

Exploratory Data Analysis

This Notebook contains the code used to analyze the data from the study

```
# Import libraries for ggplot
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --

## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v stringr 1.4.1
## v tidyr   1.2.1      v forcats 0.5.2
## v readr   2.1.3

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# Opening the csv file
data <- read_csv("TrawlCatch_SpringPreyfishBottomTrawl.csv")
```

```
## Rows: 12440 Columns: 18
## -- Column specification -----
## Delimiter: ","
## chr (3): vesselName, lifeStageName, commonName
## dbl (15): opId, year, serial, opDate, latitude, longitude, fishingTemperatur...
##
## 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.
```

Exploratory Data Analysis 1: Graphing Relative abundance of native specie(slimy sculpin) and the Round Goby

```

# Getting Rid of Unneccessary Columns that are not applicable to
# current data analysis
curated_data <- data %>%
  select(-c(opId,vesselName,serial, opDate, speed_mpsec, wingSpreadModeled_m,extraBottomContactTime_sec

names(curated_data)

## [1] "year"          "commonName" "weight_g"

# now contains the columns year, commonName(Of the Specie), and
# the total collected weight in grams

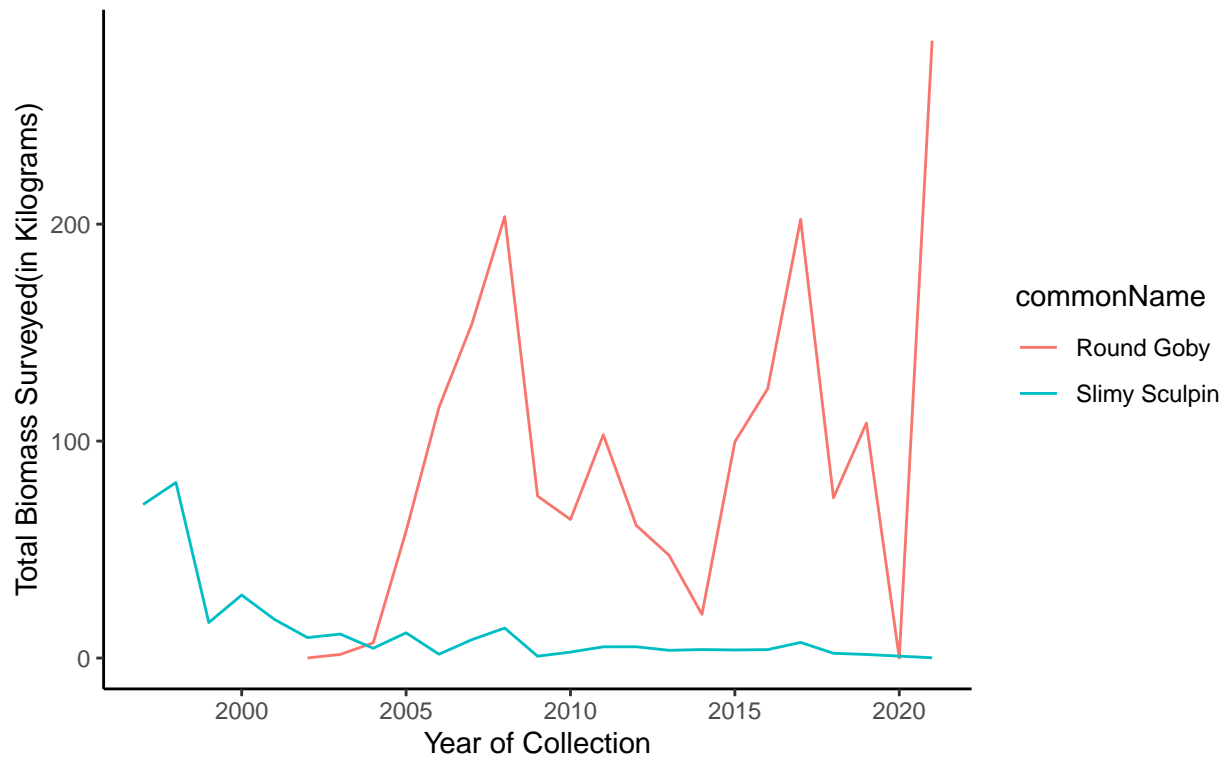
# grouping by desired species and year of collection
total_weight_by_species <- curated_data %>%
  group_by(commonName, year) %>%
  filter(commonName %in% c("Round Goby",
                           "Slimy Sculpin")) %>%
  tally(weight_g) %>%
  # current weight is measured in grams; change to kg

  mutate(kg_weight = n/1000)

# plotting by year and collected biomass; lines represent the two species
graphed <- total_weight_by_species %>%
  ggplot(aes(x= year, y = kg_weight, colour=commonName)) +
    geom_line()+ labs(title ="Yearly Collected Biomass of the Round Goby and Slimy Sculpin in La
graphed

```

Yearly Collected Biomass of the Round Goby and Slimy Sculpin in Lake Or From 1997–2022



From the Graph, we can see a decline in collected slimy sculpin biomass compared to Round Goby biomass, although the goby population appears to be in flux (with a extremely large increase in recent years). Perhaps other factors are at play? We will explore relationships such as temperature and biomass, and how different sites potentially also impacts biomass.