HW 3 Drafting Viz

AUTHOR
Ian Morris-Sibaja

PUBLISHED
February 28, 2025

Can't spell Disaster without… Pisaster??

Pisaster are keystone species in the intertidal zone. They are a type of sea star that are known to be voracious predators of mussels. When Pisaster are present, they can control the population of mussels, which in turn can have cascading effects on the rest of the ecosystem.

1. Which option do you plan to pursue? It's okay if this has changed since HW #1.

I plan to remain pursuing option 1.

2. Restate your question(s). Has this changed at all since HW #1? If yes, how so?

The overall question of: How have Pisaster, a Sea Star Genus, been affected by changing environments over the past 20 years? My three subquestions include: How has the abundance of Pisaster species changed over time in California?, How has the distribution of Pisaster species changed over time in California?, and How has the presence of Pisaster species changed over time in California?

This has been a more specified question since HW #1. I have decided to focus on Pisaster species specifically, as they are a keystone species in the intertidal zone, rather than all coastal species.

3. Explain which variables from your data set(s) you will use to answer your question(s), and how.

I plan to clean my data to species, year, longitude, and latitude. I will then group by and summarise totals to get a count of each species by year and bin both year and latitude. Then I will plot these to visualize their changes.

4. Inspirational Visualizations and explain which elements you might borrow

WE MUST PROTECT
MARINE ENVIRONMENTS

The ocean covers more than 70% of planet Earth and **supports humankind.**
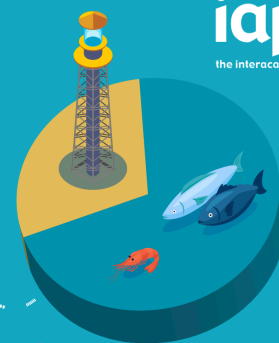
OXYGEN TO BREATH

FISH AND SEAFOOD TO EAT
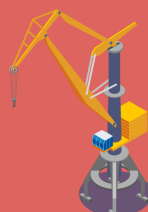
SEAWAYS FOR TRANSPORTATION

LEISURE AND HEALING PLACES TO VISIT

Unregulated and excessive human activities and climate change are **reducing biodiversity** and **threatening ecosystem services.**

HABITAT DEGRADATION AND LOSS
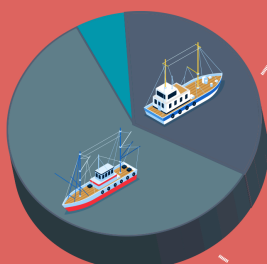
COASTAL DEVELOPMENT

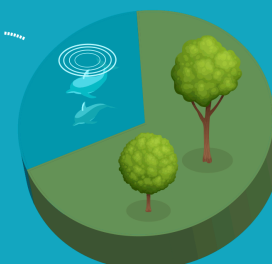**INVASIVE SPECIES**

ENVIRONMENTAL CONTAMINANTS

CLIMATE CHANGE

Most fisheries stocks are fully exploited: more than 90% of marine stocks are either overfished (34.2%) or fished at maximum sustainable levels (59.6%).
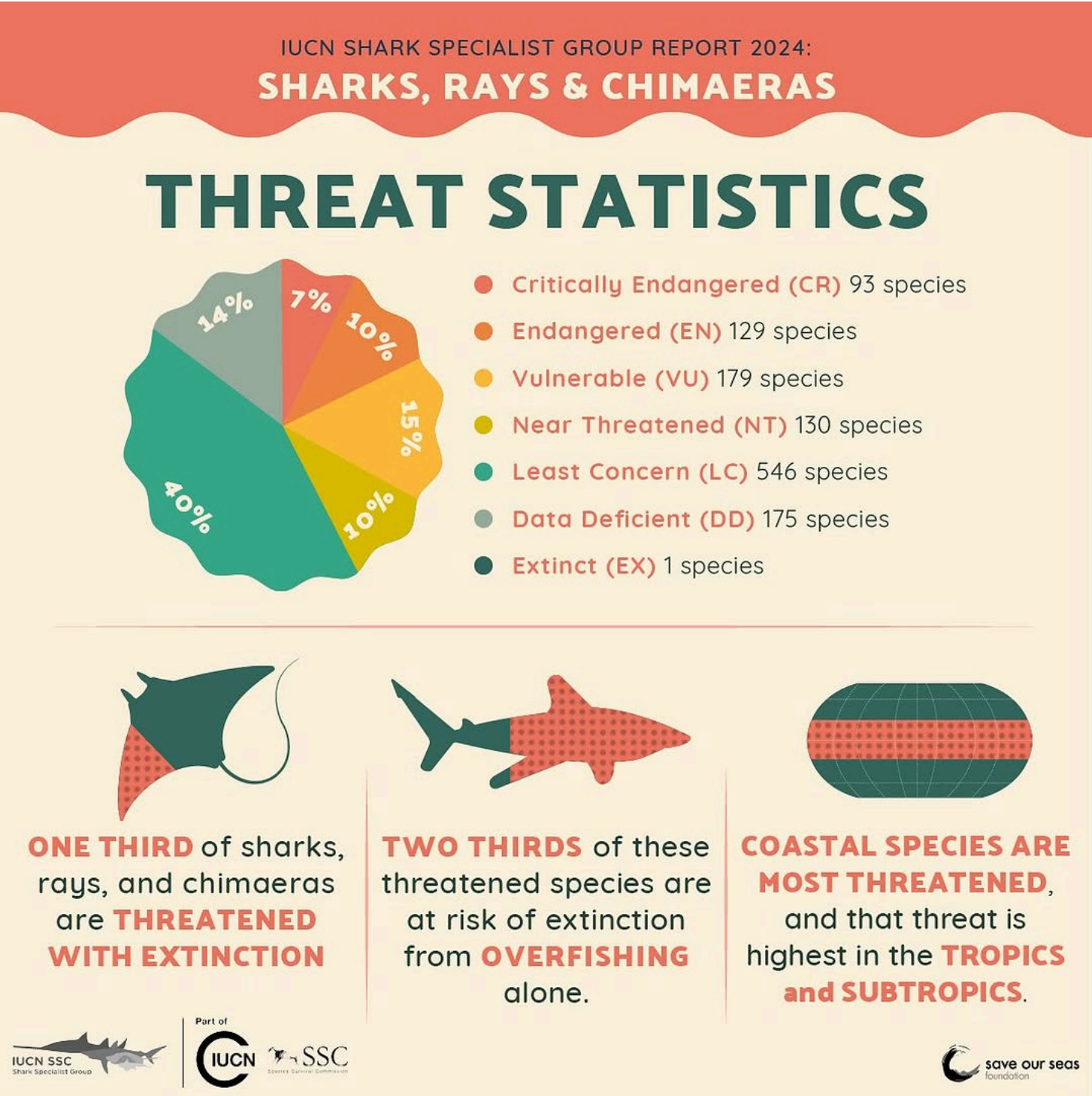
The ocean has absorbed 30% of total anthropogenic CO_2 emissions.

WORLD ACADEMIES
URGE WORLD LEADERS
TO IMPROVE OCEAN HEALTH

by stopping **habitat destruction** and the spread of **environmental contaminants**, fighting **climate change** and **overexploitation**, and adopting **science-based policies.**

STOP POLLUTION

I like how the colors are seperating the brochure in 3rds. It is pleasing to the eyes and will work well with three subquestions.

I like how the colors fill up the animal outline, I think it can help me portay abundance wihin this context.

5. Hand-draw your anticipated visualizations
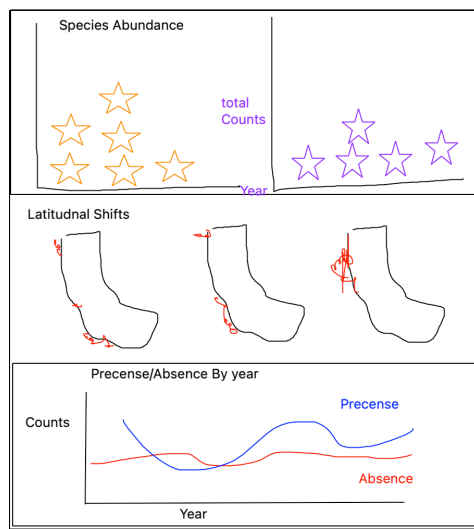
Mock Up

Set Up

Data Import

```
# Read in three excel files from MARINe biodiversity data
point_contact_raw <- read_excel(here('data', 'MARINe_biodiversity_data',
                                      'cbs_data_CA_2023.xlsx'), sheet = 'point_contact_summary
quadrat_raw <- read_excel(here('data', 'MARINe_biodiversity_data',
                               'cbs_data_CA_2023.xlsx'), sheet = 'quadrat_summary_data')
swath_raw <- read_excel(here('data', 'MARINe_biodiversity_data',
                             'cbs_data_CA_2023.xlsx'), sheet = 'swath_summary_data')

# Read in Dangermond preserve shape file
dangermond <- read_sf(here('data', 'dangermond_shapefile', 'jldp_boundary.shp'))

# Read in California state boundary
california <- spData::us_states %>%
  filter(NAME == "California")
```

Data Cleaning

Lots of cleaning based on our capston project

```
# Clean point_contact dataset
point_contact_clean <- point_contact_raw %>%
  # Remove non-matching columns
  select(!c('number_of_transect_locations', 'percent_cover')) %>%
  # Rename num of hits to total count
  rename(total_count = number_of_hits) %>%
  # Create new data collection source column
  mutate(collection_source = "point contact") %>%
```

```
  # Filter to mainland only
  filter(island == "Mainland") %>%
  # Remove certain species lumps
  filter(!species_lump %in% c("Rock", "Sand", "Tar", "Blue Green Algae", "Red Crust", "Diatom

# Clean quadrat dataset
quadrat_clean <- quadrat_raw %>%
  # Remove non-matching columns
  select(!c('number_of_quadrats_sampled', 'total_area_sampled_m2', 'density_per_m2')) %>%
  # Create new data collection source column
  mutate(collection_source = "quadrat") %>%
  # Filter to mainland only
  filter(island == "Mainland") %>%
  # Remove certain species lumps
  filter(!species_lump %in% c("Rock", "Sand", "Tar", "Blue Green Algae", "Red Crust", "Diatom

# Clean swath dataset
swath_clean <- swath_raw %>%
  # Remove non-matching columns
  select(!c('number_of_transects_sampled', 'est_swath_area_searched_m2',  'density_per_m2'))
  # Create new data collection source column
  mutate(collection_source = "swath") %>%
  # Filter to mainland only
  filter(island == "Mainland") %>%
  # Remove certain species lumps
  filter(!species_lump %in% c("Rock", "Sand", "Tar", "Blue Green Algae", "Red Crust", "Diatom
```

Merge datasets

Merge data sets for easy calculations

```
# Merge the 3 dataset together
biodiv_merge <- bind_rows(point_contact_clean, quadrat_clean, swath_clean) %>%
  filter(year < 2021)
```

```
# Group by site and species (no year)
genus_sum <- biodiv_merge %>%
  mutate(genus = word(species_lump)) %>%
  group_by(genus, species_lump, year, longitude, latitude) %>%
  summarise(num_count = sum(total_count)) %>%
  # Create column to indicate presence/absence
  mutate(presence = ifelse(num_count >= 1, 1, 0))
```

California Data

```
# Convert to WGS84 to lat long
california <- st_transform(california, crs = 4326)
```

Convert biodiv data to sf object

```
genus_sum_geo <- genus_sum %>%
  filter(presence == 1) %>%
```

```
  st_as_sf(coords = c("longitude", "latitude"), crs = st_crs(california), remove = FALSE)

# Check that the crs matches
if(st_crs(california) == st_crs(genus_sum_geo)) {
  print("The coordinate reference systems match")
} else {
  print("The coordinate reference systems do NOT match. Transformation of CRS is recommended.
}
```

```
[1] "The coordinate reference systems match"
```

Filter Genus

```
genus_count <- genus_sum %>%
  group_by(genus) %>%
  summarise(
    total_count = sum(num_count)
  ) %>%
  arrange(desc(total_count)) %>%
  slice_max(order_by = total_count, n = 10) %>%
  arrange(total_count)
```

Pisaster Abundance Plot

```
pisaster_sum <- biodiv_merge %>%
  mutate(genus = word(species_lump)) %>%
  filter(genus == "Pisaster") %>%
  group_by(species_lump, year, longitude, latitude) %>%
  summarise(num_count = sum(total_count)) %>%
  # Create column to indicate presence/absence
  mutate(presence = ifelse(num_count >= 1, 1, 0))
```

```
pisaster_sum_year <- pisaster_sum %>%
  group_by(year, species_lump) %>%
  summarise(
    total_count = sum(num_count)
  ) %>%
  arrange(year)
```

```
pisaster_sum_geo <- pisaster_sum %>%
  filter(presence == 1) %>%
  st_as_sf(coords = c("longitude", "latitude"), crs = st_crs(california), remove = FALSE)
```

```
theme_ocean <- theme_minimal(base_size = 14) +
  theme(
    panel.background = element_rect(fill = "#b3e5fc", color = NA),  # Light ocean blue backgr
    panel.grid.major = element_line(color = "#b3e5fc", linetype = 0),  # Soft wave-like grid
    panel.grid.minor = element_blank(),
    legend.background = element_rect(fill = "#ffffff", color = NA),
    legend.position = "right",
    axis.text = element_text(size = 12, color = "#01579b"),  # Deep ocean blue text
```

```r
    axis.title = element_text(size = 14, face = "bold", color = "#004d40"),
    plot.title = element_text(size = 18, face = "bold", color = "#004d40", hjust = 0.5),
    plot.subtitle = element_text(size = 14, color = "#004d40", hjust = 0.5)
  )
```

```r
pisaster_sum_perc <- pisaster_sum_year

pisaster_sum_perc$year_bin <- cut(pisaster_sum_perc$year,
                                  breaks = c(2001, 2005, 2009, 2013, 2017, 2021),
                                  include.lowest = TRUE,
                                  right = FALSE,
                                  labels = c("2001-2004",
                                             "2005-2008",
                                             "2009-2012",
                                             "2013-2016",
                                             "2017-2020"))

# Compute total_count_sum first using summarise()
pisaster_sum_perc <- pisaster_sum_perc %>%
  group_by(species_lump, year_bin) %>%
  summarise(total_count_sum = sum(total_count, na.rm = TRUE), .groups = "drop")

# Plot with corrected normalized count
ggplot(pisaster_sum_perc, aes(x = year_bin, y = total_count_sum, fill = species_lump)) +
  geom_col(position = "dodge") +  # Dodge to see separate species
  scale_fill_manual(values = c("Pisaster ochraceus" = "#6A0DAD", "Pisaster giganteus" = "#FF8
  labs(
    title = "Sea Star Abundance Trends Over Time",
    subtitle = "Fluctuations occur in Pisaster species, a genus of Sea Star",
    x = "Year",
    y = "Percentage of Total Counts",
    fill = "Species"
  ) +
  theme_ocean +
  theme(
    panel.grid.minor = element_blank(),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5)
  ) +
  facet_wrap(~species_lump, scales = "free_y", ncol = 1, nrow = 2)
```
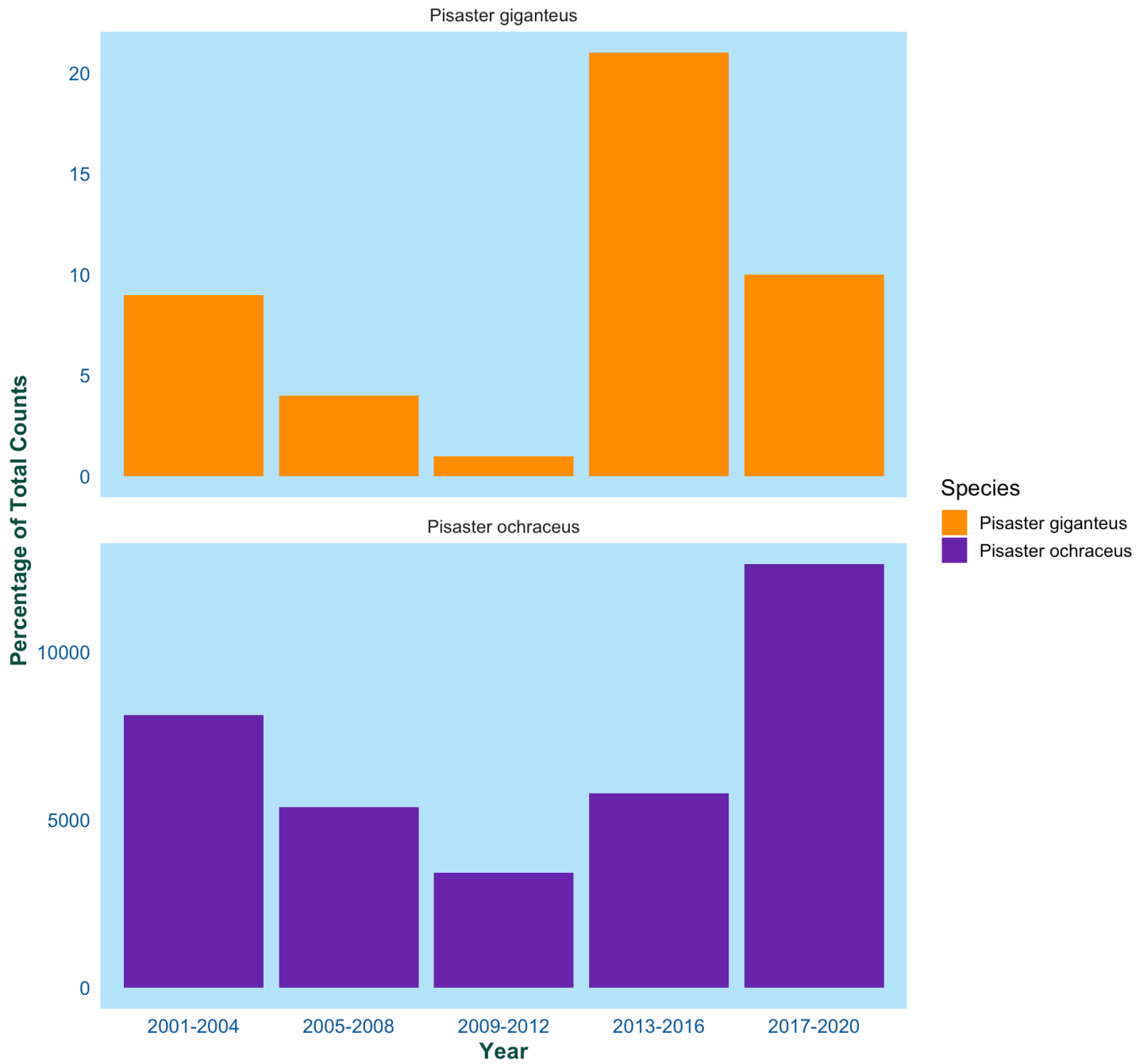
Sea Star Abundance Trends Over Time

Fluctuations occur in Pisaster species, a genus of Sea Star

Latitudinal Shift Map

```
pisaster_sum <- pisaster_sum
```

```
pisaster_sum$lat_bin <- cut(pisaster_sum$latitude,
                        breaks = seq(32, 42, 2),
                        include.lowest = TRUE,
                        labels = c("32 - 33",
                                   "34 - 35",
                                   "36 - 37",
                                   "38 - 39",
                                   "40 - 41"))
```

```
pisaster_sum$year_bin <- cut(pisaster_sum$year,
                             breaks = c(2001, 2005, 2009, 2013, 2017, 2021),
                             include.lowest = TRUE,
                             right = FALSE,
                             labels = c("2001-2004",
                                        "2005-2008",
                                        "2009-2012",
                                        "2013-2016",
                                        "2017-2020"))
```

```
ggplot(pisaster_sum, aes(x = latitude, y = year_bin, color = year_bin)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01,
                               fill = NA) +
    scale_color_cyclical(name = "year_bin", guide = "legend",
                         values = c("#6A0DAD", "#FF8C00")) +
  theme_minimal(base_size = 14) +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    legend.position = "none",
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5)
  ) +
  labs(title = "Latitudinal Shift of Pisaster Over 20 Years",
       x = "Latitude",
       y = "Year",
       fill = "Density")
```
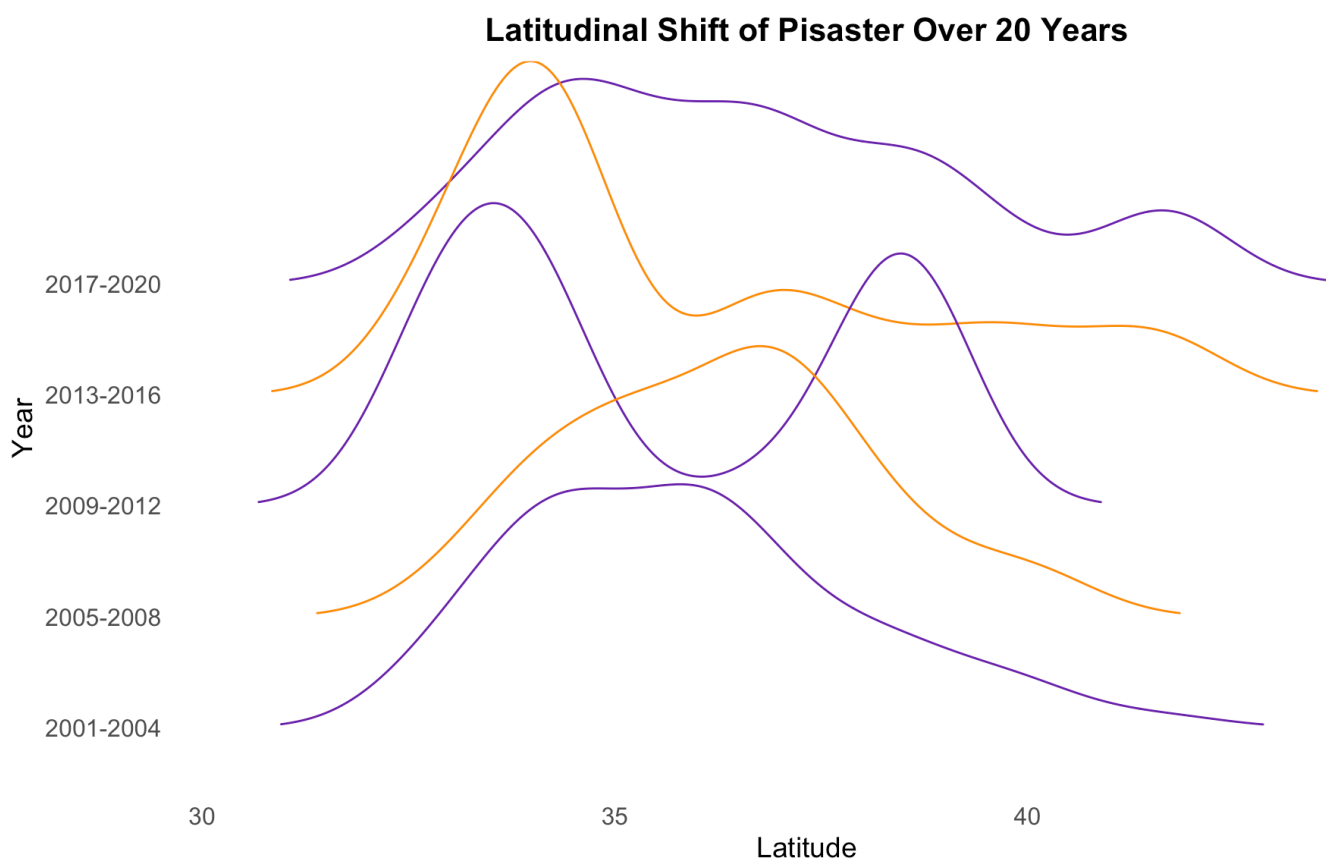
Abundance/Presence Plot

```r
# Ensure presence is a factor and year_bin is in the right format
pisaster_sum$presence <- factor(pisaster_sum$presence, levels = c(0, 1), labels = c("Absence

# Create the bar chart with proper grouping and fill aesthetics
ggplot(pisaster_sum, aes(x = presence, fill = presence, group = presence)) +
  geom_bar() +
  facet_wrap(~ year_bin, scales = "free_y") +  # Create separate plots for each year_bin
  labs(title = "Proportion of Sites with Presence vs Absence by Year",
       x = "Presence/Absence",
       y = "Count") +
  scale_fill_manual(values = c("Presence (1)" = "#6A0DAD", "Absence (0)" = "#FF8C00")) +
  theme_ocean +
  theme(
    panel.grid.minor = element_blank(),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5)
  )
```

Proportion of Sites with Presence vs Absence by Year
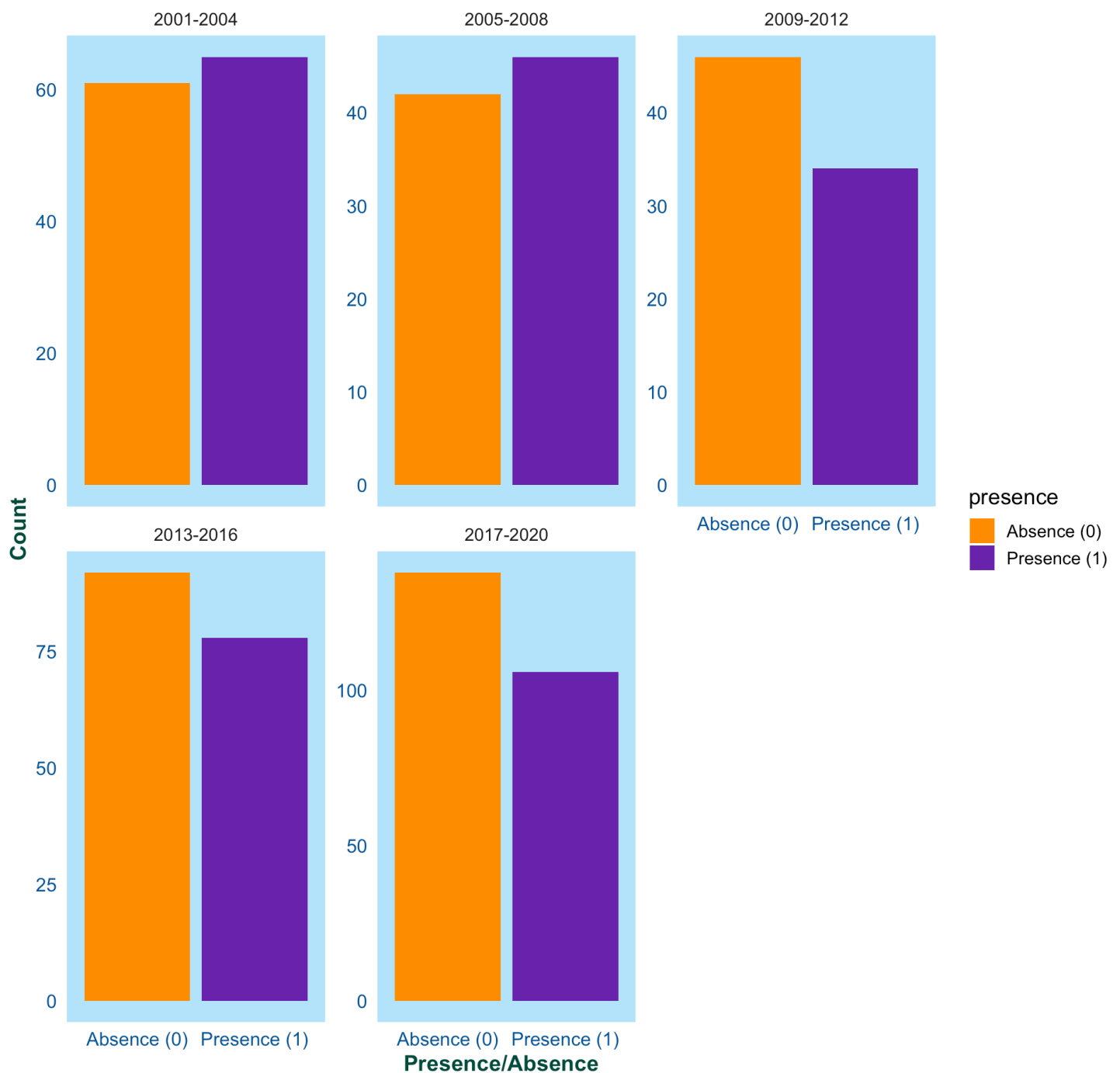
7. Answer the following questions:

a. What challenges did you encounter or anticipate encountering as you continue to build / iterate on your visualizations in R? If you struggled with mocking up any of your three visualizations (from #6, above), describe those challenges here.

For visualization 1, I need to learn how plot bars as images (aka starfishes). For visualization 2, I need to learn how to plot density ridges geographically, to represent species ranges. For visualization 3, I need to learn how change my bar plot to a line plot with the relevant data.

b. What ggplot extension tools / packages do you need to use to build your visualizations? Are there any that we haven't covered in class that you'll be learning how to use for your visualizations?

In additon to base ggplot, I plan to use gganimate, which we have not gone over yet. I will show how latitudnal plots shift over time.

    c. What feedback do you need from the instructional team and / or your peers to ensure that your intended message is clear?

I think I would like more feedback on if my concepts are worth pursuing in the first place. Additionally, I wonder what fonts and colors would make this most visually appealing.