably redundant, OME is the only one with 2024 though
```

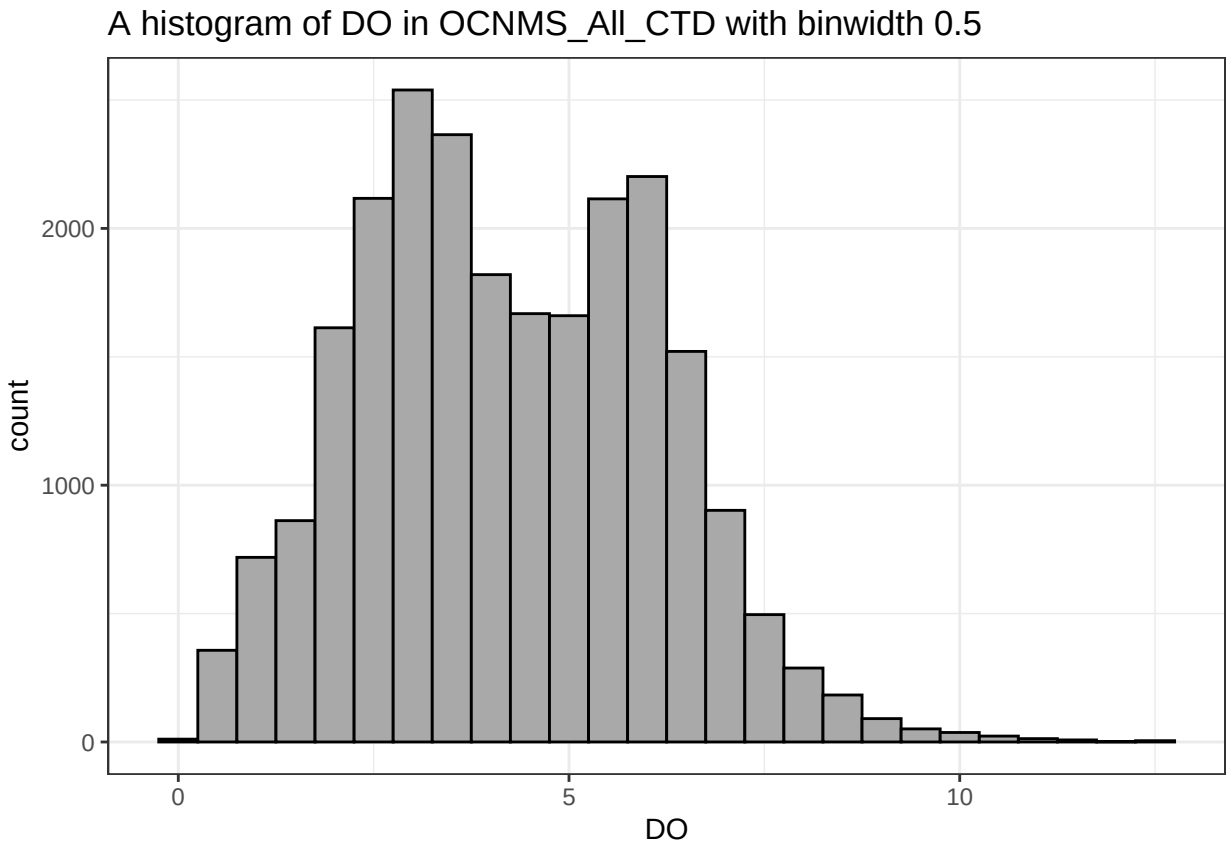
```
# While plotting depth, I found 2 rows with depths of -9999 and several other NA values.
↪  Remove.
OCNMS_OME_CTD <- OCNMS_OME_CTD %>%
  filter(depth > 0) # This removes 2 rows with NA values + -9999 depth
```

Oxygen Data Exploration

Histograms

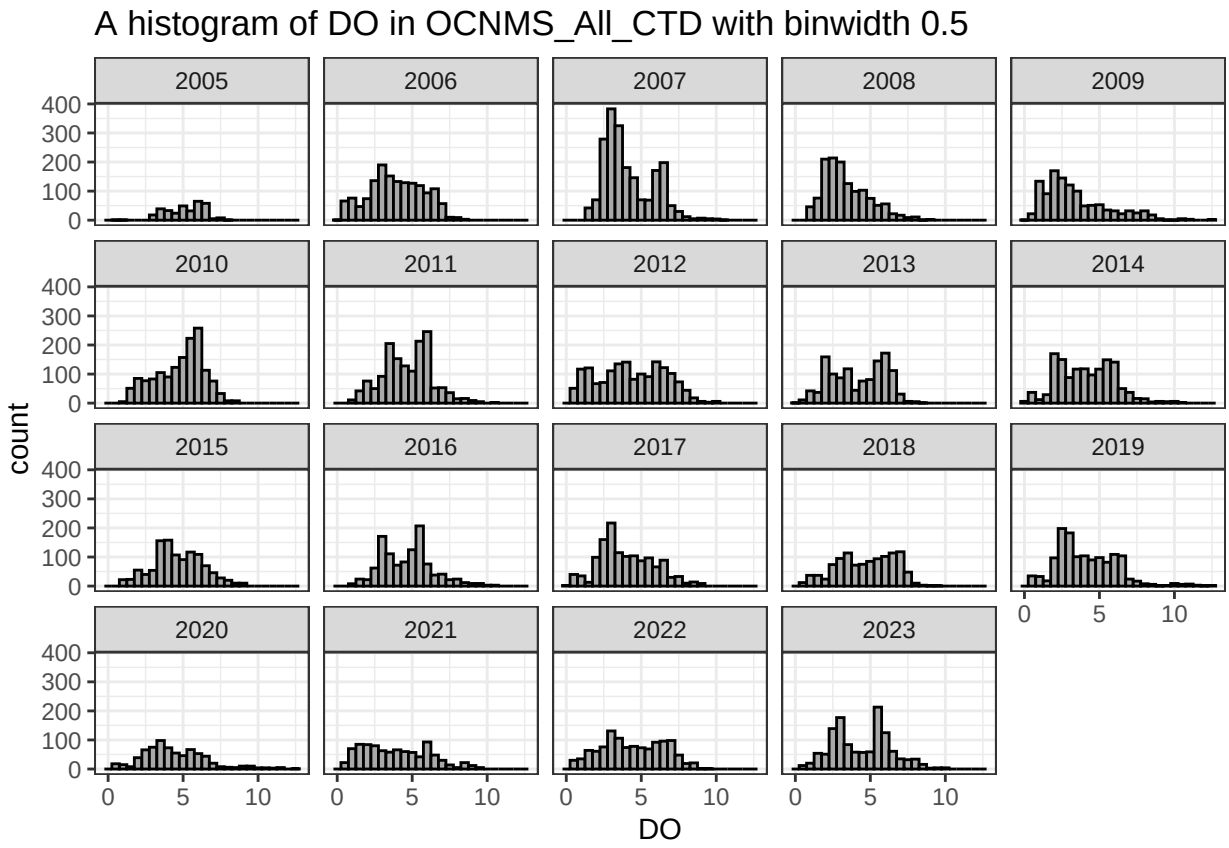
```
histogram(OCNMS_All_CTD, DO, 0.5)
```

```
## Warning: Removed 30 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



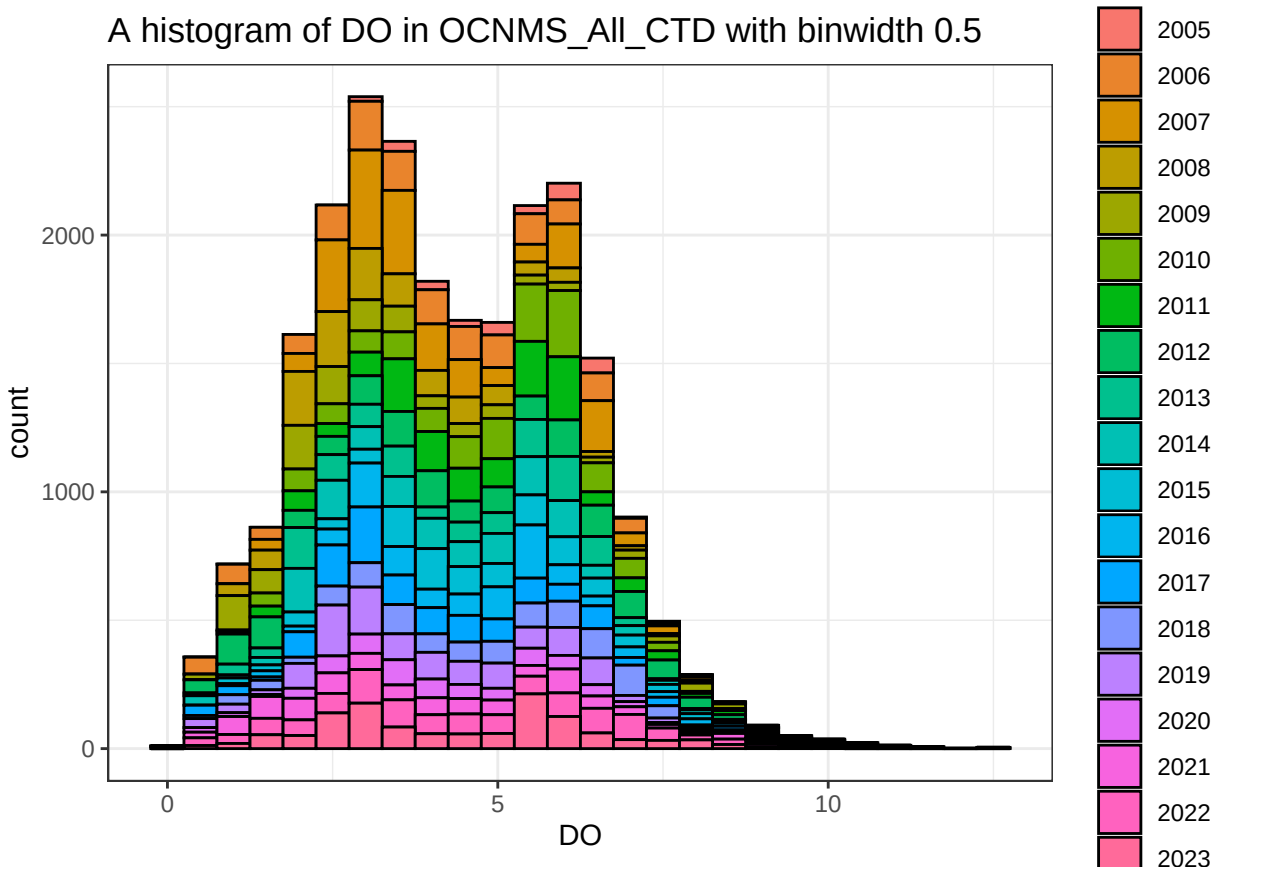

```
histogram(OCNMS_All_CTD, DO, 0.5) +
  facet_wrap(facets = vars(year))
```

```
## Warning: Removed 30 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



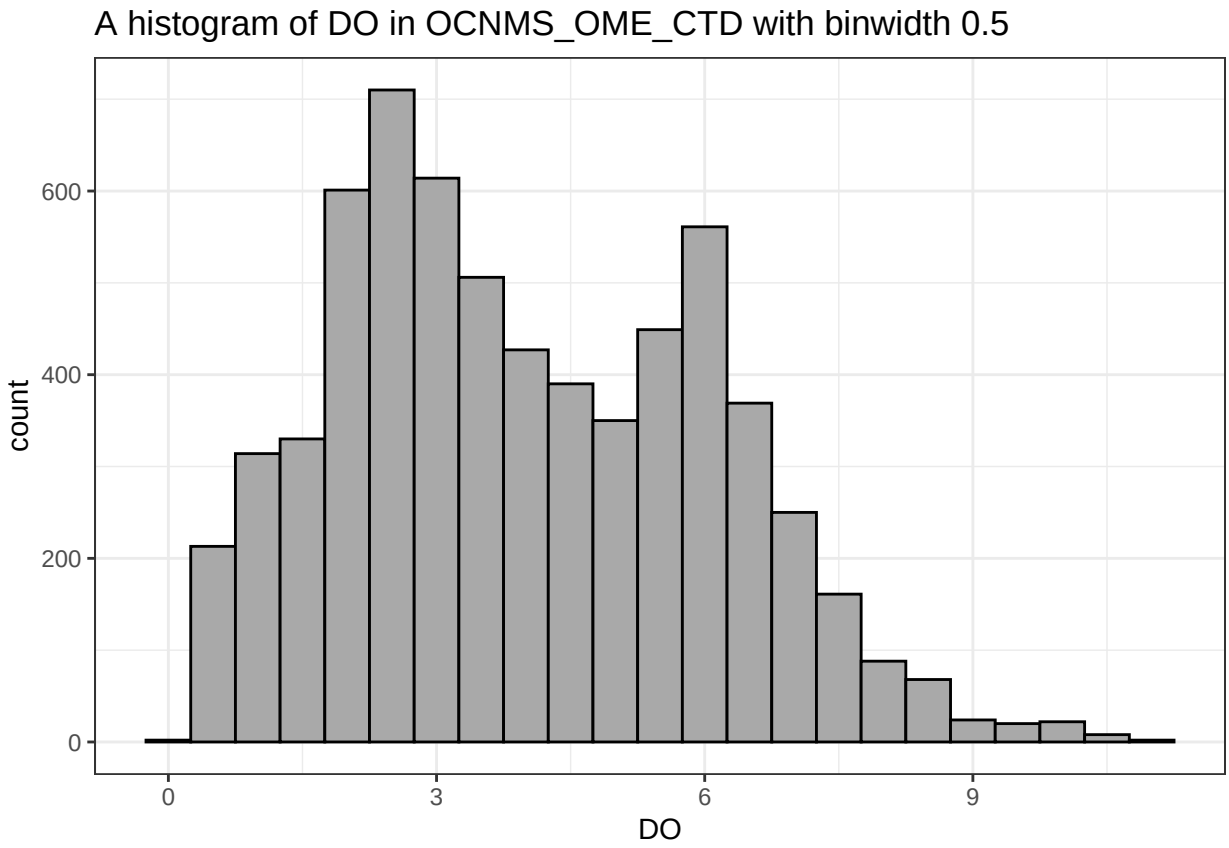

```
histogram_fill(OCNMS_All_CTD, DO, 0.5, fill = year)
```

```
## Warning: Removed 30 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



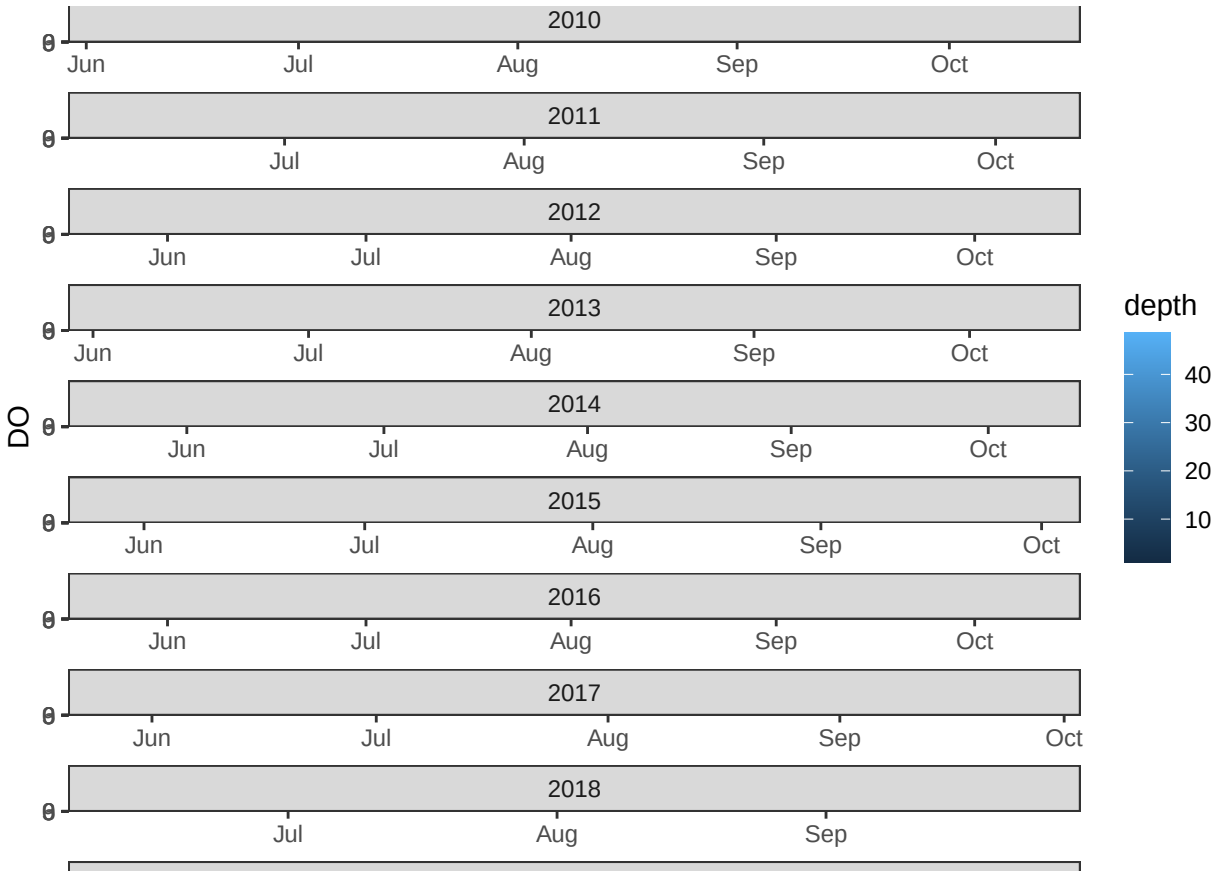

```
histogram(OCNMS_OME_CTD, DO, 0.5)
```

```
## Warning: Removed 7 rows containing non-finite outside the scale range
## (`stat_bin()`).
```




```
# Goal: One line plot per year of dissolved oxygen versus date

# Graph of depth, oxygen level, and date by year
OCNMS_OME_CTD %>%
  ggplot(aes(x = date, y = DO, color = depth)) +
    geom_point(shape = 1) +
    theme_bw() +
    facet_wrap(facets = vars(year), scales = "free_x", ncol = 1)
```

```
## Warning: Removed 7 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



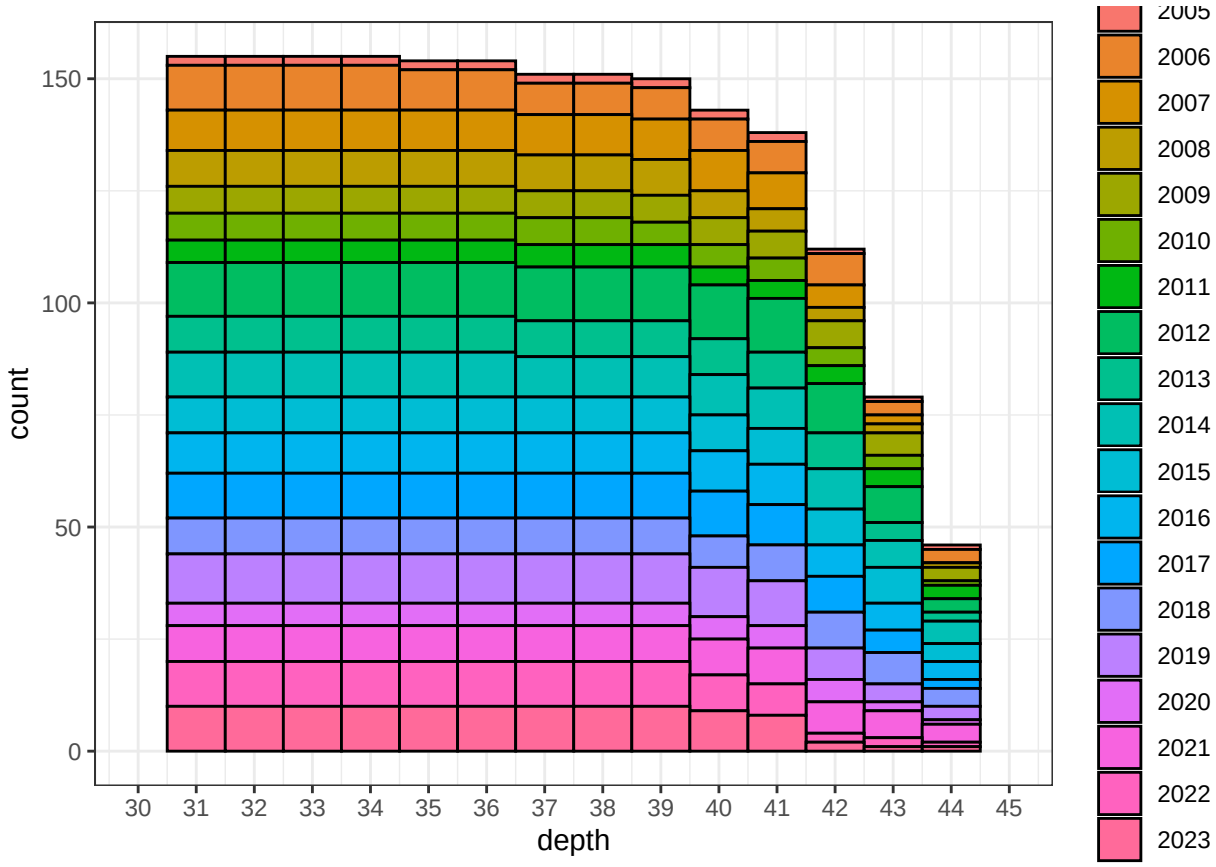

```
ggsave("OCNMS_CTD_Oxygen_FacetYear.png", path = here(pltpath), width = 1000, height =
↪  5000, units = "px")
```

```
## Warning: Removed 7 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
# What depths do we have?
depths <- OCNMS_OME_CTD %>%
  group_by(depth) %>%
  summarize(n = n())
# Specific depths do repeat, but there are a few around 42 meters. Perhaps near(depth,
↪  42, tol = 1)

ggplot(OCNMS_OME_CTD, aes(x = depth, fill = year)) +
  geom_histogram(binwidth = 1, color = "black") +
  scale_x_continuous(limits = c(30, 45), breaks = seq(from = 30, to = 45, by = 1)) +
  theme_bw()
```

```
## Warning: Removed 4562 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

```
## Warning: Removed 38 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



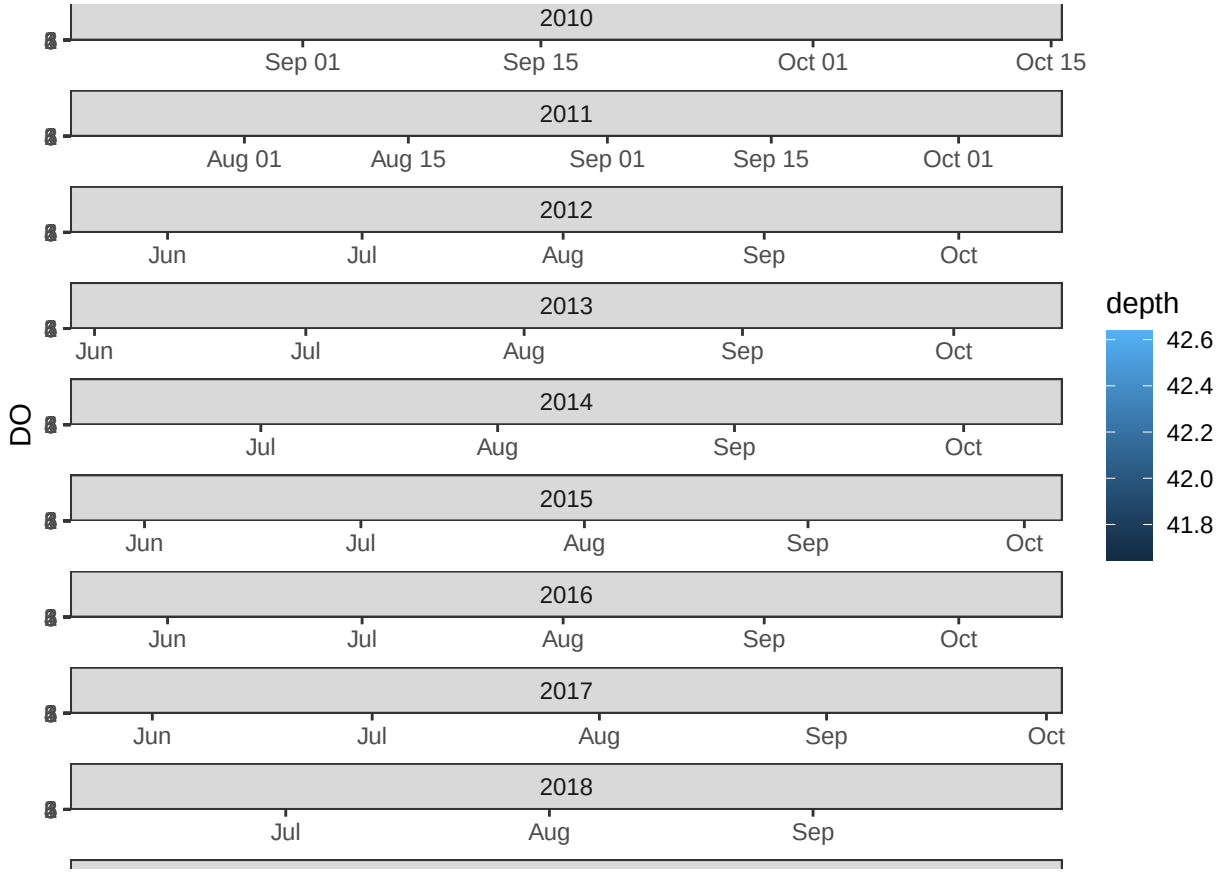

```
ggsave("OCNMS_CTD_Depths.png", path = here(pltpath), width = 3000, height = 2000, units =
↪  "px")
```

```
## Warning: Removed 4562 rows containing non-finite outside the scale range (`stat_bin()`).
## Removed 38 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```

```
# `if_else(condition, if_true_output, if_false_output, NA_val)` - used to transform
↪  vectors/data columns

# Logical vector of whether depth is within 1m of 42m
near42 <- near(OCNMS_OME_CTD$depth, 42, tol = 1)
sum(near42) # So the resulting DF should have 191 rows
```

```
## [1] 191
```

```
OCNMS_OME_CTD42m <- OCNMS_OME_CTD %>%
  filter(near(OCNMS_OME_CTD$depth, 42, tol = 1)) # 191 rows! Epic.

# Graph of oxygen levels around 42 meters for each year
OCNMS_OME_CTD42m %>%
  ggplot(aes(x = date, y = DO, color = depth)) +
    geom_line() +
    geom_point(shape = 1) +
    theme_bw() +
    facet_wrap(facets = vars(year), scales = "free_x", ncol = 1)
```




```
ggsave("OCNMS_CTD_Oxygen_FacetYear42m.png", path = here(pltpath), width = 1000, height =
↪  5000, units = "px")
```