Salmonid_Diet

Kaitlyn Kukula

2025-11-21

Guide to the data frames in analysis

salmonid_diet_processed: total counts of individual taxa observed per year and before/after dam removal for all salmon sampled chinook_diet: salmonid_diet_processed for Chinook only coho_diet: salmonid_diet_processed for Coho only salmon_taxa_totals: number of unique taxa for each year and species (used for plotting)

Setting up R session

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.1.0     v tidyr     1.3.1
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## here() starts at /Users/kaitlynkukula/KukulaHipp
##
##
## Attaching package: 'cowplot'
##
##
## The following object is masked from 'package:lubridate':
##
##     stamp

## [1] "/Users/kaitlynkukula/KukulaHipp"
```

Read in and clean data

```
salmonid_diet_raw <- read.csv("~/KukulaHipp/Data/Raw/Elwha_estuary_fish_diet/Elwha_estuary_fish_diet.csv
  mutate(Date = mdy(Date),
         Year = as.numeric(Year)) %>%
  janitor::clean_names()
```

Explore data

```
# colnames(salmonid_diet_raw)

# What fish did they study
#unique(salmonid_diet_raw$fish)

# What dates did they sample on
#unique(salmonid_diet_raw$date)
```

Wrangle data

```
# Summed all taxa consumed by salmon type in a year
salmonid_diet_processed <- salmonid_diet_raw %>%
  group_by(year, fish) %>%
  summarise(across(c(acari_larva:trichoptera_pupa), ~sum(., na.rm = TRUE))) %>%
  ungroup() %>%
  pivot_longer(c(acari_larva:trichoptera_pupa)) %>%
  pivot_wider(id_cols = c("fish", "name"),
              names_from = "year",
              values_from = "value",
              names_prefix = "value_") %>%
  mutate(taxa = str_extract(name, "^.+\\_") %>% str_remove_all("\\_")) %>%
  group_by(fish, taxa) %>%
  summarise(across(c(value_2006:value_2014), ~sum(., na.rm = TRUE))) %>%
  ungroup() %>%
  mutate(before_count = rowSums(across(value_2006:value_2007)),
         after_count = rowSums(across(value_2013:value_2014)),
         total_count = rowSums(across(before_count:after_count)))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
## `summarise()` has grouped output by 'fish'. You can override using the
## `.groups` argument.
```

Creating individual dataframes for Chinook and Coho

```
# Create a data frame for Chinook salmon diet
chinook_diet <- salmonid_diet_processed %>%
  filter(fish == "Chinook") %>%
  mutate(value_2006 = as.numeric(value_2006),
         value_2007 = as.numeric(value_2007),
         value_2013 = as.numeric(value_2013),
         value_2014 = as.numeric(value_2014))

# Create a data frame for Coho salmon diet
coho_diet <- salmonid_diet_processed %>% filter(fish == "Coho") %>%
  mutate(value_2006 = as.numeric(value_2006),
```

```
        value_2007 = as.numeric(value_2007),
        value_2013 = as.numeric(value_2013),
        value_2014 = as.numeric(value_2014))

### Create a data frame for the diet of all salmon species before dam removal
# Rows: taxa
# Columns: totals for each year
# all_salmon_diet_before <- salmonid_diet_processed %>%
#   group_by(taxa) %>%
#   summarise(across(c(value_2006:value_2007), ~sum(., na.rm = TRUE)))

### Create a data frame for the diet of all salmon species after dam removal
# Rows: taxa
# Columns: totals for each yearall_salmon_diet_after <- salmonid_diet_processed %>%
# all_salmon_diet_after <- salmonid_diet_processed %>%
# group_by(taxa) %>%
#   summarise(across(c(value_2013:value_2014), ~sum(., na.rm = TRUE)))
```

Analyze data

Is the diet of Chinook and Coho different between species and before and after dam removal?

1. Compare which taxa were the most consumed with a histogram for before and a second histogram for after Datasets: chinook_diet, coho_diet

2. Compare the diversity of taxa (did one species consume more types of taxa than the other?) before and after between the salmon species Datasets: chinook_diet, coho_diet

3. Compare the diversity of taxa consumed for *all* fish before and after the dam removal (did dam removal change the abundance and diversity of diet?) Datasets: all_salmon_diet_before, all_salmon_diet_after

4. Compare the diversity of taxa consumed for individual salmon species before and after the dam removal (did dam removal change the abundance and diversity of diet?)

Based on these results, it does appear that the dam removal caused changes to the top food sources for salmon.

What were the top taxa before and after the dam removal?

```
### Investigating the top taxa before
#
# # Creating a data frame with the top taxa before dam removal (considering year)
# chinook_top_taxa_before <-
#   # Create a combined dataset with the top 10 taxa from 2006 and top 10 taxa from 2007
#   full_join(chinook_taxa_2006 %>% head(20),
#             chinook_taxa_2007 %>% head(20),
#             by = c("fish", "taxa")) %>%
```

```
#    # Join original data to fill missing values
#    inner_join(chinook_diet %>%
#               select(fish:value_2007),
#             by = c("fish", "taxa")) %>%
#    # Select relevant columns
#    select(fish:taxa,
#           value_2006 = value_2006.y,
#           value_2007 = value_2007.y) %>%
#    # Pivot the data to create a plot
#    pivot_longer(c(value_2006:value_2007))
#
# # Plotting the data: although there are 14
# chinook_top_taxa_before_plot <- ggplot(data = chinook_top_taxa_before %>% head(20),
#                                      aes(x = reorder(taxa, -value),
#                                          y = value,
#                                          fill = name)) +
#    geom_bar(stat = "identity", position = "stack") +
#    theme(axis.text.x = element_text(angle = 45, hjust = 1))
# chinook_top_taxa_before_plot
```

Chironomidae dominate the diet before removal with consistent observations between the two years, followed by hilara and whatever that next thing is.

Investigating the top taxa after

```
# Creating a data frame with the top taxa after dam removal (considering year)
# chinook_top_taxa_after <-
#    # Create a combined dataset with the top 10 taxa from 2013 and top 10 taxa from 2014
#    full_join(chinook_taxa_2013 %>% head(20),
#            chinook_taxa_2014 %>% head(20),
#            by = c("fish", "taxa")) %>%
#    # Join original data to fill missing values
#    inner_join(chinook_diet %>%
#               select(fish:taxa, value_2013:value_2014),
#             by = c("fish", "taxa")) %>%
#    # Select relevant columns
#    select(fish:taxa,
#           value_2013 = value_2013.y,
#           value_2014 = value_2014.y) %>%
#    # Pivot the data to create a plot
#    pivot_longer(c(value_2013:value_2014))
#
# # Plotting the data
# chinook_top_taxa_after_plot <- ggplot(data = chinook_top_taxa_after %>% head(20),
#                                      aes(x = reorder(taxa, -value),
#                                          y = value,
#                                          fill = name)) +
#    geom_bar(stat = "identity", position = "stack") +
#    theme(axis.text.x = element_text(angle = 45, hjust = 1))
# chinook_top_taxa_after_plot
```

Although the diversity of taxa consumed by salmon after the dam removal increased, the diet of salmon is dominated by six groups (cyclopoida, chironomidae, and osteichthyes, whatever that is, chironomini, and tanypodinae).

4

Of the top taxa, which are present before and after removal and how did their occurrences change?

```
# # Create a data frame that finds which taxa overlap between the TOP before and after taxa
# chinook_top_taxa_overlap <- inner_join(chinook_top_taxa_before %>%
#                                        pivot_wider(names_from = "name", values_from = "value") %>%
#                                        mutate(before_count = rowSums(across(value_2006:value_2007))),
#                                      chinook_top_taxa_after %>%
#                                        pivot_wider(names_from = "name", values_from = "value") %>%
#                                        mutate(after_count = rowSums(across(value_2013:value_2014))),
#                                      by = "taxa") %>%
#   select(fish.x:value_2007, value_2013:value_2014) %>%
#   pivot_longer(c(value_2006:value_2014))
# view(chinook_top_taxa_overlap)
#
#
#
# # Plotting total before and after counts for the taxa that were present in before and after
# chinook_top_taxa_before_overlap_plot <- ggplot(data = chinook_top_taxa_overlap,
#                                       aes(x = taxa,
#                                           y = value,
#                                           fill = name)) +
#   geom_bar(stat = "identity", position = "dodge") +
#   theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
#   labs(title = "Changes to occurence of most common taxa present before and after")
# chinook_top_taxa_before_overlap_plot

# chinook_top_taxa_after_overlap_plot <- ggplot(data = chinook_top_taxa_overlap %>% filter(name == c("v
#                                       aes(x = taxa,
#                                           y = value,
#                                           fill = name)) +
#   geom_bar(stat = "identity", position = "stack") +
#   theme(axis.text.x = element_text(angle = 45, hjust = 1))
# chinook_top_taxa_after_overlap_plot
#
# chinook_top_taxa_overlap_plot <-
```

Certain taxa experienced a large change in occurrences after dam removal, but there is not a clear trend. This may be due to availability of food sources, among other things.

What happened to the taxa observed before dam removal?

```
# Plotting
# chinook_top_taxa_before_result <- left_join(chinook_top_taxa_before %>%  pivot_wider(names_from = "na
#                                                                                     values_from = "value")
#                                             head(10),
#                                           chinook_diet) %>% pivot_longer(c(value_2006:value_2014))
#
# chinook_top_taxa_before_result_plot <- ggplot(data = chinook_top_taxa_before_result,
```

```
#                                               aes(x = taxa,
#                                                   y = value,
#                                                   fill = name)) +
#    geom_bar(stat = "identity", position = "dodge") +
#    theme(axis.text.x = element_text(angle = 45, hjust = 1))
# chinook_top_taxa_before_result_plot
```

What were the levels of the taxa observed after like before removal?

```
# chinook_top_taxa_after_result <- left_join(chinook_top_taxa_after %>%
#                                              pivot_wider(names_from = "name",
#                                                          values_from = "value"),
#                                            chinook_diet) %>%
#    pivot_longer(c(value_2013:value_2007))
#
# chinook_top_taxa_after_result_plot <- ggplot(data = chinook_top_taxa_after_result,
#                                              aes(x = taxa,
#                                                  y = value,
#                                                  fill = name)) +
#    geom_bar(stat = "identity", position = "dodge") +
#    theme(axis.text.x = element_text(angle = 45, hjust = 1))
# chinook_top_taxa_after_result_plot
```

Of all taxa consumed by Chinook before the dam removal, how many new taxa were present in samples after removal? In other words, what percentage of the taxa observed in after-removal samples were new taxa?

Of all taxa consumed by Chinook after the dam removal, what percentage were present in samples before removal? In other words, what taxa were not consumed before the dam removal?

Suggested final diet code

How many unique taxa were consumed by each species before and after the dam removal?

```
salmon_taxa_totals_plot <- ggplot(data = data.frame(
  Salmon = c(rep("Chinook", 4), rep("Coho", 4)),
  Year = c(rep(c("2006", "2007", "2013", "2014"),2)),
  Count = c(sum(chinook_diet$value_2006 > 0),
            sum(chinook_diet$value_2007 > 0),
            sum(chinook_diet$value_2013 > 0),
```

```
            sum(chinook_diet$value_2014 > 0),
            sum(coho_diet$value_2006 > 0),
            sum(coho_diet$value_2007 > 0),
            sum(coho_diet$value_2013 > 0),
            sum(coho_diet$value_2014 > 0))),
  aes(x = Year, y = Count, fill = as.factor(Salmon))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Number of Dietary Taxa Observed in Salmon Diets",
       x = NULL,
       y = "Number of Species Observed in Salmon",
       fill = "Salmon Species") +
  scale_fill_manual(values = c("Chinook" = "#4d9221", "Coho" = "#91bfdb"))
salmon_taxa_totals_plot
```


Number of Dietary Taxa Observed in Salmon Diets
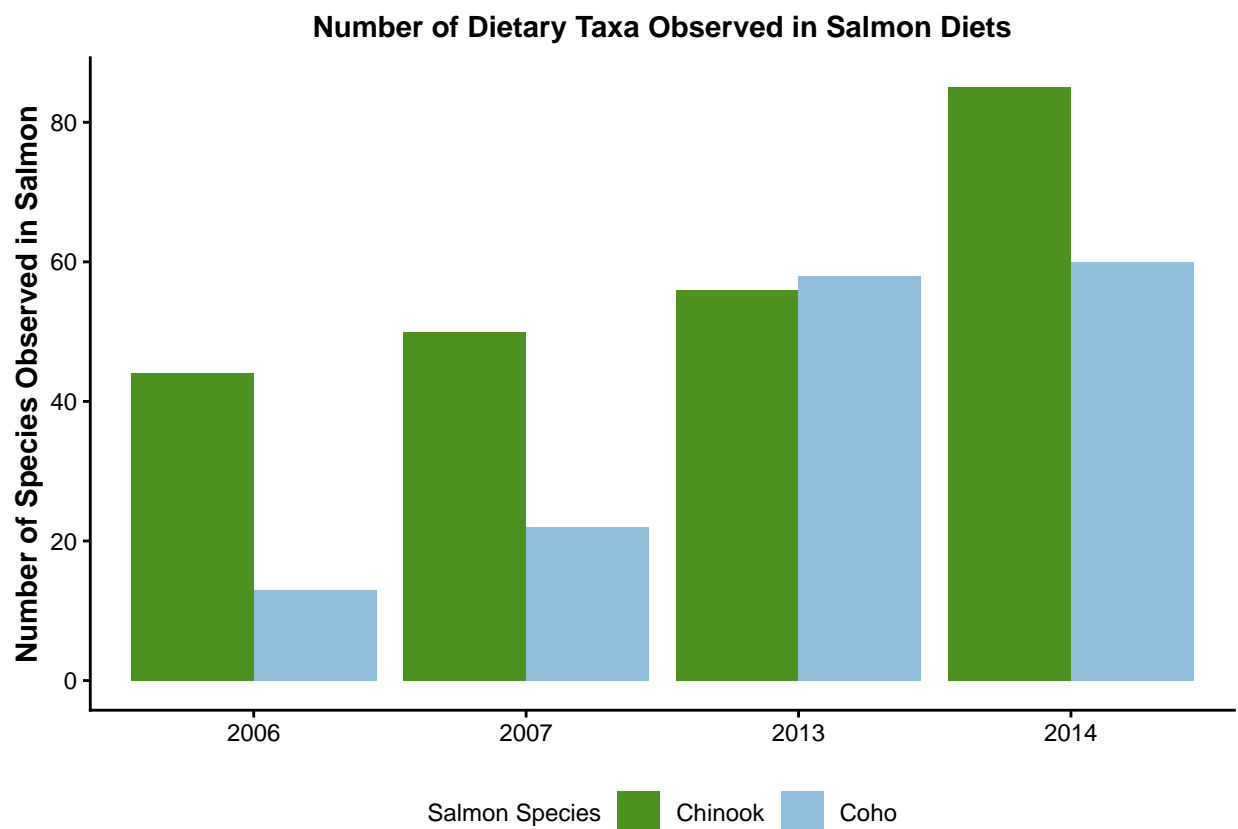
There was some increase in the diversity of taxa consumed by Chinook salmon after dam removal. In Coho salmon, the diversity of taxa consumed greatly increased after dam removal.

Taxa consumed by Chinook species yearly (exploratory plots)

```
# Chinook taxa in 2006
chinook_taxa_2006 <- chinook_diet %>%
  select(fish:value_2006) %>%
  arrange(desc(value_2006)) %>%
```

```r
  filter(value_2006 > 0)

chinook_top_taxa_plot_2006 <- ggplot(data = chinook_taxa_2006 %>% head(10),
                                     aes(x = reorder(taxa, -value_2006), y = value_2006)) +
  geom_bar(stat = "identity", fill = "black") +
  theme(axis.text.x = element_text(angle = 25, hjust = 1)) +
  labs(x = NULL,
       y = NULL,
       title = "2006")
#chinook_top_taxa_plot_2006

# Chinook taxa in 2007
chinook_taxa_2007 <- chinook_diet %>%
  select(fish:taxa, value_2007) %>%
  arrange(desc(value_2007)) %>%
  filter(value_2007 > 0)

chinook_top_taxa_plot_2007 <- ggplot(data = chinook_taxa_2007 %>% head(n = 10),
                                     aes(x = reorder(taxa, -value_2007), y = value_2007)) +
  geom_bar(stat = "identity", fill = "black") +
  theme(axis.text.x = element_text(angle = 25, hjust = 1)) +
  labs(x = NULL,
       y = NULL,
       title = "2007")
#chinook_top_taxa_plot_2007

# Chinook in 2013

chinook_taxa_2013 <- chinook_diet %>%
  select(fish:taxa, value_2013) %>%
  arrange(desc(value_2013)) %>%
  filter(value_2013 > 0)

chinook_top_taxa_plot_2013 <- ggplot(data = chinook_taxa_2013 %>% head(n = 10),
                                     aes(x = reorder(taxa, -value_2013), y = value_2013)) +
  geom_bar(stat = "identity", fill = "black") +
  theme(axis.text.x = element_text(angle = 25, hjust = 1)) +
  labs(x = NULL,
       y = NULL,
       title = "2013")
#chinook_top_taxa_plot_2013

# Chinook in 2014

chinook_taxa_2014 <- chinook_diet %>%
  select(fish:taxa, value_2014) %>%
  arrange(desc(value_2014)) %>%
  head(n = 10)

chinook_top_taxa_plot_2014 <- ggplot(data = chinook_taxa_2014,
                                     aes(x = reorder(taxa, -value_2014), y = value_2014)) +
  geom_bar(stat = "identity", fill = "black") +
  theme(axis.text.x = element_text(angle = 25, hjust = 1)) +
```
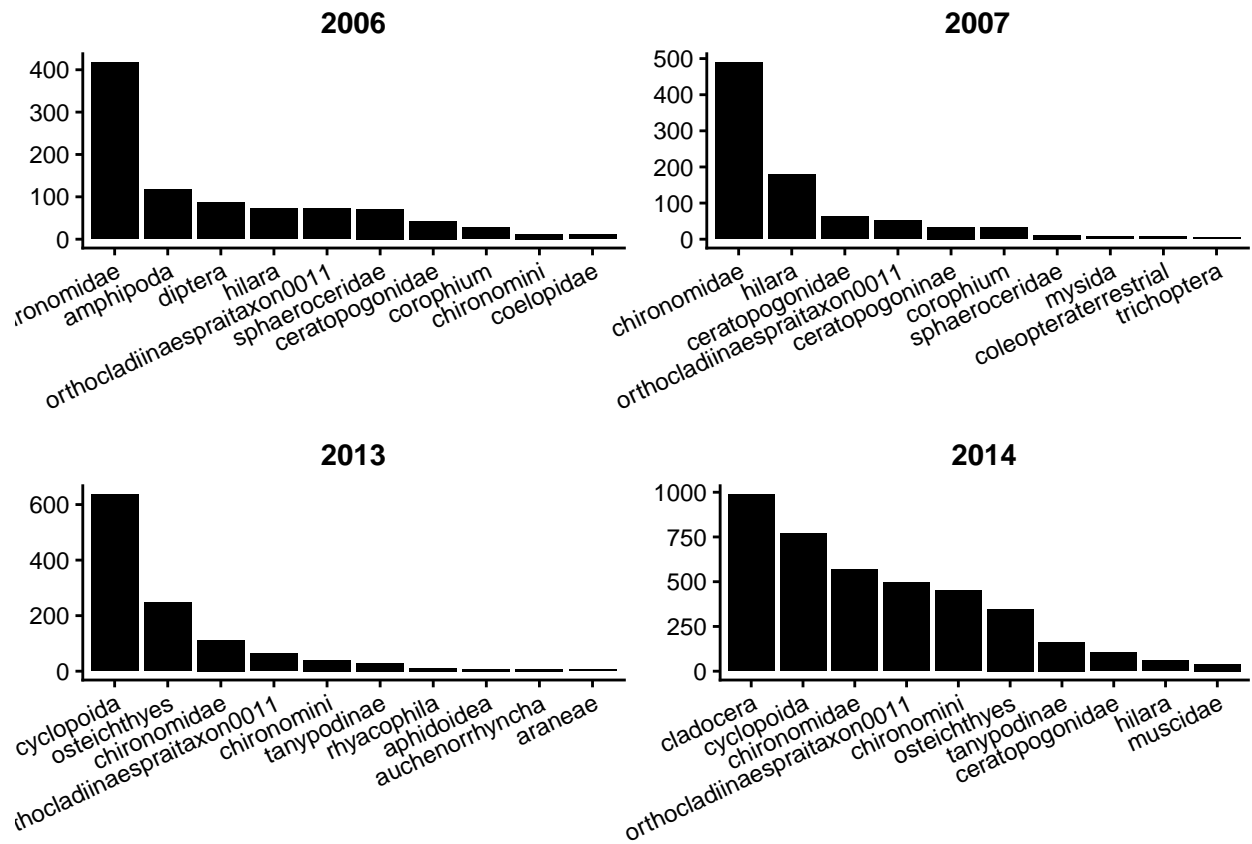
```
    labs(x = NULL,
         y = NULL,
         title = "2014")
#chinook_top_taxa_plot_2014

chinook_diet_4x4 <- plot_grid(chinook_top_taxa_plot_2006,
                              chinook_top_taxa_plot_2007,
                              chinook_top_taxa_plot_2013,
                              chinook_top_taxa_plot_2014,
                              nrow = 2,
                              ncol = 2)
chinook_diet_4x4
```


Plotting the changes in Chinook taxa

```
# Overall
chinook_top_taxa <- salmonid_diet_processed %>%
  filter(fish == "Chinook") %>%
  arrange(desc(total_count)) %>%
  head(10) %>%
  select(fish:taxa, before_count:after_count) %>%
  pivot_longer(cols = c(before_count, after_count))
```

```r
chinook_top_taxa_plot <- ggplot(chinook_top_taxa, aes(x = taxa, y = value, fill = as.factor(name))) +
  geom_bar(stat = "identity", position = "dodge") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Top Dietary Taxa",
       x = NULL,
       y = "Count",
       fill = NULL) +
  scale_fill_manual(values = c("before_count" = "#4d9221", "after_count" = "#91bfdb"),
                    labels = c("Before", "After")) +
  theme(legend.position = "none")
# chinook_top_taxa_plot

# Before
chinook_top_before_taxa <- salmonid_diet_processed %>%
  filter(fish == "Chinook") %>%
  select(taxa, before_count) %>%
  arrange(desc(before_count)) %>%
  head(10) %>%
  left_join(chinook_diet) %>%
  select(taxa:before_count, after_count) %>%
  pivot_longer(c("before_count", "after_count"))
```

Joining with 'by = join_by(taxa, before_count)'

```r
chinook_top_before_taxa_plot <- ggplot(chinook_top_before_taxa, aes(x = taxa, y = value, fill = as.factor
  geom_bar(stat = "identity", position = "dodge") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Top Before Removal Dietary Taxa",
       x = NULL,
       y = "Count",
       fill = NULL) +
  scale_fill_manual(values = c("before_count" = "#4d9221", "after_count" = "#91bfdb"),
                    labels = c("Before", "After")) +
  theme(legend.position = "none")
# chinook_top_before_taxa_plot

# After
chinook_top_after_taxa <- salmonid_diet_processed %>%
  filter(fish == "Chinook") %>%
  select(taxa, after_count) %>%
  arrange(desc(after_count)) %>%
  head(10) %>%
  left_join(chinook_diet) %>%
  select(taxa:after_count, before_count) %>%
  pivot_longer(c("before_count", "after_count"))
```

Joining with 'by = join_by(taxa, after_count)'

```r
chinook_top_after_taxa_plot <- ggplot(chinook_top_after_taxa, aes(x = taxa, y = value, fill = as.factor
  geom_bar(stat = "identity", position = "dodge") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Top After Removal Dietary Taxa",
```
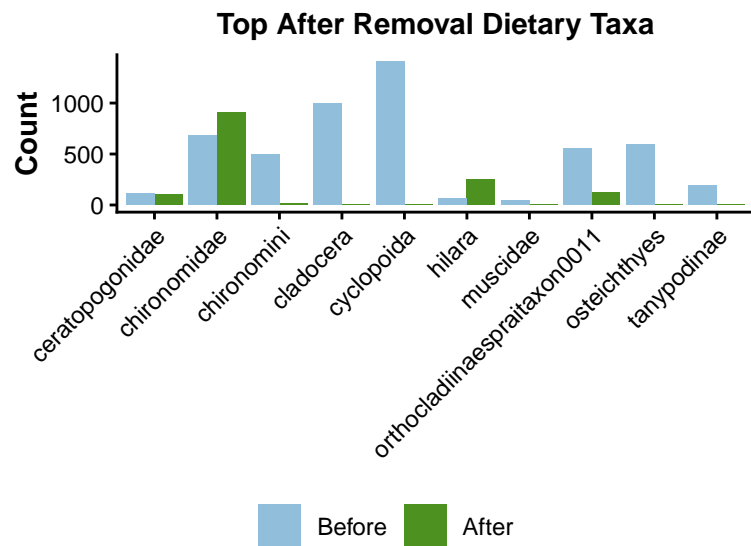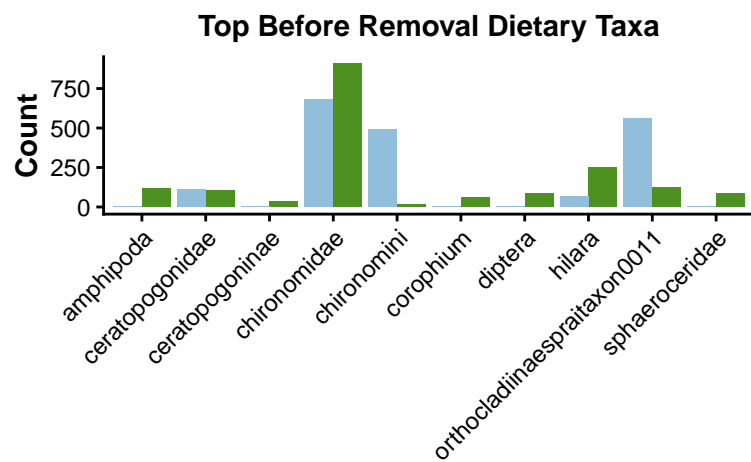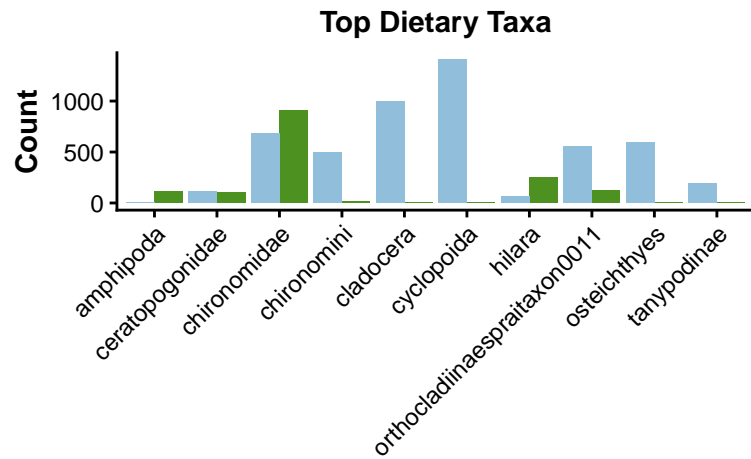
```r
      x = NULL,
      y = "Count",
      fill = NULL) +
  scale_fill_manual(values = c("before_count" = "#4d9221", "after_count" = "#91bfdb"),
                    labels = c("Before", "After"))+
  theme(legend.position = "none")
# chinook_top_after_taxa_plot

dam_state_legend <- get_legend(ggplot(chinook_top_after_taxa, aes(x = taxa, y = value, fill = name)) +
  geom_bar(stat = "identity", position = "dodge") +
    labs(fill = NULL) +
    scale_fill_manual(values = c("before_count" = "#4d9221",
                                 "after_count" = "#91bfdb"),
                      labels = c("Before", "After")))

chinook_top_taxa_3x3 <- plot_grid(chinook_top_taxa_plot,
                                  chinook_top_before_taxa_plot,
                                  chinook_top_after_taxa_plot,
                                  dam_state_legend,
                                  nrow = 4,
                                  rel_heights = c(0.5, 0.5, 0.5, 0.1))
chinook_top_taxa_3x3
```

Top Dietary Taxa

Top Before Removal Dietary Taxa

Top After Removal Dietary Taxa

Before After

NULL

```r
# salmon_taxa_totals_plot
#
# chinook_diet_4x4
```

```
#
# chinook_top_taxa_plot

# chinook_top_taxa_plot
# chinook_top_before_taxa_plot
# chinook_top_after_taxa_plot
```

Is there a relationship between observed invertebrate counts and salmonid diet?

After wrangling the invert data, use inner_join to combine with diets so you only grab similar taxa