Exploratory Data Analysis-Seasonal

William Norfolk

10/10/2019

Load the required libraries for exploratory analysis.

library(readxl)  
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 1.2.1 --

## v ggplot2 3.2.1 v readr 1.3.1  
## v tibble 2.1.3 v purrr 0.3.2  
## v tidyr 0.8.3 v stringr 1.4.0  
## v ggplot2 3.2.1 v forcats 0.4.0

## -- Conflicts ---------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(forcats)  
library(ggthemes)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(knitr)  
library(naniar)  
library(broom)  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(zoo)

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

Load the processed data from the RDS. Then take a look!

WQ\_clean\_data <- readRDS("../../data/processed\_data/processeddata.rds")  
  
glimpse(WQ\_clean\_data)

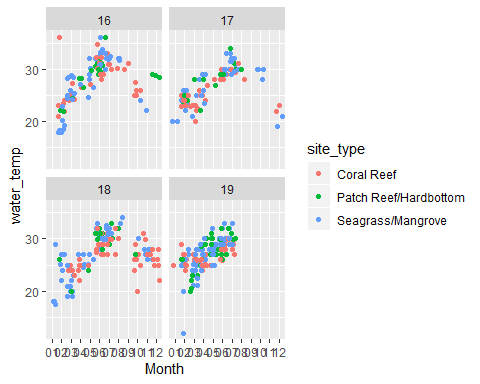
## Observations: 522  
## Variables: 15  
## $ Month <chr> "01", "02", "02", "02", "02", "02", "02", "02...  
## $ Day <chr> "08", "08", "08", "08", "09", "09", "09", "12...  
## $ Year <chr> "16", "16", "16", "16", "16", "16", "16", "16...  
## $ military\_time <dbl> 1415, 1515, 1550, 1555, 1001, 1015, 1022, 103...  
## $ location <chr> "Boat Ramp", "Grecian Dry Rocks", "Grecian Dr...  
## $ instructor\_name <chr> "Katy, Sarah, Driver", "Chelsea", "Katy, Tomm...  
## $ group\_name <chr> "NA", "McLean High School", "McLean High Scho...  
## $ ph <dbl> 8.0, 8.4, 8.2, 8.4, 8.0, 8.0, 8.0, 8.0, 8.4, ...  
## $ ammonia <dbl> 0.00, 0.00, 0.00, 0.00, 0.25, 0.00, 0.00, 0.0...  
## $ dissolved\_oxygen <dbl> 5.0, 4.0, 4.0, 6.0, 8.0, 4.0, 5.0, 6.0, 6.0, ...  
## $ water\_temp <dbl> NA, 23.5, 21.0, 36.0, 18.0, 18.0, 18.0, 18.3,...  
## $ salinity <dbl> 36, 40, 44, 35, 33, 30, 33, 35, 40, 30, 35, 2...  
## $ equipment <chr> "kit", "kit", "kit", "kit", "kit", "kit", "ki...  
## $ island\_side <chr> "ocean", "ocean", "ocean", "ocean", NA, "bay"...  
## $ site\_type <chr> "Seagrass/Mangrove", "Coral Reef", "Coral Ree...

island\_side\_filter <- filter(WQ\_clean\_data, !is.na(island\_side))  
  
site\_type\_filter <- filter(WQ\_clean\_data, !is.na(site\_type))

Now that we have baseline visuals of the differet island sides and site types, lets start to look at some seasonal changes and see if we can identify the presence of Hurricane Irma in the data.

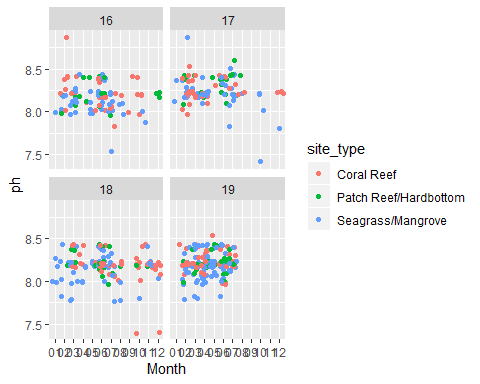
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = water\_temp, color = site\_type)) + facet\_wrap(~Year)

## Warning: Removed 12 rows containing missing values (geom\_point).



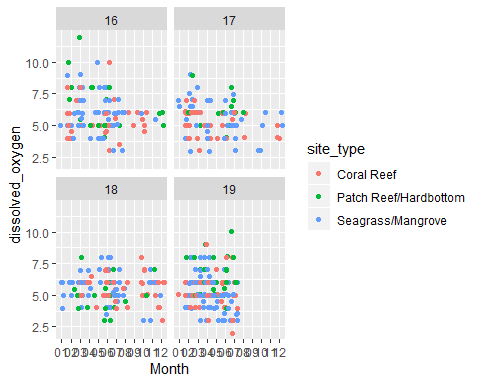
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = ph, color = site\_type)) + facet\_wrap(~Year)

## Warning: Removed 6 rows containing missing values (geom\_point).



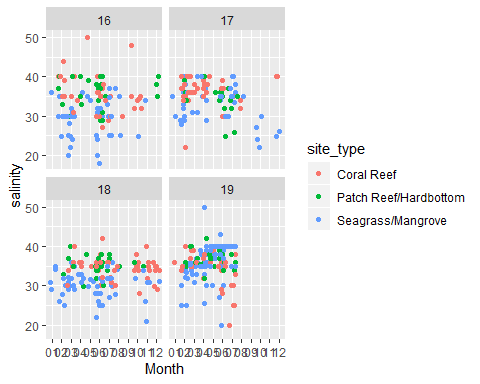
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = dissolved\_oxygen, color = site\_type)) + facet\_wrap(~Year)

## Warning: Removed 4 rows containing missing values (geom\_point).



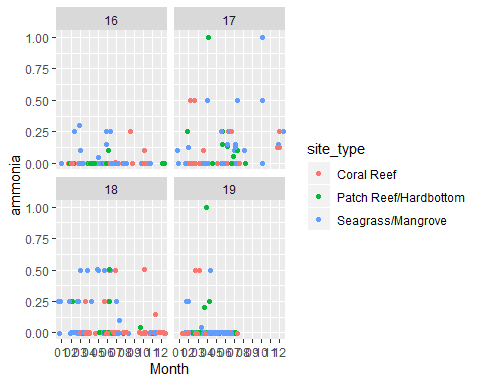
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = salinity, color = site\_type)) + facet\_wrap(~Year)

## Warning: Removed 15 rows containing missing values (geom\_point).



site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = ammonia, color = site\_type)) + facet\_wrap(~Year)

## Warning: Removed 10 rows containing missing values (geom\_point).



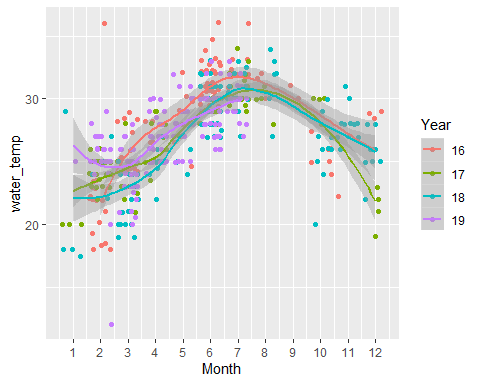
WQ\_clean\_data$Month <- as.numeric(as.character(WQ\_clean\_data$Month))

WQ\_clean\_data %>% ggplot(aes(x = Month, y = water\_temp, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 12 rows containing non-finite values (stat\_smooth).

## Warning: Removed 12 rows containing missing values (geom\_point).

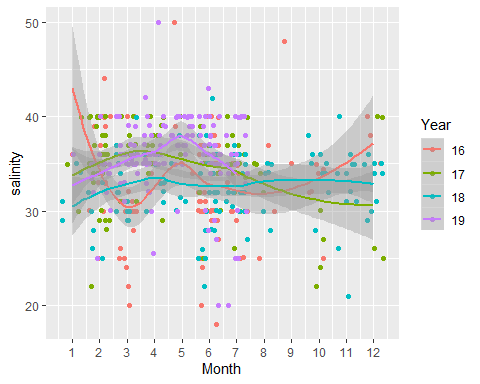


WQ\_clean\_data %>% ggplot(aes(x = Month, y = salinity, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 16 rows containing non-finite values (stat\_smooth).

## Warning: Removed 16 rows containing missing values (geom\_point).

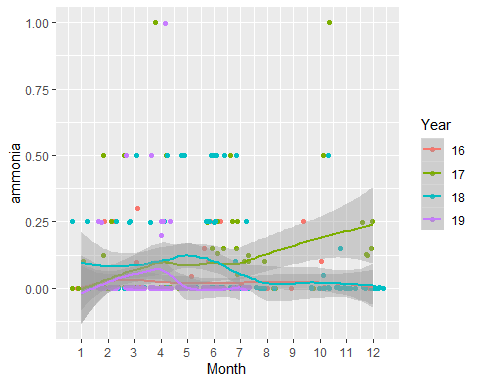


WQ\_clean\_data %>% ggplot(aes(x = Month, y = ammonia, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 10 rows containing non-finite values (stat\_smooth).

## Warning: Removed 10 rows containing missing values (geom\_point).

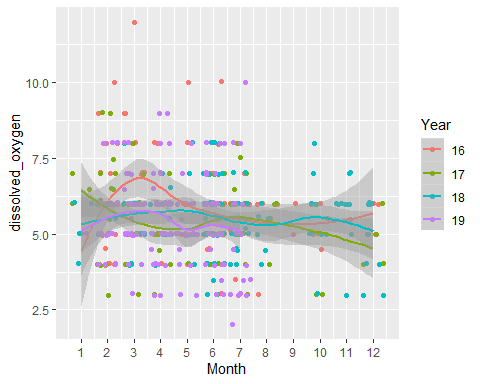


WQ\_clean\_data %>% ggplot(aes(x = Month, y = dissolved\_oxygen, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

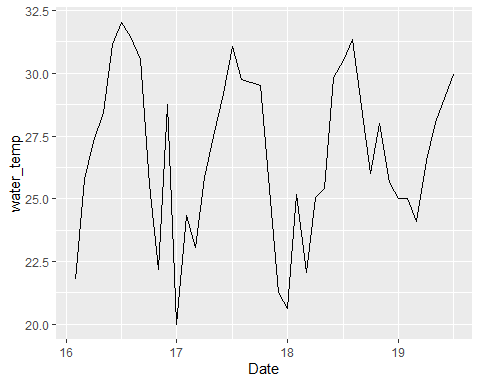
## Warning: Removed 4 rows containing non-finite values (stat\_smooth).

## Warning: Removed 4 rows containing missing values (geom\_point).



mean\_water\_temp <- aggregate(water\_temp ~ Month + Year, WQ\_clean\_data, mean)   
  
  
  
mean\_water\_temp$Date <- as.yearmon(paste(mean\_water\_temp$Month, mean\_water\_temp$Year, sep = "."), format = "%m.%Y")  
  
mean\_water\_temp %>% ggplot(aes(x = Date, y = water\_temp)) + geom\_line()

## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.



mean\_salinity <- aggregate(salinity ~ Month + Year, WQ\_clean\_data, mean)   
  
mean\_salinity$Date <- as.yearmon(paste(mean\_salinity$Month, mean\_salinity$Year, sep = "."), format = "%m.%Y")  
  
mean\_salinity %>% ggplot(aes(x = Date, y = salinity)) + geom\_line()

## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.

