

# tidy\_coral\_analysis

*Tatiana*

*June 19, 2019*

Libraries

```
library(tidyverse)
library(janitor) # install.packages('janitor')
library(skimr) # install.packages('skimr')
library(stringr) # added when we needed it for benthic data
```

Load data

```
## benthic data
benthic_url <- 'https://www.nodc.noaa.gov/archive/arc0054/0104255/1.1/data/0-data/cd08/1003080aAla03m.C'

## buoy data
buoy_url <- 'http://www.ndbc.noaa.gov/view_text_file.php?filename=mokh1h2010.txt.gz&dir=data/historical'
```

```
benthic_raw <- read_csv(benthic_url)
```

```
## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   `ID Name` = col_character(),
##   Point = col_double(),
##   X = col_double(),
##   Y = col_double(),
##   `File Name` = col_character(),
##   `Total Points` = col_double(),
##   `ID Date` = col_character()
## )

## See spec(...) for full column specifications.
```

```
head(benthic_raw)
```

```
## # A tibble: 6 x 25
##   `Site Name` Station `Frame No.` `Image Date` `ID Name` `ID Code` Point
##   <lgl>         <lgl>   <lgl>         <lgl>         <chr>    <lgl>    <dbl>
## 1 NA          NA      NA          NA          Pocillop~ NA        1
## 2 NA          NA      NA          NA          Turf alg~ NA        2
## 3 NA          NA      NA          NA          Corallin~ NA        3
## 4 NA          NA      NA          NA          Corallin~ NA        4
## 5 NA          NA      NA          NA          Corallin~ NA        5
## 6 NA          NA      NA          NA          Turf alg~ NA        6
## # ... with 18 more variables: X <dbl>, Y <dbl>, Intensity <lgl>,
## #   Red <lgl>, Green <lgl>, Blue <lgl>, `File Name` <chr>, `Total
## #   Points` <dbl>, `ID Date` <chr>, `Site ID` <lgl>, `Site Code` <lgl>,
## #   `Time Code` <lgl>, Institution <lgl>, `User Name` <lgl>,
## #   Habitat <lgl>, WQS <lgl>, Length <lgl>, Depth <lgl>
```

clean\_names () this function delete the space between names and replace them with "\_"

```
## the `janitor` package's `clean_names` function
```

```
benthic <- benthic_raw %>%
  janitor::clean_names()
```

```
names(benthic)
```

```
## [1] "site_name"      "station"        "frame_no"       "image_date"
## [5] "id_name"        "id_code"        "point"          "x"
## [9] "y"             "intensity"      "red"            "green"
## [13] "blue"          "file_name"      "total_points"   "id_date"
## [17] "site_id"       "site_code"      "time_code"      "institution"
## [21] "user_name"     "habitat"        "wqs"            "length"
## [25] "depth"
```

```
benthic <- benthic %>%
  select(id_name, point, x, y, id_date)
```

remove the “#” before the numbers

```
benthic <- benthic %>%
  mutate(date = stringr::str_remove_all(id_date, "#"))
```

```
head(benthic)
```

```
## # A tibble: 6 x 6
##   id_name      point      x      y id_date      date
##   <chr>      <dbl> <dbl> <dbl> <chr>      <chr>
## 1 Pocillopora meandrina      1  1773  1000 #2010-03-12# 2010-03-12
## 2 Turf algae                2  2308   194 #2010-03-12# 2010-03-12
## 3 Coralline algae           3  1700  1782 #2010-03-12# 2010-03-12
## 4 Coralline algae           4  2470   584 #2010-03-12# 2010-03-12
## 5 Coralline algae           5   314  1145 #2010-03-12# 2010-03-12
## 6 Turf algae                6   198  1660 #2010-03-12# 2010-03-12
```

Explore data. Package skimr has a function to summarise data

```
summary(benthic)
```

```
##   id_name      point      x      y
## Length:4925    Min.   : 1    Min.   : 1    Min.   : 1.0
## Class :character 1st Qu.: 7    1st Qu.: 627   1st Qu.: 489.0
## Mode :character  Median :13   Median :1266  Median : 958.0
##                Mean   :13   Mean   :1271   Mean   : 963.8
##                3rd Qu.:19   3rd Qu.:1918  3rd Qu.:1434.0
##                Max.   :25   Max.   :2560   Max.   :1920.0
##   id_date      date
## Length:4925    Length:4925
## Class :character Class :character
## Mode :character Mode :character
##
##
##
```

```
skimr::skim(benthic)
```

```
## Skim summary statistics
## n obs: 4925
```

```
## n variables: 6
##
## -- Variable type:character -----
## variable missing complete      n min max empty n_unique
##      date          0      4925 4925  10  10      0        6
##     id_date        0      4925 4925  12  12      0        6
##     id_name        0      4925 4925   4  21      0       16
##
## -- Variable type:numeric -----
## variable missing complete      n      mean      sd p0 p25 p50 p75 p100
##      point          0      4925 4925   13      7.21  1   7   13   19   25
##          x          0      4925 4925 1271.09 743.82  1 627 1266 1918 2560
##          y          0      4925 4925  963.8  546.82  1 489  958 1434 1920
##      hist
## <U+2587><U+2586><U+2586><U+2586><U+2586><U+2586><U+2586><U+2586>
## <U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587>
## <U+2586><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587>
```

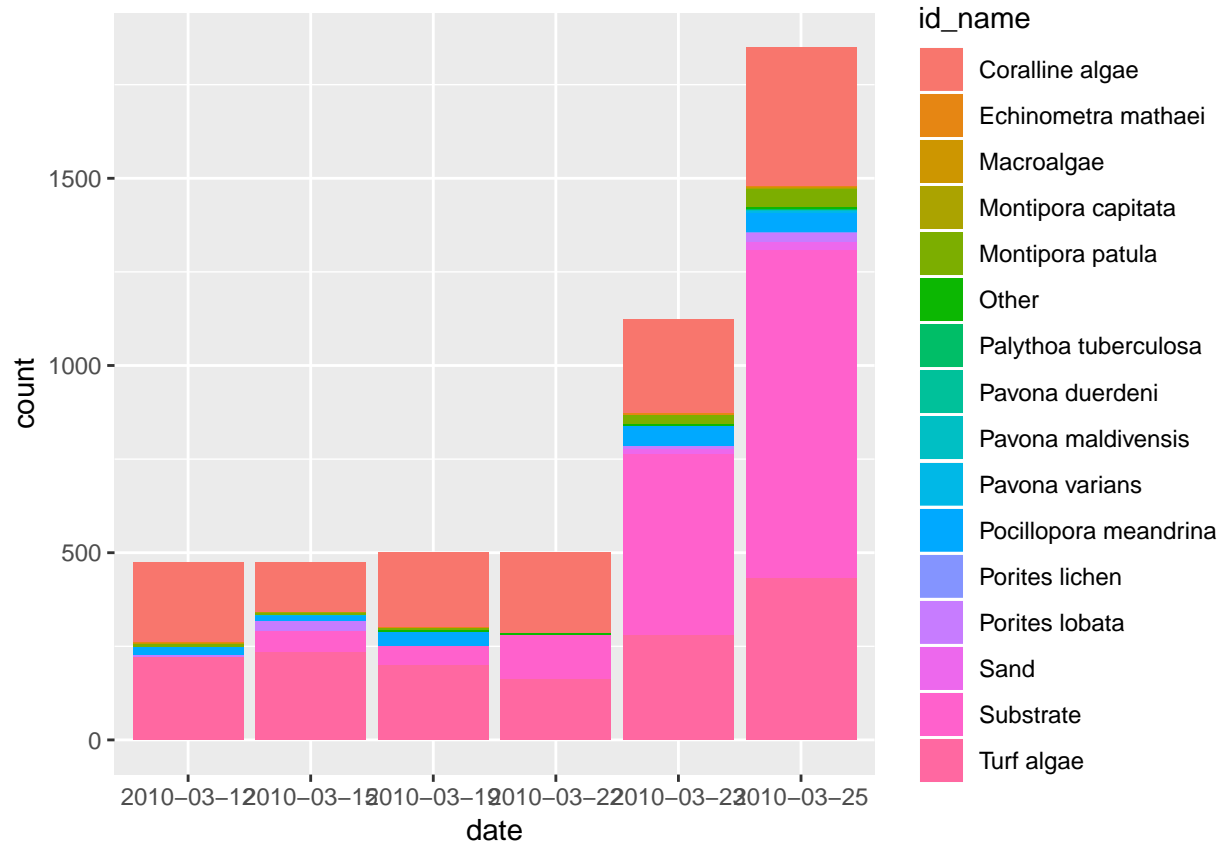
Which species are represented

```
unique(benthic$id_name)
```

```
## [1] "Pocillopora meandrina" "Turf algae"
## [3] "Coralline algae"      "Macroalgae"
## [5] "Montipora capitata"   "Montipora patula"
## [7] "Echinometra mathaei" "Porites lobata"
## [9] "Sand"                 "Substrate"
## [11] "Pavona varians"       "Other"
## [13] "Palythoa tuberculosa" "Pavona maldivensis"
## [15] "Pavona duerdeni"     "Porites lichen"
```

plot And to get a sense of our data let's just have a quick plot of species count by date:

```
ggplot(benthic, aes(date, fill=id_name))+
  geom_bar()
```



```
#Buoy data
```

```
buoy <- readr::read_csv(buoy_url)
```

```
## Parsed with column specification:
```

```
## cols(
```

```
##   `#YY` MM DD hh mm WDIR WSPD GST WVHT DPD APD MWD PRES ATMP WTMP DEWP VIS TIDE` = col_c
```

```
## )
```

```
head(buoy) # hmmm this doesn't look right! Why not?
```

```
## # A tibble: 6 x 1
```

```
##   `#YY` MM DD hh mm WDIR WSPD GST WVHT DPD APD MWD PRES ATMP WTMP~
```

```
##   <chr>
```

```
## 1 #yr mo dy hr mn degT m/s m/s m sec sec deg hPa degC degC ~
```

```
## 2 2010 01 01 00 00 999 99.0 99.0 99.00 99.00 99.00 999 1012.4 999.0 25.4 ~
```

```
## 3 2010 01 01 00 06 999 99.0 99.0 99.00 99.00 99.00 999 1012.4 999.0 25.4 ~
```

```
## 4 2010 01 01 00 12 999 99.0 99.0 99.00 99.00 99.00 999 1012.2 999.0 25.4 ~
```

```
## 5 2010 01 01 00 18 999 99.0 99.0 99.00 99.00 99.00 999 1012.2 999.0 25.4 ~
```

```
## 6 2010 01 01 00 24 999 99.0 99.0 99.00 99.00 99.00 999 1012.1 999.0 25.4 ~
```

```
buoy_raw <- read_table(buoy_url)
```

```
## Parsed with column specification:
```

```
## cols(
```

```
##   `#YY` = col_character(),
```

```
##   MM = col_character(),
```

```
##   DD = col_character(),
```

```
## hh = col_character(),
## mm = col_character(),
## `WDIR WSPD GST` = col_character(),
## WVHT = col_character(),
## DPD = col_character(),
## APD = col_character(),
## MWD = col_character(),
## PRES = col_character(),
## ATMP = col_character(),
## WTMP = col_character(),
## DEWP = col_character(),
## VIS = col_character(),
## TIDE = col_character()
## )
```

```
head(buoy_raw)
```

```
## # A tibble: 6 x 16
##   `#YY` MM    DD    hh    mm    `WDIR WSPD GST` WVHT  DPD    APD    MWD
##   <chr> <chr> <chr> <chr> <chr> <chr>          <chr> <chr> <chr> <chr>
## 1 #yr   mo    dy    hr    mn    degT m/s  m/s    m    sec  sec  deg
## 2 2010  01    01    00    00    999 99.0 99.0    99.00 99.00 99.00 999
## 3 2010  01    01    00    06    999 99.0 99.0    99.00 99.00 99.00 999
## 4 2010  01    01    00    12    999 99.0 99.0    99.00 99.00 99.00 999
## 5 2010  01    01    00    18    999 99.0 99.0    99.00 99.00 99.00 999
## 6 2010  01    01    00    24    999 99.0 99.0    99.00 99.00 99.00 999
## # ... with 6 more variables: PRES <chr>, ATMP <chr>, WTMP <chr>,
## #   DEWP <chr>, VIS <chr>, TIDE <chr>
```