

# BDataFridayScript

Andy Jones

8/13/2020

```
knitr::opts_chunk$set(echo = TRUE,warning = FALSE, message = FALSE)

#Loading libraries
library(tidyverse)
library(umap)
library(patchwork)
library(BBmisc)

#Pulling in the data and formatting dates
B_Data <- read_csv("~/Downloads/B_Data.csv",
  col_types = cols(EFFORT_ENTRY_DATE_GMT = col_datetime(format = "%Y-%m-%d %H:%M:%S"),
    END_HAUL_DATE_GMT = col_datetime(format = "%Y-%m-%d %H:%M:%S"),
    SAIL_DATE_LCL = col_datetime(format = "%Y-%m-%d %H:%M:%S"),
    START_HAUL_DATE_GMT = col_datetime(format = "%Y-%m-%d %H:%M:%S"),
    START_HAUL_LAT = col_number(), START_SET_DATE_GMT = col_datetime(format = "%Y-%m-%d %H:%M:%S")))

#Making an effort ID column
B_Data <- B_Data %>% mutate(EFFORT_ID=as.numeric(paste(TRIP_ID,EFFORT_NUM,sep='')))

#Looking at which species are most commonly caught
Species_tally <- B_Data %>% group_by(COMMON_NAME) %>% tally() %>% arrange(-n)

Species_tally

## # A tibble: 31 x 2
##   COMMON_NAME      n
##   <chr>          <int>
## 1 TILEFISH (GOLDEN TILEFISH)  935
## 2 EEL,CONGER              795
## 3 DOGFISH,CHAIN           789
## 4 HAKE,SPOTTED            745
## 5 CRAB,JONAH              591
## 6 DOGFISH,SPINY           524
## 7 SKATE,BARNDOOR          404
## 8 DOGFISH,SMOOTH          321
## 9 LOBSTER,AMERICAN        178
## 10 HAKE,SILVER (WHITING)    149
## # ... with 21 more rows

#Making a function to explore the data with UMAP
Trip_Weight_Sum_UMAP <- function(Species=c('TILEFISH (GOLDEN TILEFISH)'),
  Variable='YEAR',
  neighbor='20') {
```

```

#Setting up the data in a 'wide' format and replacing NAs with zeros
B_Data %>% group_by(EFFORT_ID,COMMON_NAME) %>%
  filter(COMMON_NAME %in% Species) %>%
  dplyr::summarise(SUM_WEIGHT=sum(HAIL_AMOUNT)) %>%
  pivot_wider(names_from = COMMON_NAME,values_from = SUM_WEIGHT) -> data_sum
data_sum %>% select(-EFFORT_ID) %>%
  mutate(
    across(everything(), ~replace_na(.x, 0))
  ) %>%
  umap(.,n_neighbor=neighbor) -> a

#looking at the data for debugging
#data_sum %>% bind_cols(a$layout %>% as_tibble()) %>%
  #left_join(.,B_Data %>% select(EFFORT_ID,Variable) %>% distinct()) %>% print()

#Plotting out the UMAP results and saving it to an object
data_sum %>% bind_cols(a$layout %>% as_tibble()) %>%
  left_join(.,B_Data %>% select(EFFORT_ID,YEAR) %>% distinct()) %>%
  mutate(YEAR=as.factor(YEAR)) %>%
  ggplot(aes_string('V1', 'V2',colour=Variable)) +
  geom_point(alpha = 0.7,position = position_dodge2(width = 0.1)) +
  stat_ellipse(level = 0.99) +
  theme(legend.justification = 'top') +
  labs(colour='',x='',y='') +
  scale_colour_viridis_d() -> umap_plot

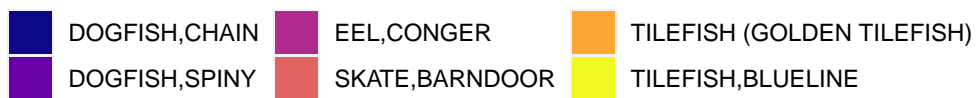
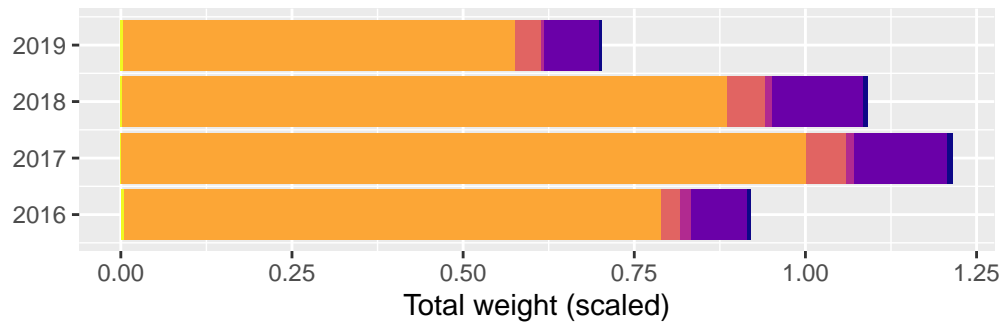
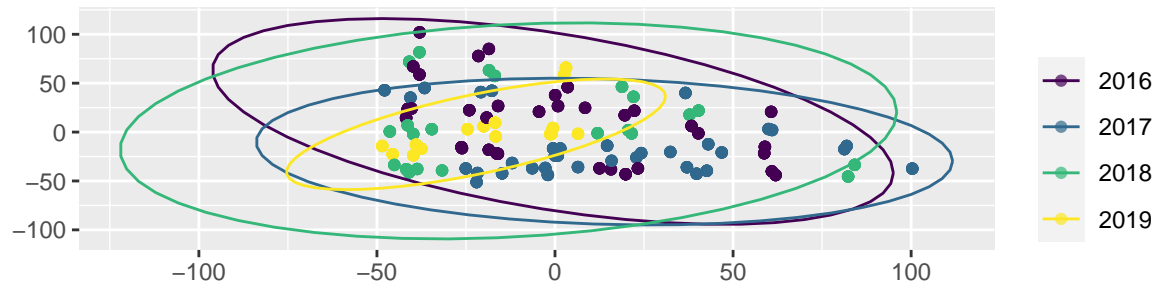
#Using the original data again to make a bar plot and saving it
B_Data %>% group_by(YEAR,COMMON_NAME) %>%
  filter(COMMON_NAME %in% Species) %>%
  dplyr::summarise(SUM_WEIGHT=sum(HAIL_AMOUNT)) %>%
  ungroup() %>%
  #group_by(YEAR) %>%
  mutate(SCALED_WEIGHT=normalize(SUM_WEIGHT,method='range',range=c(0,1))) %>%
  ggplot(aes_string(x=Variable,y='SCALED_WEIGHT',fill='COMMON_NAME')) + geom_bar(stat='identity') +
  coord_flip() +
  theme(legend.position="bottom") +
  labs(fill='',x='',y='Total weight (scaled)') +
  scale_fill_viridis_d(option = 'plasma') -> bar_plot

#putting the two plots together
umap_plot / bar_plot

}

#Trying out the function
Trip_Weight_Sum_UMAP(Species=c('TILEFISH (GOLDEN TILEFISH)',
                                'DOGFISH,SPINY','TILEFISH,BLUELINE',
                                "EEL,CONGER","DOGFISH,CHAIN",
                                "SKATE,BARNDORR"),
  Variable='YEAR',
  neighbor=10)

```



```
#Second set of species with the function
Trip_Weight_Sum_UMAP(Species=Species_tally$COMMON_NAME %>% head(),
  Variable='YEAR',
  neighbor=5)
```

