Data Exploration

```
# libraries used
library(ggplot2)
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(ggfortify)
library(tidyverse)
```

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

## v tibble  3.1.8     v purrr   0.3.5
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.3     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
fish.data <- read.csv("TrawlCatch_SpringPreyfishBottomTrawl.csv")
fish.data.raw <- fish.data %>%
  mutate(year=as.factor(year)) #didn't want year to be a continuous variable, I wanted them as discrete
fish.data <- fish.data.raw %>%
  filter(!is.na(fishingTemperature_C), !is.na(latitude), !is.na(longitude), commonName!= "No fish caught
#I am also removing unidentified or misc. fishes
```

```
fish.data %>%
  group_by(commonName) %>%
  tally() %>%
  arrange(desc(n)) #this helped us see how many observations we had per species
```

```
## # A tibble: 63 x 2
##    commonName                  n
##    <chr>                   <int>
##   1 Alewife                6659
##   2 Rainbow smelt          4646
##   3 Slimy sculpin          2328
##   4 Lake trout             1635
##   5 Round goby             1135
##   6 Dreissena spp.         1102
##   7 Johnny darter          1048
##   8 Trout-perch             714
##   9 Deepwater sculpin       618
##  10 Threespine stickleback   482
## # ... with 53 more rows
```

```
fish.list <- as.data.frame(unique(fish.data$commonName)) #this just made a data frame of the list so we
fish.data.exonat <- fish.data %>% #Here we made a new column that marks each fish species as exotic or
  mutate(inv.status = case_when(
    endsWith(commonName, "Alewife") ~ "exotic",
    endsWith(commonName, "Sea lamprey") ~ "exotic",
    endsWith(commonName, "Chinook salmon") ~ "exotic",
    endsWith(commonName, "Rainbow trout (Steelhead)") ~ "exotic",
    endsWith(commonName, "Carp") ~ "exotic",
    endsWith(commonName, "Brown trout") ~ "exotic",
    endsWith(commonName, "Rainbow smelt") ~ "exotic",
    endsWith(commonName, "Coho salmon") ~ "exotic",
    endsWith(commonName, "White perch") ~ "exotic",
    endsWith(commonName, "Blueback herring") ~ "exotic",
    endsWith(commonName, "Chain pickerel") ~ "exotic",
    endsWith(commonName, "Round goby") ~ "exotic",
    endsWith(commonName, "Tubenose goby") ~ "exotic",
    endsWith(commonName, "Threespine stickleback") ~ "native",
    endsWith(commonName, "Emerald shiner") ~ "native",
    endsWith(commonName, "Lake whitefish") ~ "native",
    endsWith(commonName, "Deepwater sculpin") ~ "native",
    endsWith(commonName, "Lake trout") ~ "native",
    endsWith(commonName, "Burbot") ~ "native",
    endsWith(commonName, "Slimy sculpin") ~ "native",
    endsWith(commonName, "Emerald shiner") ~ "native",
    endsWith(commonName, "Cisco (lake herring)") ~ "native",
    endsWith(commonName, "Whitefishes") ~ "native",
    endsWith(commonName, "Johnny darter") ~ "native",
    endsWith(commonName, "Trout-perch") ~ "native",
    endsWith(commonName, "Yellow perch") ~ "native",
    endsWith(commonName, "Spottail shiner") ~ "native"
    ))
fish.data.exonat %>%
  filter(is.na(inv.status)) %>%
  group_by(commonName) %>%
  tally() %>%
  arrange(desc(n))
```

```
## # A tibble: 37 x 2
##    commonName                     n
##    <chr>                      <int>
##  1 Dreissena spp.              1102
##  2 White bass                    65
##  3 White sucker                 65
##  4 Rockbass                     64
##  5 American eel                 60
##  6 Walleye                      53
##  7 Freshwater drum              49
##  8 Vegetation/plant material    43
##  9 Brown bullhead               31
## 10 Pumpkinseed                  29
## # ... with 27 more rows
```

```
#Checking to see which ones I hadn't researched yet to make sure I did not miss any important ones.
#Dreissena are mussels and we are only focused on fishes so we will be cutting those out anyway
#We ignore everything below 200 observations on this list because they do not have enough observations
fish.data.exonat %>%  #now that we have labeled each species, we can display our native species of inte
  filter(inv.status=="native") %>%
  group_by(commonName) %>%
  tally() %>%
  arrange(desc(n))
```

```
## # A tibble: 13 x 2
##    commonName                 n
##    <chr>                  <int>
##  1 Slimy sculpin           2328
##  2 Lake trout              1635
##  3 Johnny darter           1048
##  4 Trout-perch              714
##  5 Deepwater sculpin        618
##  6 Threespine stickleback   482
##  7 Yellow perch             396
##  8 Spottail shiner          271
##  9 Lake whitefish           131
## 10 Emerald shiner           110
## 11 Cisco (lake herring)      65
## 12 Burbot                    19
## 13 Whitefishes                1
```

```
#based on this, we can choose only species with more than 300 observations. In this case that means Yel
```

```
fish.data.clean <- fish.data.exonat %>% #this is now the data we are interested in, including only the
  filter(commonName=="Yellow perch" | commonName=="Threespine stickleback" | commonName=="Deepwater scul

head(fish.data.clean)
```

```
##    opId year                                         vesselName serial   opDate
## 1 30247 1992 Kaho                                                    2 19920421
## 2 30248 1992 Kaho                                                    3 19920421
## 3 48124 1984 Kaho                                                   35 19840420
## 4 48124 1984 Kaho                                                   35 19840420
## 5 48124 1984 Kaho                                                   35 19840420
## 6 48125 1984 Kaho                                                   36 19840420
##   latitude longitude fishingTemperature_C fishingDepth_m towTime_min
## 1 43.38000 -77.51833                  2.9            150          10
## 2 43.38667 -77.55833                  2.6            130          10
## 3 43.37167 -78.75000                  2.7             55          10
## 4 43.37167 -78.75000                  2.7             55          10
## 5 43.37167 -78.75000                  2.7             55          10
## 6 43.37833 -78.75000                  2.9             65          10
##   speed_mpsec wingSpreadModeled_m extraBottomContactTime_sec
## 1    1.251712            9.081622                   9.191982
## 2    1.251712            9.081138                   9.185680
## 3    1.251712            8.979826                   8.960727
## 4    1.251712            8.979826                   8.960727
## 5    1.251712            8.979826                   8.960727
## 6    1.251712            9.029780                   9.039374
##   areaSampledDoors_m2           lifeStageName    commonName   n weight_g
## 1           11803.029 Life Stage Not Recorded Slimy sculpin  33      309
## 2           10730.085 Life Stage Not Recorded Slimy sculpin 193     1919
## 3            7756.949 Life Stage Not Recorded    Lake trout  20     7768
## 4            7756.949 Life Stage Not Recorded Johnny darter 256      380
## 5            7756.949 Life Stage Not Recorded Slimy sculpin 280     1337
## 6            8093.266 Life Stage Not Recorded    Lake trout   2      390
##   inv.status
## 1     native
## 2     native
## 3     native
## 4     native
## 5     native
## 6     native
```

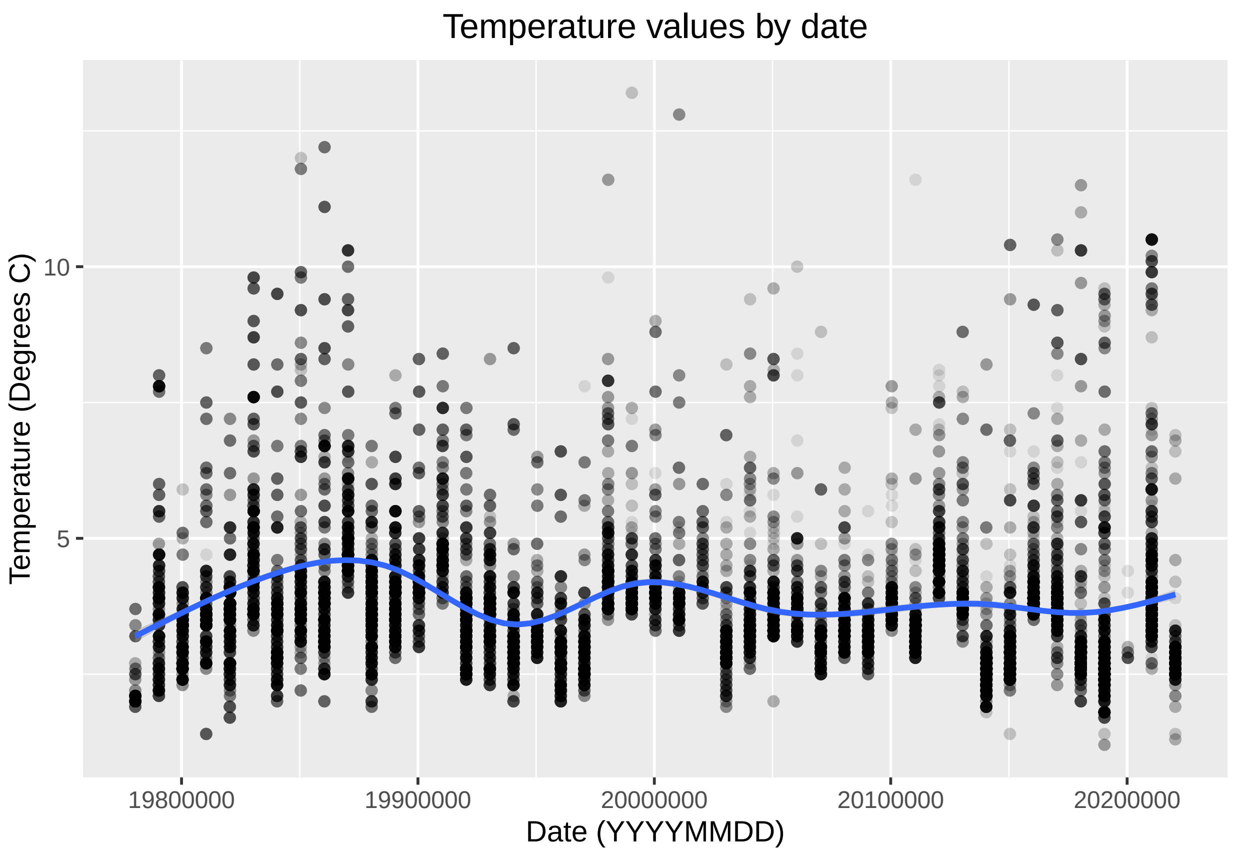
```
unique(fish.data$year) #we have data from 1997 to 2022
```

```
##  [1] 1992 1984 1985 2004 1983 1979 1978 1996 1995 1998 1999 1994 1986 1993 2000
## [16] 1988 1987 1997 1982 1981 1980 1990 1989 2005 2006 2003 2002 2001 1991 2009
## [31] 2011 2007 2010 2013 2008 2012 2014 2017 2018 2015 2016 2022 2019 2021 2020
## 45 Levels: 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 ... 2022
```

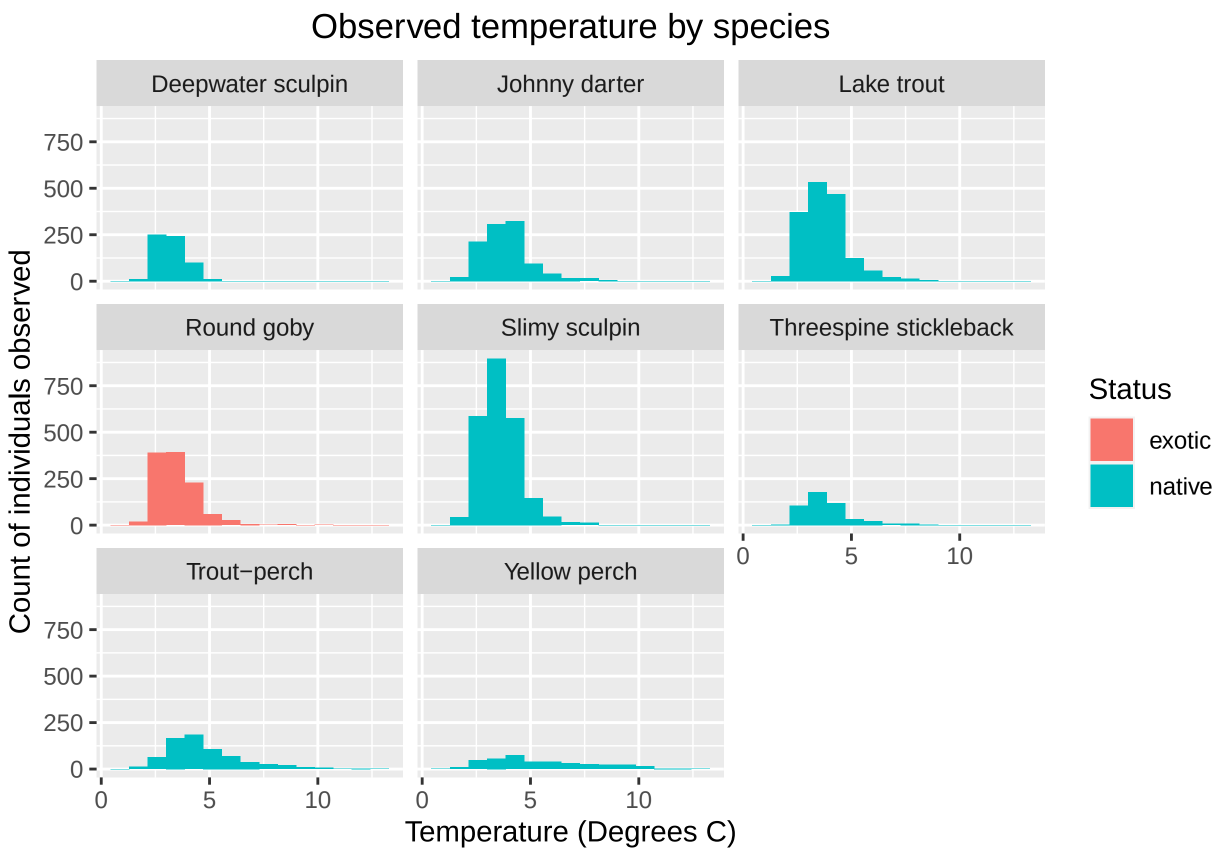
```
#the dates should be converted into a more readable format. I just don't know how to do that so I need

ggplot(fish.data, aes(x=opDate, y=fishingTemperature_C)) + geom_point(alpha=0.1) + geom_smooth() + labs
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



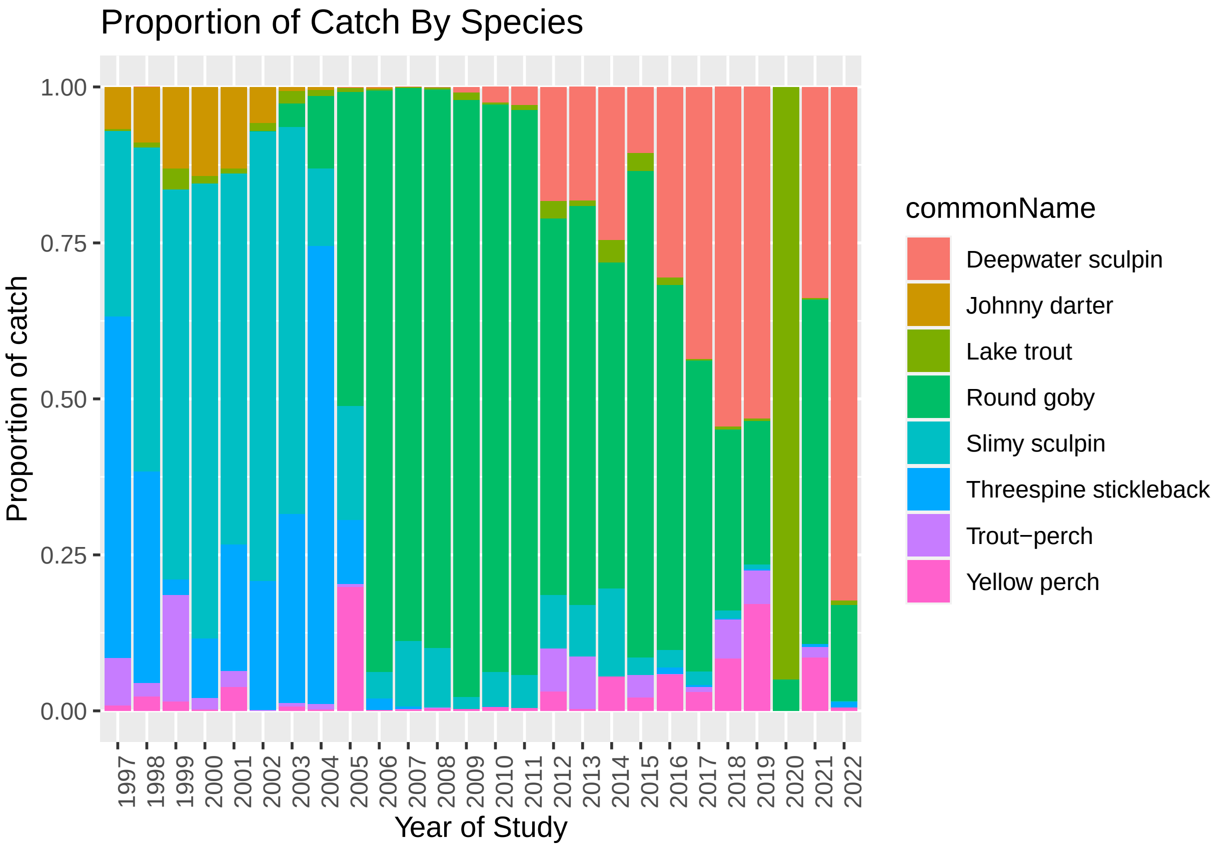
```
ggplot(fish.data.clean, aes(x=fishingTemperature_C, fill=inv.status)) + geom_histogram(bins=15) + facet_
```



```
#plotting the count of observations of each species depending on the temperature.
```

Lets visualize the percentage of each species

```
# Proportion of the total catch from the first siting in 1997 -->
# Based on abundance
fish.data.clean %>%
  group_by(commonName, year) %>%
  # filtering out year based on first time a goby was sighted --> in 1997
  filter(year %in% seq(1997, 2022)) %>%
  tally(n) %>%  # tallying up occurances of each species
  ggplot(aes(x=year, y=n, fill=commonName)) + geom_bar(position="fill", stat="identity") + labs(title="
```


Proportion of Catch By Species
Proportion of catch
1.00
0.75
0.50
0.25
0.00
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year of Study
commonName
Deepwater sculpin
Johnny darter
Lake trout
Round goby
Slimy sculpin
Threespine stickleback
Trout−perch
Yellow perch